The Telecommunications and Data Acquisition Progress Report 42-115

July–September 1993

Joseph H. Yuen
Editor

November 15, 1993

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.

Note From the Editor

An error on page iii of the last issue of this quarterly report, issue 42-114, misidentified the maiden name of Dr. Edward C. Posner's widow.

The correct name of Dr. Edward C. Posner's widow is Sylvia Kouzel Posner.

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) in planning, supporting research and technology, implementation, and operations. Also included are standards activity at JPL for space data and information systems and reimbursable DSN work performed for other space agencies through NASA. The preceding work is all performed for NASA's Office of Space Communications (OSC). The TDA Office also performs work funded by another NASA program office through and with the cooperation of OSC. This is the Orbital Debris Radar Program with the Office of Space Systems Development.

The TDA Office is directly involved in several tasks that directly support the Office of Space Science (OSS), with OSC funding DSN operational support. In radio science, *The TDA Progress Report* describes the spacecraft radio science program conducted using the DSN. For the High-Resolution Microwave Survey (HRMS), the report covers 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.

Finally, tasks funded under the JPL Director's Discretionary Fund and the Caltech President's Fund that 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:

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

OSC Cooperative Tasks:
- Orbital Debris Radar Program

OSS Tasks:
- Radio Science; High-Resolution Microwave Survey;
- Goldstone Solar System Radar; Radio Astronomy

Discretionary Funded Tasks

Contents

OSC TASKS
DSN Advanced Systems
TRACKING AND GROUND-BASED NAVIGATION

COMMUNICATIONS, SPACECRAFT–GROUND

DSN Systems Implementation
CAPABILITIES FOR EXISTING PROJECTS

NETWORK UPGRADE AND SUSTAINING

A Review of GPS-Based Tracking Techniques for TDRS Orbit Determination

B. J. Haines, S. M. Lichten, R. P. Malla, and S.-C. Wu
Tracking Systems and Applications Section

This article evaluates two fundamentally different approaches to the Tracking and Data Relay Satellite (TDRS) orbit determination utilizing Global Positioning System (GPS) technology and GPS-related techniques. In the first, a GPS flight receiver is deployed on the TDRS. The TDRS ephemerides are determined using direct ranging to the GPS spacecraft, and no ground network is required. In the second approach, the TDRSs broadcast a suitable beacon signal, permitting the simultaneous tracking of GPS and Tracking and Data Relay Satellite System satellites by ground receivers. Both strategies can be designed to meet future operational requirements for TDRS-II orbit determination.

I. Introduction

The Tracking and Data Relay Satellite System (TDRSS) is used by NASA to support positioning and data relay activities for a wide variety of Earth-orbiting spacecraft [1]. The present operational system is composed of two geosynchronous Tracking and Data Relay Satellites (TDRSs), TDRS-E and TDRS-W at 41 and 171 deg west longitude, respectively: a central ground station located at White Sands, New Mexico (WS); and remote tracking sites at Ascension Island, American Samoa, and Alice Springs, Australia. Accurate real-time positioning of the TDRSs is fundamental to the proper operation of the system and is achieved via the relay of coherent signals broadcast by unmanned transponders at the remote tracking sites. These remote beacons are collectively referred to as the Bilateration Ranging Transponder System (BRTS). Range and Doppler observations from BRTS are routinely scheduled by the central ground processing facility at White Sands, where they are used in conjunction with models of the forces perturbing the spacecraft motion to determine the TDRS positions. Evaluation of the TDRS ephemerides suggests that orbit consistency is maintained to better than 70 m using the operational BRTS method [2]. This level of precision is adequate for current applications; however, the technique requires valuable TDRS antenna time that could otherwise be used for servicing user spacecraft.

In recognition of the need for improved tracking for the next-generation TDRSS (TDRSS-II), a number of alternative methods have been explored [3–6]. Some of the techniques studied were originally developed for DSN tracking of high Earth and elliptical orbiters. The application of these techniques to the TDRS orbit-determination problem is not necessarily straightforward, due to the geosynchronous TDRS orbit and certain unique characteristics of the TDRS signals. The demand for improved accuracies provides an important motivation for these efforts. This requirement, however, is balanced by the ap-

peal of a simple, reliable, and autonomous system that requires no disruption of TDRSS user services and delivers the ephemerides in near-real time. One technique which promises the potential to meet these sometimes conflicting demands relies on technology from the U.S. Department of Defense Global Positioning System (GPS). Previous efforts addressing this option have produced encouraging results. Wu [7] proposed two GPS-related techniques for determining the orbits of high-altitude Earth satellites. He envisioned a wide variety of possible applications; hence, the breadth of the study prevented a thorough treatment of TDRSS. Recent efforts have focused directly on TDRSS, but software limitations precluded a complete evaluation [3]. In this article, the authors build on these earlier studies by revisiting their assumptions and revising them to insure they reflect the current state of the art. The new assumptions form the basis of a covariance study that exploits software and methodology that have evolved over the past decade as part of a program at the Jet Propulsion Laboratory (JPL) to support GPS-based tracking of Earth orbiters.

Results for two distinct solution strategies, as prescribed by Wu [7], are reported. In the first, a GPS receiver is deployed on the TDRSs, and the ephemerides are determined using direct measurements from the GPS to the TDRSs. In the second, the TDRSs broadcast wide-beam beacon signals which permit the simultaneous tracking of GPS and TDRSS satellites from a small ground network.

II. GPS-Based Techniques for Orbit Determination

For both military and civilian customers, the principal application of GPS is the precise positioning of ground sites and of moving vehicles near the Earth's surface [8]. The space segment of this system, which is due for completion early in 1994, will consist of 21 satellites and 3 active spares orbiting in 6 uniformly spaced orbit planes inclined at 55 deg with respect to the equator. The satellites, which are at an altitude of about 20,200 km, transmit unique navigational signals centered on two L-band carrier frequencies (L1 at 1575.42 MHz and L2 at 1227.60 MHz). Each carrier is modulated with pseudorandom square-wave codes: a coarse acquisition (C/A) code on L1, and a precise (P) code on both L1 and L2. An additional Y-code may be used to encrypt the P-code (antispoofing or AS).

A GPS receiver generates a replica of these codes and correlates them with the received signals, from which a pseudorange to each visible spacecraft can be inferred. (Pseudorange is simply a range biased by the unknown offset between the spacecraft and receiver clocks.) In the simplest (and lowest accuracy) form of GPS positioning, the receiver uses these pseudorange measurements together with ephemeris and clock information broadcast by the respective GPS spacecraft to determine its location. A minimum of four satellites must be in view of the receiver in order for the user to solve for the three components of position and the clock offset. The accuracy with which the user can determine its position is dependent on a number of factors; principal among them is the geometric configuration of the satellites in view. The quality of the broadcast ephemeris and clock information, which can be intentionally degraded as part of selective availability (SA), is also an important factor.

The same principles can be applied to the positioning of low Earth orbiters equipped with GPS receivers. Because the applications in this area are primarily in the field of precise geodesy, a more robust approach is generally required. In particular, multidirectional pseudorange and carrier phase measurements collected simultaneously at ground stations and the user spacecraft can be combined over suitable intervals of time—typically a few hours to several days—in order to determine the ephemerides of the orbiter [9–11]. The simultaneous measurements from the ground stations can be combined to nearly eliminate the effects of clock-error SA degradation, while also mitigating the effects of errors in the GPS ephemerides.

What makes this approach especially attractive is that the robust observation geometry permits orbit solutions without dynamic model constraints on the spacecraft motion [12]. (Errors in dynamic models are the principal limitations in traditional approaches to satellite orbit determination.) Where advantageous, however, dynamic models can still be exploited to improve accuracy [13]. Plans for a number of U.S. and international missions include flight-hardened, high-performance GPS receivers. Two such missions, the joint U.S.–French TOPEX/POSEIDON satellite and NASA's Extreme Ultraviolet Explorer, were launched in mid-1992 and have performed superbly. The accuracy of the TOPEX/POSEIDON orbits determined at JPL from GPS tracking is believed to be at the 3-cm-rms level in altitude and 10 cm rms or better in along- and cross-track components.[1]

While the application of GPS for the positioning of low Earth orbiters has received considerable attention, this is

[1] T. P. Yunck, W. I. Bertiger, S.-C. Wu, Y. Bar-Sever, E. J. Christensen, B. J. Haines, S. M. Lichten, R. J. Muellerschoen, Y. Vigue, and P. Willis, "First Assessment of GPS-Based Reduced Dynamic Orbit Determination on TOPEX/POSEIDON," submitted to *Geophys. Res. Ltr.*, August 1993.

not the case for high Earth orbiters, particularly geosynchronous spacecraft such as TDRS. The GPS constellation illuminates the Earth from an altitude of 20,200 km and, therefore, is better suited for low Earth orbiters. Since the TDRSs are located above the GPS constellation, they must look down to receive GPS signals spilled over the limb of the Earth from satellites on the other side of the planet. The configuration, hereinafter referred to as "down-looking GPS" in keeping with Wu [7], is shown in Fig. 1.

Although an observer traveling with TDRS would be able to establish a direct line of sight to many GPS satellites, the number of useful GPS spacecraft is limited to those that fall within an annular region delineated on the inside by the Earth's blockage and on the outside by the beamwidth of the GPS signals. The half-width of the main beams are 22 and 27 deg, respectively, at L1 and L2 frequencies, while the angle subtended by the Earth at GPS altitude is 27 deg. Together these constraints imply that, on average, the signals from only one GPS satellite can be seen from a geosynchronous altitude at any given time [7]. Of course this entirely precludes the possibility of kinematic positioning, and the orbits must be determined dynamically. For a spacecraft at geosynchronous altitude, however, the perturbative accelerations due to the nonspherical Earth are highly attenuated and the effects of atmospheric drag are negligible. As a result, the proper modeling of the forces acting on a spacecraft is much less problematic than it is for a low Earth orbiter.

Aside from these special limitations, the overall strategy for down-looking GPS is not unlike that for the up-looking variation used by low Earth orbiters. In particular, the determination of the orbit can be made using simultaneous observations formed with data collected at ground stations or directly, without the aid of a ground network. The benefit gained from the use of simultaneous observations, however, is somewhat limited owing to visibility constraints. Simultaneous observations of the same two GPS spacecraft from geosynchronous orbit and the ground are possible less than half the time even with the most optimistic scenarios [3,7]. Implicit in both approaches, therefore, is a greater vulnerability to clock errors and to the effects of SA if the flight receiver is not equipped with a decryption module. Despite these problems, the down-looking GPS approach is quite attractive for TDRS orbit determination because of the high level of autonomy and the greater potential for achieving real-time results.

An alternative strategy requires that the high Earth orbiter transmit a suitable signal which can be monitored at the same ground stations observing GPS satellites [7,14]. This method has been referred to as "inverted GPS" because the major factor affecting the orbit accuracy is the number of ground stations, rather than GPS satellites, in common view of the user spacecraft (Fig. 2). Inverted GPS, also referred to as GPS-like tracking (GLT), promises the highest accuracies for geosynchronous tracking because any number of ground sites may be visible from the TDRSs [7]. Coincident observations of the GPS satellites from the ground are desired in order to enable estimation of clock biases. As is the case for down-looking GPS, dynamic models of the forces governing the orbital motion are used to supplement the geometric content of the measurements.

III. Common Stategy

The assumptions forming the foundation of this study are governed by guidelines that have been advanced by NASA for future TDRS-II orbit determination, e.g., [3,4]. These guidelines reflect a balance between the demands for increased accuracy and system autonomy. For this effort, the figure of merit for the accuracy is 50 m in total position (1σ). It was assumed that this level of accuracy should be met in nominal operations with 24 hr of tracking, although we also examined the feasibility of achieving 50 m after only 2 hr of tracking (for the cases where the trajectory is to be recovered rapidly after a station-keeping maneuver). For system autonomy, the primary drivers include minimized impact on TDRSS user services, minimized human intervention during normal operations, and, for the inverted technique, a simple ground network. We began with the premise that the inverted technique, GLT, would provide the best accuracy, and focused on identifying compromises that would ensure greater autonomy. Conversely, for the down-looking approach, effort was devoted to determining ways to improve the accuracy.

The Orbit Analysis and Simulation Software (OASIS) package developed at JPL served as the primary evaluation tool. The OASIS system is designed to provide a flexible, versatile, and efficient covariance analysis tool for Earth-satellite navigation and GPS-based geodetic studies [15]. It has been used extensively for spacecraft orbit error analysis, and its factorized Kalman filter strategies [16] also form the basis for the GPS Inferred Positioning System (GIPSY) software used in the reduction of actual GPS data for recovering geodetic baselines and improving satellite orbits.

For both strategies, a full 24-satellite GPS constellation was assumed. The TDRSS-II satellites were assumed to be at the same locations as the present TDRS-W and

TDRS-E. The actual TDRSS-II constellation will contain additional satellites, but they should be clustered in the same vicinities as the current spacecraft. The results, therefore, should not be significantly different for these additional satellites. The next sections detail specific error models applied in the two solution strategies, along with the results. Covariance analysis results portray the actual expected errors only to the extent that the a priori models are authentic. In order to address the possibility of unanticipated errors, a somewhat conservative set of a priori assumptions was adopted.

IV. Inverted GPS (GPS-Like Tracking)

A. Assumptions

As a starting point, we propose some small ground networks suitable for the simultaneous tracking of GPS and TDRSS spacecraft. An initial stated goal for TDRS-II orbit determination was to confine all stations to the continental U.S. [3]. This constraint was subsequently relaxed [4]; it nonetheless remains essential to identify a minimum network that will deliver the desired orbit accuracy. For this effort, we selected various station configurations from the six-site global GPS network that has been established to support the TOPEX/POSEIDON mission. Three of the six sites are colocated with DSN stations at Goldstone, California; Madrid, Spain; and Canberra, Australia. The remaining three are at Santiago, Chile; Usuda, Japan; and Hartebeesthoek, South Africa. An additional receiver at the TDRSS ground control station at White Sands was assumed for some of the variations. The visibility of these sites from TDRS-E and TDRS-W, respectively, is shown in Fig. 3.

It is instructive to note that these sites are presently used to support well-established NASA programs. Each is equipped with a JPL Rogue digital receiver capable of simultaneously tracking pseudorange and carrier phase from eight GPS spacecraft [17]. Although the receivers are designed to operate unattended, staff are always on call at these sites should any problems develop. For this study, we assumed that the Rogue receivers at each of the tracking sites were retrofitted so that a TDRS beacon signal could be tracked continuously on one of the eight channels (Fig. 4). Note that GPS receivers have already been used in demonstrations to track Pioneer Venus and Magellan at X- (8000–12,500 MHz) and S-bands (2000–4000 MHz).[2]

A critical design parameter for the inverted GPS technique is the measurement characteristic of the TDRS beacon signal. Several options for the design of an advanced beacon signal have been considered [3,4]. For the present study, ranging tones broadcast by the TDRSs at Ku-band (12.5–18 GHz) served as the nominal configuration for the transmission. A major advantage of exploiting the high-frequency Ku-band is the relatively small signal delay due to ionospheric refraction. Equivalent range delays at Ku-band vary from less than 1 to 20 cm depending on the level of solar activity. Ionospheric calibration based on the GPS dual-frequency L-band data collected at the various tracking sites can then be applied in modeling the delay to better than 1 cm in range. A similar activity is already underway at the DSN sites, where the GPS data are used to calibrate ionospheric delays for deep-space tracking [18].

The proposed Ku-band signal could, in theory, provide pseudorange measurements with a random noise component of 1 cm averaged over 30 min, assuming a 100 MHz bandwidth.[3] In practice, the implementation of new Rogue hardware to downconvert the Ku-band signal to GPS frequencies (L-band) would introduce an additional error because separate signal paths would be used for the TDRS and GPS signals. This instrumental error would manifest itself as a slowly varying delay offset in the TDRS pseudorange residuals. Preliminary analysis indicates the effect would be bounded by about 1 nsec (amounting to 30 cm in range delay) and would modulate with a period of about one-half of a day. Because of the long period, the error appears as a constant bias over a typical measurement interval, permitting it to be modeled as a stochastic process in OASIS. Several variations from these nominal characteristics were explored in order to assess how deviations from these assumptions would impact the TDRS orbit accuracies. Results and additional details are presented in the next section.

The noise of the ionosphere-corrected GPS P-code pseudorange and carrier phase measurements was set at 25 and 1 cm, respectively, for 30-min measurement intervals. As Rogue receivers are presently providing this level of precision for 6-min measurement intervals (cf., Fig. 5), these estimates are quite conservative. The higher levels of data noise, however, are intended to accommodate periods when the receivers must track using codeless techniques because AS is turned on. Additional assumptions applied in OASIS for evaluating the inverted GPS technique are

[2] K. Zukor, "Comparison of NCB and TurboRogue Phase Measurements," JPL Interoffice Memorandum 335.1-91-016 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 15, 1991.

[3] L. E. Young, personal communication, Tracking Systems and Applications Section, Jet Propulsion Laboratory, Pasadena, California, 1992.

listed in Table 1. We assumed the a priori knowledge of the GPS ephemerides was very poor and solved for the 24-GPS and 2-TDRS epoch states together. Additional estimated parameters included a single solar-radiation pressure coefficient for each TDRS, and GPS solar-radiation pressure coefficients and carrier phase biases. Clock errors were estimated as stochastic white noise processes with a reference frequency standard at Goldstone, an approach which is analogous to (but more general than) using doubly differenced measurements. A random-walk process noise parameter was used to model the zenith troposphere delay at each of the stations [19].

The sensitivities of the TDRS orbit to errors in several important nonestimated parameters were also computed. These nonestimated, or "consider," parameters can be included in covariance studies in order to yield more realistic error estimates. The additional error contributions from the consider parameters are added to formal errors from the filter, which contain only the effects of data noise. The consider parameters and their associated errors (1σ) are also shown in Table 1. Note that these errors for consider parameters represent fixed systematic errors [16]. Most important among them are the tracking station coordinates and Earth orientation parameters. For individual components of the DSN station positions, errors of 3 cm were assumed. Recent analyses suggest that centimeter-level accuracies are already being achieved for the locations of GPS antennas at the two DSN sites in the Northern Hemisphere [20]. Coordinates for non-DSN sites were assigned conservative errors of 10 cm. Uncertainties in the X- and Y-pole positions were set at 25 cm, while the error in the variation of the Earth's rotation as manifest in UT1–UTC was set at 6.0×10^{-4} sec. In a unified GPS/TDRSS solution strategy at JPL, these Earth orientation parameters could be adjusted to reduce these errors by at least an order of magnitude. By using higher errors, we allow for a real-time system where accuracy may be degraded.

The lumped effects of errors in the Earth's gravity model were represented by 25 percent of the difference between the Goddard Earth Models (GEMs)-10 and -L2 [21,22]. Our own analysis suggests that for many applications this representation is comparable to the errors in the GEM-T3 gravity field [23], a state-of-the-art model developed in support of the TOPEX/POSEIDON mission. Owing to the extremely high altitude of a geosynchronous orbiter, the gravity model errors have only a minor effect on the TDRS orbit determination in comparison with other sources.

B. Results for Routine Orbit Determination

Consider first a nominal case which is characterized by the TDRS Ku-band beacon design outlined in the previous section and a minimal ground network consisting of the three DSN sites and White Sands. Figure 6 depicts the mapped orbit error (1σ) for TDRS-W as a function of time past the epoch of the arc. The formal (computed) error reflecting the effects of data noise is shown along with the systematic error from nonestimated (consider) parameters. The maximum rss total error was adopted as the basis for comparing various strategies in relation to the TDRSS requirement. For the 24-hr period in question, the total position error for TDRS-W never exceeds 15 m, well under the 50-m requirement.

In interpreting Fig. 6, it is instructive to note that the total error is dominated by the formal (computed) error contribution, indicating the results may be highly sensitive to the assumptions made for the TDRS beacon signal. To address this concern, limiting cases were examined. In the first, the TDRS signal was degraded by increasing the magnitude of the systematic contribution from 30 to 100 cm. Introducing this increase allows the partial accommodation of unmodeled ionospheric refraction errors, in addition to aggravated instrumental effects. For instance, if the TDRS beacon broadcast at S-band instead of Ku-band, the calibration of the ionospheric delay would yield accuracies of only a few decimeters. For the case of this degraded beacon, the maximum total error grew to 41 m (Fig. 7), a value which is still lower than the 50-m requirement.

In the second case, the systematic contribution was removed entirely, but the noise was increased by a factor of 25 (from 1 to 25 cm for 30-min averaging). Inasmuch as the GPS pseudorange signals were also assigned a data noise of 25 cm, this approach is analogous to the situation in which the TDRSs are equipped with actual GPS beacons. The maximum total rss error was 10 m, an improvement over the nominal case, showing that the 25-fold increase in the noise contribution was more than balanced by the elimination of the slowly varying bias (cf., Fig. 7). Taken together, these results indicate that the greatest concern for the TDRS beacon signal lies in the minimization of the systematic, slowly varying bias introduced by the different path lengths for the GPS and TDRS signals.

It is also instructive to investigate how the period of these systematic errors in the TDRS beacon signal affects the orbit determination. To answer this question, different values were assigned to the time constant for the 30-cm bias, and the formal position error for TDRS-W was computed at epoch. (Recall that the nominal $1/e$ folding time

constant, τ, was one-half of a day.) The results, depicted in Fig. 8, indicate that the worst accuracies are experienced when the period of the systematic error is about 5 hr. As the time constant of the systematic error decreases below 5 hr, the orbit error also decreases until the limiting case of white noise is reached. This phenomenon is evidently a consequence of increased decoupling with other parameter errors, even though a smaller τ represents higher process noise. Likewise, as the period approaches 1 day, the orbit error decreases as the systematic error appears more like a single constant bias over the entire 24-hr arc.

We now examine the effects of various tracking network configurations. While it is adequate for observing TDRS-W at 171 deg west, the minimum network consisting of stations at the three DSN sites and White Sands is not well-suited for tracking TDRS-E at 41 deg west. The situation is best illustrated in Fig. 3. TDRS-W is viewed by two DSN sites (Goldstone and Canberra) plus White Sands. Although the distance between the two American stations is rather short, the overall baseline orientation is adequate for providing the necessary geometric diversity in the observations. In contrast, TDRS-E is viewed by only Madrid and White Sands. (The elevation of TDRS-E above the horizon at Goldstone is about 2 deg, rendering any observations collected there unreliable.) The network consists of a single, long baseline which can provide TDRS-E orbit accuracies no better than 300 m. Even in a best-case scenario, in which it is assumed that useful observations can be made from Goldstone, the maximum orbit error for TDRS-E cannot be brought below the 50-m level without tuning of Earth orientation parameters. For tracking TDRS-E, it is, therefore, necessary to consider an augmented tracking network.

The simplest augmented network is a five-station configuration consisting of the three DSN sites, White Sands, and the TOPEX/POSEIDON site in Santiago, Chile. While the tracking geometry for TDRS-W remains identical to the nominal case, the situation for TDRS-E is dramatically improved. The introduction of the Santiago site implies that TDRS-E is observed by three well-distributed stations. Indeed, Table 2 reveals that with this five-station network the TDRS-E orbit can be determined to the sub-5-m level, a factor of three better than the TDRS-W orbit.

As a final case, consider the six-station TOPEX/POSEIDON network. This configuration supplies the most robust and consistent geometry for observing both spacecraft—TDRS-E is observed by Madrid, Hartebeesthoek, and Santiago, while TDRS-W is viewed by Canberra, Goldstone, and Usuda. It is noteworthy that no tracking from White Sands is involved, a scenario which is attractive because (1) among all the sites discussed, White Sands is the only location not presently part of the operational NASA GPS network, and (2) in many of the strategies, tracking of both TDRS-W and TDRS-E is required from White Sands, implying that the single TDRS channel in the reconfigured GPS receiver would have to be shared. Figure 9 shows the orbit accuracies for TDRS-W and TDRS-E throughout a 24-hr simulated arc with tracking from the full TOPEX/POSEIDON network. The accuracies achieved are better than 5-m for both spacecraft, an order of magnitude better than the 50-m requirement.

C. Results for Trajectory Recovery and Prediction

The TDRSS spacecraft are actively maneuvered as part of routine station-keeping activities. In order to keep disruption of user services to a minimum, it is desirable to recover the trajectory as quickly as possible after the thrust maneuvers. This section explores the capability of the inverted technique for determining the TDRS positions to better than 50 m within 2 hr of a thrust event. Two different approaches are adopted: In the first, a complete recovery of the TDRS epoch state immediately after the maneuver is performed. No a priori information on the TDRS trajectory is assumed. In contrast to the nominal approach outlined in the previous section, however, the GPS orbits are well determined from routine tracking for 12 hr prior to the maneuver. In the second approach, a three-component velocity increment at the maneuver time is used to augment the TDRS state vector; thus, the thrust maneuver is determined as part of the orbit determination process.

Figure 10 depicts the TDRS-W orbit accuracy as a function of time after the thrust event for these two approaches. Two different tracking configurations are also considered. For complete orbit state recovery with the nominal tracking network, the 50-m requirement is nearly met after 2 hr. Using the full TOPEX/POSEIDON network, sub-40-m accuracy can be achieved after only 2 hr of tracking. Assuming that the three-component velocity increments can adequately model the thrust event, and moreover that the time of the maneuver is known, the 50-m requirement can easily be met with minimal tracking.

Finally, consider how long the quality of the TDRSs' trajectories can be maintained after cessation of tracking. To examine this, the TDRS-W orbit state was predicted forward for 3 days following the end of the 24-hr definitive orbit determination interval. The results, shown in Fig. 11, suggest that the 50-m requirement would continue to be satisfied, even with a total loss of tracking for 3 days.

V. Down-Looking GPS

A. Assumptions

For the down-looking GPS tracking option, we elected not to introduce any NASA tracking from the ground. The enhancement in accuracy that might be achieved with only a very limited number of differential observations is outweighed by the benefit of the increased autonomy associated with no ground sites. The estimation strategy for nondifferential down-looking GPS is quite different from that for the inverted option, owing in large part to the weak observability. Many of the parameters, such as the solar radiation pressure coefficient and the GPS orbit states, cannot be recovered reliably with the limited set of observations. Moreover, tracking in the nondifferential mode implies that the GPS measurements are sensitive to the effects of the intentional dithering of the GPS clocks and ephemerides (SA). For the nominal case, then, it was assumed that the onboard flight receiver would be a military-class instrument with a decryption module. Note that the introduction of this type of flight instrument on TDRSS-II spacecraft should not pose a problem since considerable military data are already processed through TDRSS. It is additionally assumed that the receiver would represent an advanced design capable of 35-cm pseudorange measurements with averaging over 15 min.

Table 3 lists the nominal set of a priori assumptions for the down-looking approach. The TDRSs' epoch positions and clock errors (bias and linear drift) served as the only estimated parameters. Solar radiation pressure was considered at 5 percent, a value which is conservative in comparison with the 2 percent value that is representative of current modeling efforts [24]. GPS satellite epoch states and clock errors were also considered. For the nominal case, in which it was assumed that the flight receiver was equipped with a decryption module, the GPS ephemeris and clock errors were set at the few-meters level. For the degraded case, these values were increased by a factor of four or more to account for the effects of SA [25].

B. Results for Routine Orbit Determination

Figure 12 shows the position error for TDRS-W as a function of time for nominal 24-hr tracking. Because the down-looking technique considered herein does not rely on ground tracking, the overall results are invariant to the position of the satellite and should not be much different for TDRS-E. The results suggest that with the decryption module on the TDRSs, the down-looking technique yields orbit accuracies at the sub-10-m level. In contrast, without the module, the position error reaches 80 m, and the 50-m requirement is not met. Longer data spans are not expected to provide appreciably higher accuracies; after 24 hr, the TDRS position errors approach the limiting values governed by the GPS ephemeris and clock errors.

Because the TDRS orbit errors for the down-looking approach are dominated by errors in unadjusted parameters, it is instructive to examine a simple error budget. Figure 13 shows the breakdown of the TDRS-W orbit error for the 24-hr arc. For the nominal case (with decryption), the limiting error sources are the GPS clocks and ephemerides. The data noise contribution from the filter estimation is negligible, owing to the high quality of the pseudorange measurements. For the case in which the receiver is not equipped to handle SA degradation, the GPS errors increase several-fold. In addition, the data noise contribution from the filter estimation becomes quite significant. This increase reflects the dithering of the GPS clocks, which can introduce apparent range errors as high as 60 m into the pseudorange observables [25].

C. Results for Trajectory Recovery and Prediction

The figure of merit for evaluating the trajectory recovery capability of down-looking GPS is simply the shortest interval of tracking that can provide a sub-50-m position error for TDRS. In this context, rapid recovery of the trajectory after station keeping can be achieved only if the flight receiver is equipped with a decryption module. Without the module, the TDRS position error after 2 hr of tracking is in excess of 4 km; approaching the 50-m requirement requires at least 24 hr of tracking. With the module, the 50-m requirement can be met with tracking as short as 4 hr (Fig. 14).

The nature of the predicted orbit error for TDRS-W was not explicitly examined for the down-looking case. Note that predicted orbit error is a function of (1) the error in the satellite state at the beginning of the predictive interval (also called the initial condition error) and (2) the errors in the dynamic models used to integrate the satellite position. To the extent that the initial condition errors for the down-looking and inverted approaches are roughly equivalent in magnitude, the predictive errors should also be similar. In this context, we conclude that the 50-m requirement cannot be met during the predictive interval unless the flight receiver is equipped with a decryption module. Without the module, the errors in the initial conditions estimated with 24-hr of tracking prior to the predictive interval would exceed the 50-m threshold. With the module, a sub-15-m initial condition error is achieved after 24 hr of tracking, and the pattern of the predicted error would likely be similar to that shown in Fig. 11.

VI. Conclusions

Two GPS-based strategies for tracking the geosynchronous TDRSs have been explored. Direct tracking of the TDRSs from the GPS constellation promises the greatest autonomy since no ground network is required. For this strategy, the primary impairment is the poor geometry—the TDRSs must look down to find signals broadcast from GPS satellites on the other side of the Earth. The situation is exacerbated by sensitivity of the TDRS orbit accuracy to SA, because measurements from the ground cannot be exploited to form differential observations which are free from these effects. In order to circumvent this difficulty, the TDRS-IIs can carry military-qualified GPS flight receivers which are designed to decrypt the degraded signals. The results suggest that, equipped in this manner, a GPS receiver should be able to provide the TDRS positions autonomously to better than 15 m for routine 24-hr tracking. Implicit in this result is the assumption that nominal Department of Defense operations are maintained. Moreover, if this technique is adopted, the effects of the long GPS-to-TDRS transmission paths and near-Earth grazing need to be further examined.

An alternative approach relies on simultaneous tracking of TDRS and GPS beacon signals from the ground (GPS-like tracking). If accuracy is the prime concern, then this inverted technique is the best suited for tracking geosynchronous orbiters. However, the introduction of a ground network makes it less autonomous than its down-looking counterpart. This study relied on a small number of current NASA GPS tracking sites and assumed that the receivers operating at those sites would be retrofitted to track TDRSS-II spacecraft on one of the eight channels that are normally reserved for GPS. Moreover, it was assumed that the TDRSS-II spacecraft would be configured to broadcast continuously a suitable wide-beam beacon signal, preferably at Ku-band to mitigate the effects of ionospheric refraction. The results suggest that data collected at the ground sites introduce a robust differential observation geometry that promises to deliver few-meter accuracies for TDRS with as few as six global stations. Smaller networks could still meet the 50-m TDRSS accuracy requirement, but each satellite must be observed by a minimum of three stations that are moderately well distributed.

The TDRSS-II orbit determination activities could be incorporated into routine GPS data processing that is currently done at JPL to support ongoing NASA programs. The mechanisms for near-real-time operations are already in place, as the GPS data from these remote sites are transmitted to JPL on a daily basis for automated processing.

Acknowledgment

The authors would like to thank Larry Young for providing the fundamental design and signal characteristics of the proposed GPS/TDRSS Rogue satellite receiver.

References

[1] W. J. Holmes, Jr., "NASA's Tracking and Data Relay Satellite System," *IEEE Communications Magazine*, vol. 16, no. 5, pp. 13–20, September 1978.

[2] J. Teles, Y. Nakai, and M. V. Samii, "TDRS Tracking and Orbit Determination," paper presented at CNES Intl. Conference on Space Dynamics for Geostationary Satellites, Toulouse, France, October 28–30, 1985.

[3] A. Long, S. Liu, R. Potash, L. Floyd, A. Jacobsen, K. Cunningham, A. Kapoor, C. Kwadrat, J. Radel, and J. McCarthy, *Advanced Tracking Systems Design and Analysis*, Contractor Report CSC/TM-88/6060, Computer Sciences Corporation, Silver Spring, Maryland, September 1989.

[4] A. Weinberg, D. D. Elrod, D. F. Eggerts, and A. Jacobsen, *Assessment of Candidate ATDRS Tracking Techniques*, Contractor Report TR91002, Stanford Telecommunications, San Jose, California, May 1991.

[5] J. Ray, C. Knight, N. Zelensky, P. Liebrecht, and A. Jones, "VLBI Tracking of the TDRS," *J. Astronaut. Sci.*, vol. 36, no. 4, pp. 347–364, October–December 1988.

[6] S. Nandi, C. Edwards, and S. C. Wu, "TDRSS Orbit Determination Using Short-Baseline Differenced Carrier Phase," paper presented at NASA Goddard Flight Mechanics Estimation Theory Symposium, Greenbelt, Maryland, May 1992.

[7] S. C. Wu, "Differential GPS Approaches to Orbit Determination of High-Altitude Satellites," AAS Paper 85-430, presented at the Astrodynamics Specialists Conference, Vail, Colorado, August 1985.

[8] B. W. Parkinson, "Overview," *Global Positioning System, Papers Published in Navigation*, edited by P. M. Janiczek, Washington, DC: Institute of Navigation, pp. 1–2, 1980.

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

[10] T. P. Yunck, S. C. Wu, S. M. Lichten, W. I. Bertiger, U. J. Linqwister, and G. Blewitt, "Toward Centimeter Orbit Determination and Millimeter Geodesy With GPS," *Proceedings of the 5th International Geodetic Symposium on Satellite Positioning*, vol. 1, Las Cruces, New Mexico, pp. 272–281, March 13–17, 1989.

[11] 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. Geoscience Remote Sensing*, vol. GE-28, no. 1, pp. 108–116, January 1990.

[12] T. P. Yunck and S. C. Wu, "Non-Dynamic Decimeter Tracking of Earth Satellites Using the Global Positioning System," AIAA Paper 86-0404, presented at the 24th Aerospace Sciences Meeting, Reno, Nevada, January 6–9, 1986.

[13] S. C. Wu, T. P. Yunck, and C. L. Thornton, "Reduced Dynamic Technique for Precise Orbit Determination of Low-Earth Satellites," *J. Guidance Contr. and Dyn.*, vol. 14, no. 1, pp. 24–30, January–February 1991.

[14] T. P. Yunck and S. C. Wu, "Ultra-Precise Orbit Determination by GPS," AAS Paper 83-315, presented at the Astrodynamics Specialists Conference, Lake Placid, New York, August 22–25, 1983.

[15] S. C. Wu and C. L. Thornton, "OASIS—A New GPS Covariance and Simulation Analysis Software System," paper presented at the First Int. Symposium on Precise Positioning with GPS, IUGG, Rockville, Maryland, April 15–19, 1985.

[16] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, Orlando, Florida: Academic Press, 1977.

[17] J. B. Thomas, *Functional Description of Signal Processing in the Rogue GPS Receiver*, JPL Publication 88-15, Jet Propulsion Laboratory, Pasadena, California, June 1, 1988.

[18] J. M. Srinivasan, T. K. Meehan, and L. E. Young, "Code and Codeless Ionospheric Measurements With NASA's Rogue GPS Receiver," *Proceedings of the Institute of Navigation GPS-89 Conference*, Colorado Springs, Colorado, pp. 451–453, September 27–29, 1989.

[19] S. M. Lichten and J. S. Border, "Strategies for High-Precision Global Positioning System Orbit Determination," *J. Geophys. Res.*, vol. 92, no. B12, pp. 12751–12762, November 1987.

[20] G. Blewitt, M. B. Heflin, F. H. Webb, U. L. Linqwister, and R. P. Malla, "Global Coordinates With Centimeter Accuracy in the International Terrestrial Reference Frame Using the Global Positioning System," *Geophys. Res. Let.*, vol. 19, no. 9, pp. 853–856, September 1992.

[21] F. J. Lerch, S. M. Klosko, R. E. Laubscher, and C. A. Wagner, "Gravity Model Improvement Using GEOS-3 (GEM-9 and 10)," *J. Geophys. Res.*, vol. 84, no. B8, pp. 3897–3915, August 1979.

[22] F. J. Lerch, S. M. Klosko, and B. G. Patel, "A Refined Gravity Model From Lageos (GEM-L2)," *Geophys. Res. Letters*, vol. 9, no. 11, pp. 1263–1266, November 1982.

[23] F. J. Lerch, R. S. Nerem, B. H. Putney, T. L. Felsentreger, B. V. Sanchez, S. M. Klosko, G. B. Patel, R. G. Williamson, P. S. Chinn, J. C. Chan, K. E. Rachlin, N. L. Chandler, J. J. McCarthy, J. A. Marshall, S. B. Luthcke, D. W. Pavlis, J. W. Robbins, S. Kapoor, and E. C. Pavlis, *Geopotential Models of the Earth from Satellite Tracking, Altimeter, and Surface Gravity Observations: GEM-T3 and GEM-T3S*, NASA Technical Memorandum 104555, Washington, DC, January 1992.

[24] M. Bobrowsky, P. Y. Kay, A. K. Drew, S. L. Hoge, and O. O. Cuevas, "Short- and Long-Term Determination of the TDRS Solar Reflectivity Parameter Using the Goddard Trajectory Determination System," AAS Paper 87-531, presented at the Astrodynamics Specialists Conference, Kalispell, Montana, August 10–13, 1987.

[25] E. S. Davis, W. G. Melbourne, and T. P. Yunck, "GPS Applications to Space-Based Remote Sensing Missions: Coping With Denial of Accuracy," *Proceedings of the 2nd International Symposium of Positioning With GPS*, Ottawa, Canada, pp. 25–33, September 1990.

Table 1. Error models for inverted GPS (GLT).

A-priori assumptions for estimated parameters	
Parameter	Value
TDRS position (X, Y, Z)	5 km
TDRS velocity (X, Y, Z)	50 m/sec
TDRS solar radiation pressure	5%
GPS position (X, Y, Z)	100 m
GPS velocity (X, Y, Z)	1 m/sec
GPS solar radiation pressure	25%
GPS Y-bias	10^{-12} m/sec^2
GPS carrier phase biases	1000 km
GPS/TDRS/station clocks	1000 μsec, white noise
Zenith troposphere	40 cm + 12 cm/day
Consider parameters	
Parameter	Value
DSN station coordinates	3 cm
Non-DSN station coordinates	10 cm
Earth's gravitational constant (GM)	2 ppb
Lumped Earth gravity field	25% of GEM-10 − GEM-L2
X-, Y-pole motion	25 cm
UT1–UTC	6×10^{-4} sec

Table 2. TDRSS orbit error for various tracking strategies.

Tracking network	TDRS-W maximum error	TDRS-E maximum error
3 DSN + White Sands	14	>300
3 DSN + White Sands + Santiago	14	4
3 DSN + 3 TOPEX/POSEIDON	4	4

Table 3. Error models for down-looking GPS.

A-priori assumptions for estimated parameters	
Parameter	Value
TDRS position (X, Y, Z)	10 km
TDRS velocity (X, Y, Z)	1 m/sec
TDRS clock bias	33 μsec
TDRS clock drift	3 nsec/sec

Consider parameters		
Parameter	Value	
	With decryption	Without decryption
TDRS solar radiation pressure	5%	5%
GPS position (rss total)	7 m	30 m
GPS clock error	6 nsec	60 nsec
Earth's gravitational constant (GM)	2 ppb	2 ppb

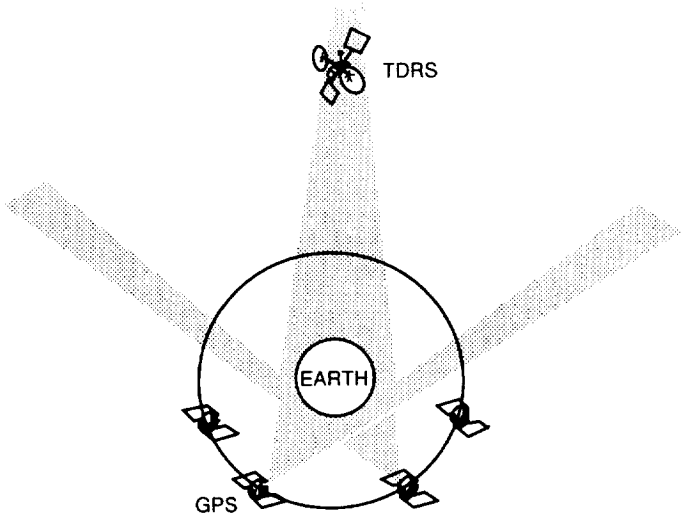


Fig. 1. Two-dimensional view of down-looking GPS tracking configuration: geosynchronous TDRS with a GPS receiver sees GPS signals spilled over the limb of the Earth.

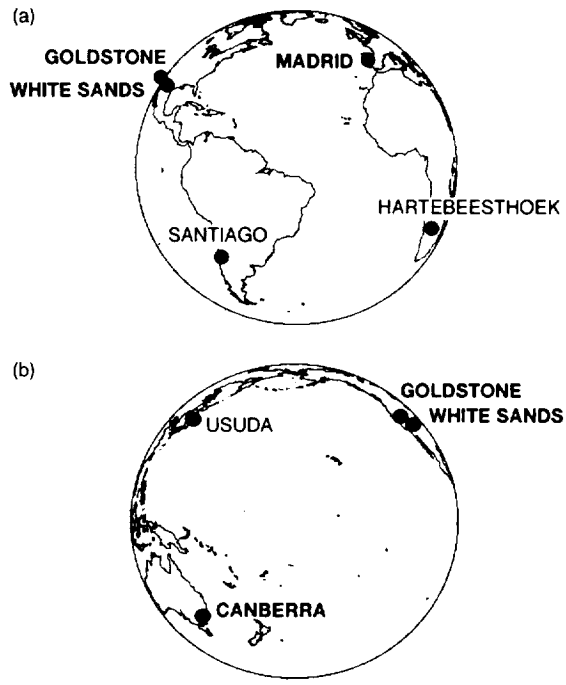


Fig. 3. Visibility of the proposed GPS ground network for TDRS tracking: perspectives of the Earth from geosynchronous orbit positions of (a) TDRS-E and (b) TDRS-W. The minimum network considered, consisting of stations at the DSN sites and the White Sands TDRSS ground control center, is shown in boldface type.

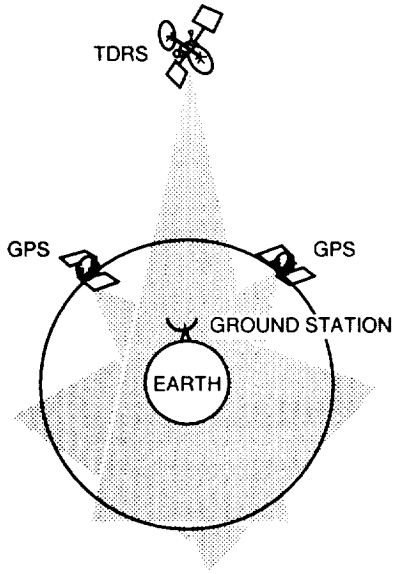


Fig. 2. Inverted GPS tracking configuration: TDRS and GPS beacon signals tracked simultaneously from the ground.

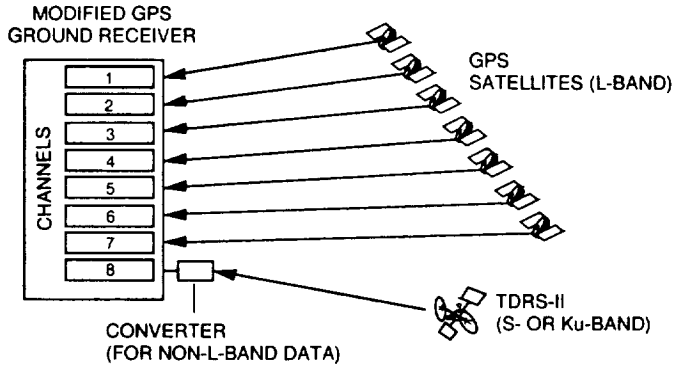


Fig. 4. The eight-channel GPS receiver modified for TDRS tracking on one channel.

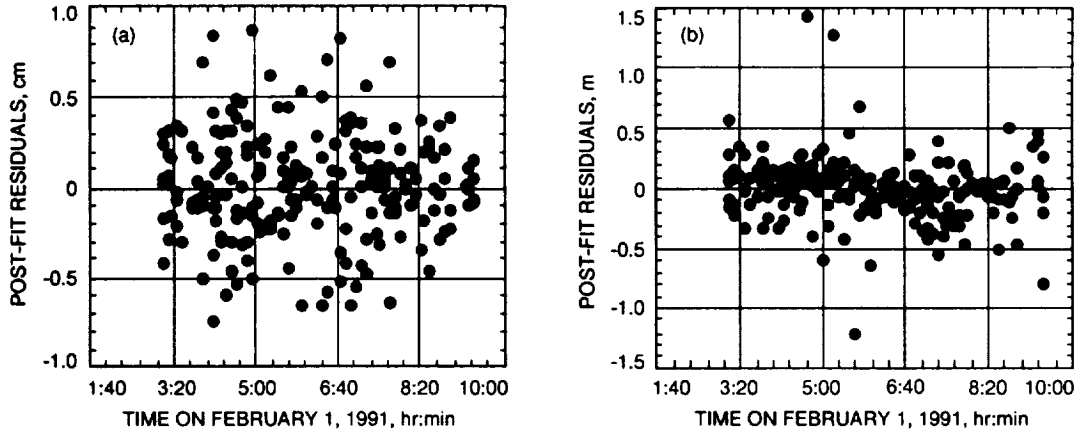


Fig. 5. Actual post-fit GPS data residuals from the Rogue receiver at Goldstone for (a) carrier phase rms = 2.9 mm and (b) pseudorange rms = 0.26 m. The GPS measurements are at a 6-min rate.

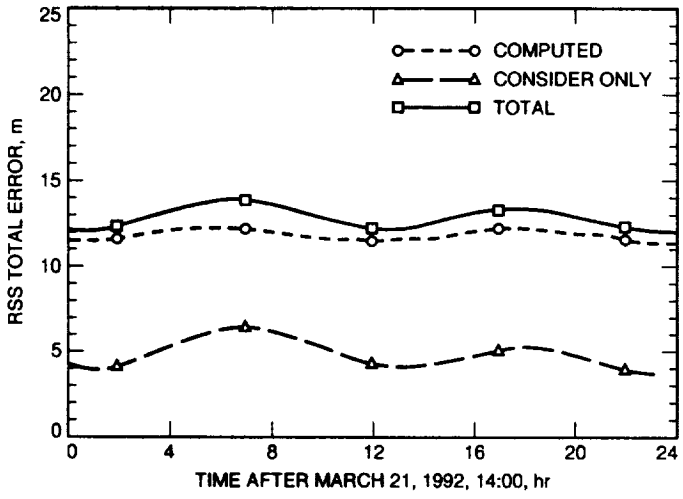


Fig. 6. TDRS-W position error for a 24-hr arc. The orbit is determined using the inverted technique with tracking from two DSN sites and White Sands. TDRS-W carries a nominal Ku-band beacon.

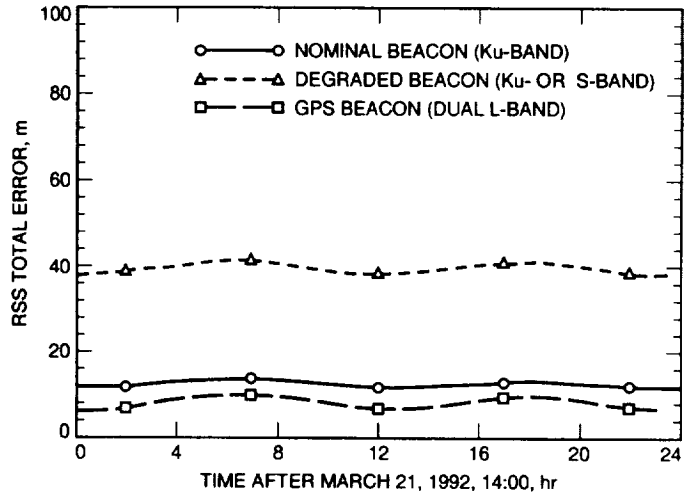


Fig. 7. TDRS-W position error over a 24-hr arc for hypothetical beacon signals. The orbit is determined using the inverted technique with tracking from two DSN sites and White Sands.

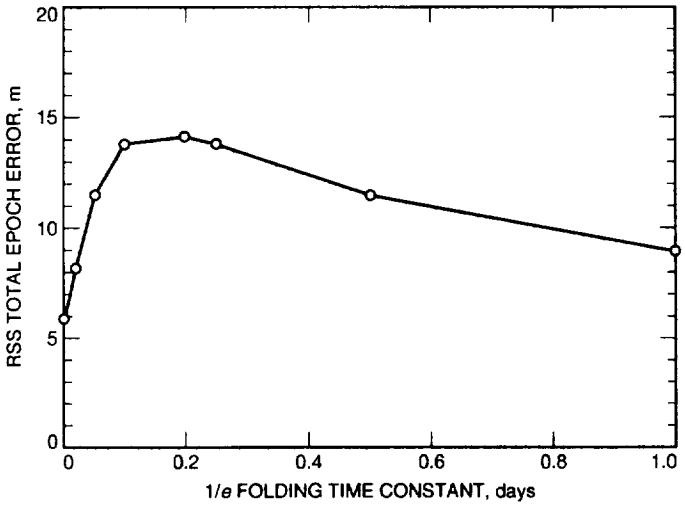


Fig. 8. TDRS-W position error for various systematic signal errors in the TDRS beacon. The time constant refers to the period of the systematic error. The error is bounded by 1 nsec (about 30 cm in range).

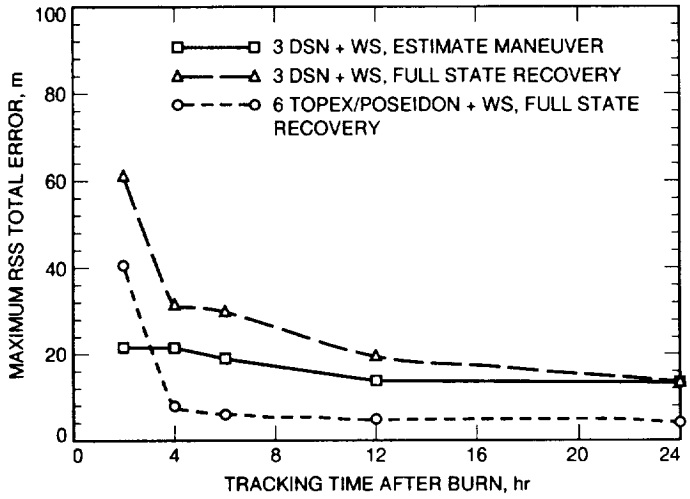


Fig. 10. TDRS-W position error after a station-keeping maneuver for the inverted technique. The orbit after the maneuver is determined using either full TDRS-W state recovery or via the solution of three-component velocity increments.

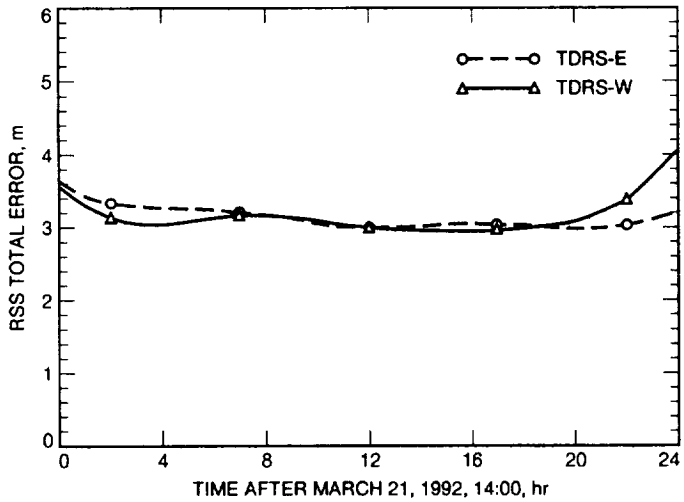


Fig. 9. TDRS-E and TDRS-W position errors for a 24-hr arc. The orbit is determined using the inverted technique with tracking from six TOPEX/POSEIDON sites. Both satellites carry nominal Ku-band beacons.

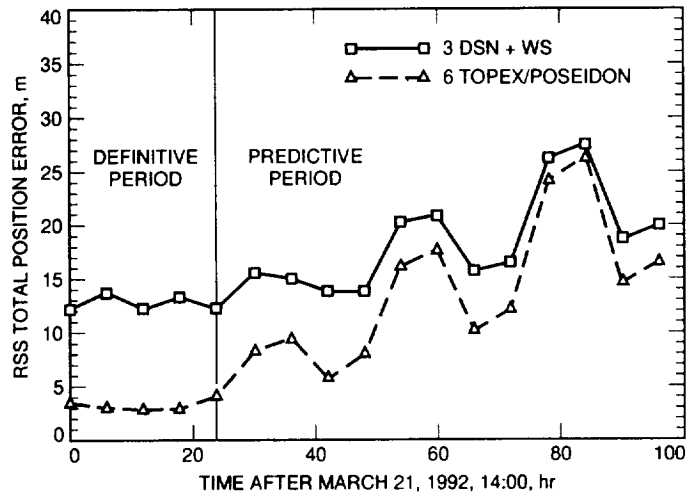


Fig. 11. TDRS-W position error as a function of time for definitive and predictive orbit determination. Actual observations are reduced only for the first day.

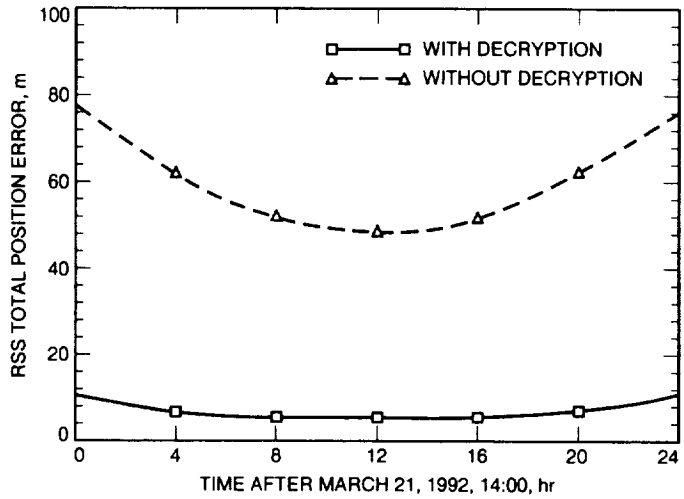


Fig. 12. TDRS-W position error for a 24-hr arc. The orbit is determined using the down-looking technique.

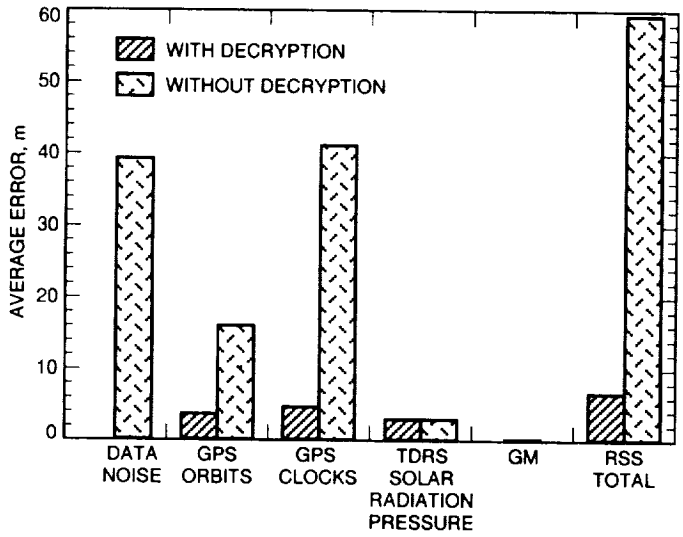


Fig. 13. Breakdown of the TDRS-W position error for a 24-hr arc. The orbit is determined using the down-looking technique.

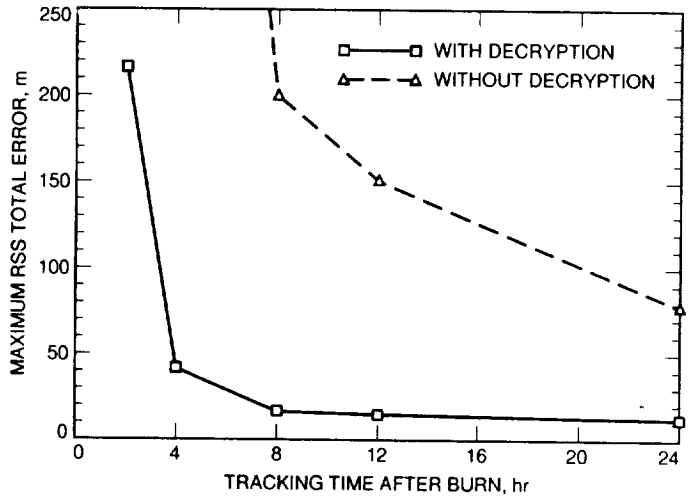


Fig. 14. TDRS-W maximum position error after a station-keeping maneuver for the down-looking technique. The orbit after the maneuver is determined using full TDRS-W orbit state recovery.

New Approaches for Tracking Earth Orbiters Using Modified GPS Ground Receivers

S. M. Lichten, L. E. Young, S. Nandi, B. J. Haines, and C. E. Dunn
Tracking Systems and Applications Section

C. D. Edwards
TDA Technology Development Section

A Global Positioning System (GPS) flight receiver provides a means to precisely determine orbits for satellites in low-to-moderate altitude orbits. Above a 5000-km altitude, however, relatively few GPS satellites are visible. New approaches to orbit determination for satellites at higher altitudes could reduce DSN antenna time needed to provide navigation and orbit determination support to future missions. Modification of GPS ground receivers enables a beacon from the orbiter to be tracked simultaneously with GPS data. The orbit accuracy expected from this GPS-like tracking (GLT) technique is expected to be in the range of a few meters or better for altitudes up to 100,000 km with a global ground network. For geosynchronous satellites, however, there are unique challenges due to geometrical limitations and to the lack of strong dynamical signature in tracking data. We examine two approaches for tracking the Tracking and Data Relay Satellite System (TDRSS) geostationary orbiters. One uses GLT with a global network; the other relies on a small "connected element" ground network with a distributed clock for short-baseline differential carrier phase (SB$\Delta\Phi$). We describe an experiment planned for late 1993, which will combine aspects of both GLT and SB$\Delta\Phi$, to demonstrate a new approach for tracking the Tracking and Data Relay Satellites (TDRSs) that offers a number of operationally convenient and attractive features. The TDRS demonstration will be in effect a proof-of-concept experiment for a new approach to tracking spacecraft which could be applied more generally to deep-space as well as near-Earth regimes.

I. Introduction

The Global Positioning System (GPS) has ushered in a new era for satellite tracking and precise orbit determination. The TOPEX/POSEIDON GPS precise orbit-determination demonstration has shown that sub-10-cm orbit accuracy is attainable for a low Earth orbiter equipped with a high-quality GPS flight receiver [1]. The key to high-accuracy positioning with a GPS flight receiver is the simultaneous common view of multiple GPS satellites as seen from the orbiting receiver and a network of ground receivers. As the altitude of the Earth orbiter car-

rying the GPS flight receiver increases, however, the number of GPS satellites in view decreases. Above a 5000-km altitude, typically less than four GPS satellites are simultaneously in view from the Earth orbiter [2]. This situation can be helped somewhat by designing the flight instrument so that GPS satellites can be tracked from the opposite side of the Earth—the "down-looking differential GPS" approach [2,3]. This introduces some additional complications, however, such as greater space loss and weaker GPS signals, higher atmospheric distortion from the near-grazing Earth geometry, contamination of the signal with radiation from the Earth, and complications regarding the placement of the GPS antenna on the satellite.

Orbit determination for Earth orbiters carrying GPS flight receivers above a 5000-km altitude has also been studied for highly elliptical orbits [4]. Decimeter-quality performance can in principle be achieved provided that perigee is below 10,000 km. However, for satellites at much higher altitudes, such as geosynchronous or higher, the GPS visibility for either the upward-looking TOPEX/POSEIDON-style flight instrument or a down-looking instrument is poor.

New techniques using GPS technology for tracking satellites in high Earth and elliptical orbits are being studied in the DSN Advanced Systems Program. One approach dispenses with the GPS flight receiver, employing instead a simple beacon aboard the satellite to transmit a signal which can be tracked along with the GPS signals by the ground GPS receivers. This GPS-like tracking (GLT) beacon approach (also sometimes called "inverted GPS" since the differential positioning of the user spacecraft is made relative to the GPS satellites rather than to the ground stations) has been studied in a preliminary way by Wu et al. [3] and Haines et al. [5]. Another approach would use short-baseline differential carrier phase (SB$\Delta\Phi$), which might be particularly convenient for tracking a geosynchronous orbiter [6]. Both GLT and SB$\Delta\Phi$ are operationally attractive because they are amenable to automated processing schemes and could provide spacecraft tracking and navigation without large DSN antenna time.

As described in [3] and [5], for GLT a small number of modified GPS ground receivers would be distributed in a global network for tracking a GPS-like beacon on the high Earth orbiter.[1] The beacon would, preferably, transmit a simple set of ranging tones, which would be detected in GPS ground receivers modified so that they simultaneously track the user satellite along with GPS satellites. The result would be a precise differential orbit-determination capability for the high Earth orbiter. In contrast, the SB$\Delta\Phi$ technique utilizes short (approximately 100-km) baselines around a central site, relying primarily on carrier phase observables (although some ranging information is required for geosynchronous tracking). As originally proposed by Nandi et al. [6], the SB$\Delta\Phi$ approach utilizes a single distributed clock for the local network of ground receivers. The distributed clock could be provided, for example, by fiber-optic links. Although the overall accuracy with the SB$\Delta\Phi$ technique is expected to be somewhat degraded by the geometrical limitations imposed by the short baselines (relative to the global network), there is considerable operational convenience—particularly for near-real-time operation.

In this article, we focus on the SB$\Delta\Phi$ and GLT approaches. Analyses will be presented for Earth orbiters at altitudes from 40,000–100,000 km, where we show that orbit accuracy of a few meters could be achieved. We describe a demonstration experiment with Tracking and Data Relay Satellite System (TDRSS) geosynchronous orbiters that will take place late in 1993. In the TDRSS experiment, the ground network will be restricted by the limited Tracking and Data Relay Satellite (TDRS) ground footprint, so we plan to use a blend of SB$\Delta\Phi$ and GLT to demonstrate a proof of concept for this unique application of GPS technology for tracking satellites at high altitude. For the TDRS tracking demonstration, the goal is to provide a near-real-time capability for tracking to an accuracy of several tens of meters.

II. Review of Differential GPS Tracking Fundamentals

The GPS constellation is designed so that typically 6 to 10 navigation satellites can be tracked above 5 deg elevation from any ground site. When complete, the constellation will include 24 operational satellites. The GPS satellites transmit carrier signals at 1.228 and 1.575 GHz (L-band), which are modulated by a pseudorandom noise code, the P-code (precision code), at 10.23 MHz. Two frequencies are provided so that ionospheric signal delays can be calibrated. A second code, the C/A (clear acquisition) code, is somewhat noisier than the P-code due to its lower frequency at 1.023 MHz and the lack of dual-band ionospheric correction (see Fig. 1).

The GPS codes include a navigation message with GPS clock and orbit information which can be utilized for real-time point positioning by users equipped with GPS receivers. GPS pseudorange to four satellites determines

[1] It is assumed that placing an actual GPS transmitter on a NASA satellite would be impractical for a number of reasons. On the other hand, a *GPS-like* (or GPS-compatible) beacon would not be difficult.

three position coordinates plus the user clock offset from GPS time. The term "pseudorange" is used since the range calculation is based on the difference between the transmit and receive times and will include any offset between the transmitter and receiver clocks. With the P-code, user positions can be determined in a point positioning mode in near-real time to about 10 m. In normal operation, the Department of Defense turns on selective availability (SA) for most GPS satellites. SA introduces what can appear to be a clock dither and alterations to the broadcast ephemeris. Certain authorized users will be equipped with keys to correct for these effects, but other users will see transmitter clock variations of the order of 30–50 m with the broadcast ephemeris degraded to a similar level of accuracy. Few scientific and nonmilitary GPS users will be equipped with the keys for SA. Simultaneous GPS tracking from multiple receivers can differentially eliminate GPS and receiver clock offsets, including the apparent clock dithering from SA. Because the broadcast ephemeris provides orbit information at the level of 5–10 m (which could degrade to 50–100 m with implementation of another component of SA), for the highest precision, the user must estimate and improve the GPS orbits *and* use differential GPS techniques, as shown in Fig. 2. While Fig. 2 shows how transmitter and receiver clock offsets can be differentially eliminated, there are distinct advantages to explicitly estimating clocks in a Kalman filter. One advantage is that the pseudorange measurements provide a natural filter constraint to improve the determination of the clock offsets and phase biases; another is that introduction of data correlations through explicit differencing is avoided.

The primary data type for highest precision in non-real-time GPS applications is the carrier phase, which can be tracked with sub-centimeter precision in modern GPS receivers. The carrier phase, continuously tracked over hours, provides a precise time history of biased range (also referred to as "range change") which can be used to obtain precise orbit solutions (Fig. 2). The GPS carrier phase is ambiguous by some integer multiple of carrier wavelengths, but the carrier measures range change to sub-centimeter precision. Pseudorange data, even though typically 1–2 orders of magnitude noisier than phase data, are still useful, since they can be used in a Kalman-type filter to constrain the clocks and carrier phase integer ambiguities, which in general must be estimated. It is the simultaneous tracking from multiple receivers that allows for estimation of clock offsets to very high precision. After 12 hr of tracking, sub-meter GPS orbit accuracy can be achieved [7] and, more recently, with data arcs of 30 hr, GPS orbits are determined daily at JPL with a precision of 30–50 cm [8]. Figure 3 shows how GPS orbit quality has been assessed recently, with slightly overlapping 30-hr arcs.

Antispoofing (AS) may be routinely activated once the GPS constellation is fully operational. AS encrypts P-code, which can then be observed only in receivers with decryption modules; ordinary receivers can receive only the noisier C/A-L1 pseudorange measurements and L1 carrier phase. This would ordinarily severely limit positioning accuracy with GPS for users without authorization to use security modules in ground receivers. However, a number of receivers can produce "codeless" observables, resulting in high-quality carrier phase data with ionospheric correction. Some receivers can produce moderately precise codeless pseudorange as well. The codeless techniques include cross-correlation and squaring algorithms that enable precise observables, corrected for the ionosphere without requiring classified information, to be obtained. The results in Fig. 3 were obtained with receivers which have a codeless capability; however, since AS was actually off during those tracking sessions, the results were determined with P-code observables.

III. Differential GPS Tracking With an Orbiting Receiver

Figure 4 shows the geometry for differential GPS tracking with an orbiting flight receiver. For low Earth orbiters at altitudes of a few thousand kilometers or less, good visibility of GPS satellites is provided with an upward-looking flight instrument. An example is the TOPEX/POSEIDON satellite, for which sub-10-cm orbit accuracy has been demonstrated [1]. However, as the altitude increases above 3000 km, GPS visibility falls off dramatically (Fig. 5). Down-looking GPS, where the flight instrument looks down past the Earth's limb to track GPS satellites from the opposite side, provides more GPS measurements for altitudes between a few thousand and 10,000 km. However, above 10,000 km, there are typically less than two GPS satellites in view, making high-precision GPS differential positioning more difficult.

One possibility for a down-looking differential GPS application would be to equip the TDRSS geosynchronous satellites with GPS flight receivers to provide orbit determination for TDRSs. This scenario was studied through a series of covariance analyses [3,5] as part of an assessment for alternative tracking techniques for TDRSs. Those studies, utilizing a square-root information Kalman-type filter, showed that a geosynchronous TDRS carrying a hypothetical GPS receiver could in principle routinely achieve sub-40-m orbit accuracy less than 4 hr after a maneuver only if the flight instrument carried decryption for SA, which would require uploads of classified information. Without removal of SA, 4 hr after a maneuver the

predicted orbit errors were much higher—several hundred meters, reducing to 60 m after 24 hr. While such performance might be acceptable to some geosynchronous users, the goal of the studies was to identify for TDRSS new technologies which could operationally and conveniently support a 50-m requirement within a few hours after a maneuver. Thus, SA correction would be required, and this was considered to be inconvenient from an operational point of view.

An alternative approach for satellites above 5000 km is to use a GPS-like beacon on board for tracking from a GPS ground network. The advantage of this GLT option is illustrated by the counts shown in Fig. 5, where 10 ground stations are assumed to be evenly distributed around the globe. The key to exploiting the precision inherent in GPS tracking in this case is to use the same GPS ground receivers for tracking the user satellite simultaneously with multiple GPS satellites. The tracking performance might be comparable to that expected from having a hypothetical GPS satellite moved from the usual 20,200-km-altitude orbit up to geosynchronous altitude at 37,000 km, Fig. 6(a). This option is discussed in the following section.

IV. GPS-Like Tracking for TDRS

What performance would one expect from a hypothetical GPS satellite at geosynchronous altitude, such as in Fig. 6(a)? As a starting point, we can look at present-day GPS tracking accuracy (Fig. 3), routinely obtained at JPL with the GPS Inferred Positioning System (GIPSY)-Orbit Analysis and Simulation Software (OASIS), a prototype for a future operational system. The GPS tracking data are processed with a turnaround of 3–10 days, but this is limited primarily by data transmittal delays from some of the approximately 30 globally distributed tracking sites. Since some of the sites routinely transmit data several minutes after a session, in principle turnaround time could be reduced to less than 1 day. The daily precise GPS orbits are produced from the same square-root filtering software used in the covariance analyses for hypothetical tracking scenarios about to be described. In the covariance analyses, assumptions for estimating parameters in the filter and for data quality were matched as closely as possible to those presently used for real GPS data analysis. If one were simply to raise the altitude of one GPS satellite from 20,200 to 37,000 km, we calculate that the orbit accuracy would degrade from 30–50 cm (presently achieved with real GPS data) to about 3 m [3,5]. The degradation results primarily from the very limited dynamical information when tracking a geostationary orbiter (GPS satellites move in 12-hr-period orbits).

Additional analysis was carried out to determine the minimum number of ground stations needed to track the geostationary "GPS" satellite. At this point, the analysis was modified to incorporate a GPS-like beacon (instead of a true GPS beacon). For a GPS-like beacon, we assumed a series of tones would be broadcast at Ku-band (12.5–18 GHz) with a 100-MHz bandwidth. These tones would be spaced so that an equivalent one-way range data type would be produced with 5-cm data noise over a 1-min averaging interval. As shown in Fig. 6(b), GPS ground receivers would be modified so that one out of the eight channels normally used to track GPS satellites would be equipped with a separate front end and a small, inexpensive antenna, enabling simultaneous tracking of seven GPS satellites and the one user satellite (TDRS in this case). As will be described later, a demonstration version of such a ground receiver is presently being built at JPL. There are many advantages of using such a modified GPS receiver. The simultaneous GPS tracking enables centimeter-level calibrations to be produced for receiver coordinates, troposphere and ionosphere delays, Earth orientation variations, and clock synchronization. The disadvantage is that since the ground receiver utilizes a different front end for GPS and the user satellite, a bias can be expected. Upon further investigation, we have found that this bias can be expected to wander slowly over about a day with a range of ± 1 nsec (± 30 cm). This additional noise source is, in fact, a significant error source for the final result. With such a GPS-like beacon, a minimal network consisting of three NASA Deep Space Network (DSN) sites plus White Sands, New Mexico (WS), can support TDRS orbit accuracy of 10–15 m [5].[2] With six ground sites plus White Sands, the performance improves to 3–5 m (Fig. 7), and either network can support 50-m orbit determination within 2 hr after a maneuver (Fig. 8). The assumptions of the analysis are described in [5]. In all cases, a global network of GPS ground receivers provides the data strength for estimation of GPS orbits (Table 1). The filtering setup is nearly identical to that used presently for routine daily GPS orbit determination. This ensures that a realistic scenario is adopted for the error analysis. However, for this analysis, between four and seven GPS sites equipped with modified GPS receivers are capable of tracking both TDRS and GPS satellites (Fig. 8). In addition, the assumed errors for Earth orientation (UT1 and polar motion) were purposely made large to reflect knowledge of predicted values in real time.

[2] The analysis here assumes a global beacon signal from TDRS. As discussed below, continuously available TDRS phase and range data are available in a relatively small footprint centered around the southwestern United States. Hence, this study is applicable to a future TDRS configuration with a global beacon, not to the present configuration.

V. Short-Baseline Differential Phase (SB$\Delta\Phi$) for TDRS Tracking

The GPS-like beacon studies described above assumed a global distribution of ground sites for tracking a TDRS Ku-band beacon. For present-day TDRS satellites, however, the available carrier and ranging signals are predetermined by the current flight hardware, and these signals are restricted to a relatively small footprint (less than 1000 km) around White Sands, New Mexico.[3] In addition to Ku-band carrier phase, two-way range and Doppler are available at White Sands. The two-way range data are affected by relatively large uncalibrated time-varying errors and are typically assigned data weights of 30 m. An analysis in [6] took into account the data quality for the present-day TDRS carrier phase and two-way range within the limited footprint, and also included realistic error assumptions for effects of troposphere delay variations, ionosphere delays, systematic errors from solar radiation pressure mismodeling, and data noise. The ground network was assumed to consist of three stations within the footprint that are separated by between 1 and 100 km with a distributed clock. The distributed clock provides a common frequency reference over a fiber-optic link, providing considerable common error cancellation [9]. This setup is similar to that of connected element interferometry (CEI) [9] except that, for this TDRS analysis, we assumed that the phase biases between stations and the integer cycle ambiguity for the carrier phase are *not* resolved.

The station-differenced carrier phase observables from this tracking network provide a precise measure of plane-of-sky TDRS position. The two-way range data are necessary to determine the third component which, for a TDRS tracked from White Sands, includes much of the longitude component. Figure 9 shows the results of the covariance analysis for short-baseline differential phase (SB$\Delta\Phi$) plus the two-way range at White Sands. Although the short baseline tracking does not generally do as well as tracking from a global network, as Fig. 9 shows, SB$\Delta\Phi$ could in principle deliver 20-m orbit accuracy in a near-real-time operational system, depending on the quality of the two-way range data. The present-day two-way range is routinely assigned a weight of 30 m. We are investigating the sources of error for these data and hope eventually to be able to improve the calibrations needed to produce higher quality range data. While the goal is to achieve 1-m ranging, the TDRS tracking goal of 50 m could be met even with 5-m-quality two-way range. If 1-m ranging were available, the limiting orbit errors would be from unmodeled differential troposphere delay variations.

[3] The bilateral ranging transponder system (BRTS) for TDRS includes two-way signals with a global network of transponder sites. These data are presently used for orbit determination. We did not include those sites in our orbit analysis because our goal was to develop an alternate tracking system in which all the tracking data can be collected locally and processed in an automated way. For this reason, we restricted our TDRS data to the current footprint near White Sands, New Mexico.

VI. A TDRS Demonstration for SB$\Delta\Phi$ Using GPS-Like Tracking

In this section, we describe a demonstration experiment which will take place in late 1993. The demonstration combines the principles of both GLT and SB$\Delta\Phi$. Three GPS TurboRogue ground receivers are presently being modified at JPL to enable simultaneous tracking of TDRS Ku-band carrier phase in one channel and GPS L-band phase and pseudorange in the other seven channels. Small (meter-sized) inexpensive antennas will be used for the TDRS tracking, while the GPS signals will come through the regular omni antennas. Since creating a new fiber-optic distributed clock over 100 km around White Sands would be impractical for a proof-of-concept demonstration, instead we plan to use the GPS receivers themselves to provide clock synchronization for the three ground receivers near White Sands. Previous experimental work by Dunn et al. [10] has shown that sub-nanosecond GPS clock synchronization can be provided over intercontinental distances with high-quality GPS receivers. Because of the close proximity of the GPS receivers for this experiment, we expect that the time transfer will be substantially better than 1 nsec due to common error cancellation. We plan to space the receivers evenly around White Sands at several-hundred-kilometer distances. As discussed above, the quality of the White Sands two-way range is critical to the final accuracy achieved for the TDRS orbit. With minor software modifications, the TDRS carrier phase and two-way range data will be processed simultaneously in the GIPSY–OASIS II processor presently used for precise GPS and TOPEX/POSEIDON orbit determination. The data from the three new TDRS/GPS tracking sites near White Sands will be combined with GPS data from about 15 other globally distributed sites, which will provide a precise reference frame and enable sub-meter determination of the GPS orbits.

The use of GPS ground receivers modified for tracking the TDRS carrier phase (plus GPS) within the footprint has a number of advantages in addition to the relatively low cost. These advantages include "built-in" high-accuracy clock synchronization and troposphere, ionosphere (from the dual-band GPS signals at 1.2 and 1.6 GHz), and station location calibrations. In our experiment, the GPS receivers will perform the clock synchro-

nization function in place of the fiber-optic link for a distributed clock in SB$\Delta\Phi$. Without the restrictions imposed by using a fiber-optic link, the baselines can be extended beyond the 100-km spacing, up to, essentially, the limit of the TDRS footprint (between 500 and 1000 km). The longer the baseline, the better the derived TDRS orbit due to the stronger geometry. Our goal for the TDRS/GPS demonstration is 50-m orbit accuracy, but the final result may be somewhat worse in the longitude component due to some uncertainty in the quality of the White Sands two-way range.

As an operational system, the small GPS/TDRS receiver network could provide turnaround of an hour or two if GPS predictions were used from a previous day for GPS orbits. A more accurate, fully combined solution with the GPS/TDRS data plus global GPS data would be available within a day or two. Reliable turnaround would be needed immediately (2 hr) after a TDRS maneuver to meet present-day operational requirements. The present-day prototype GPS analysis software at JPL is largely automated and could be the basis for such an operational system.

VII. GPS-Like Tracking for Satellites Above Geosynchronous Altitudes

Tracking for orbit determination and navigation support at altitudes above the geosynchronous has been provided by observatories such as the DSN. Although conventional tracking data types (such as two-way Doppler and range) which are routinely used at the DSN could in principle provide 10–20-m orbit accuracy [11], this requires significant amounts of antenna time from either the DSN or subnets of smaller (approximately 10-m-sized) antennas. The best performance is expected from data types which require simultaneous tracking from more than one ground antenna [11], are more costly, and require careful scheduling of observations. For future orbiting astrophysical observatories expected to do very-long-baseline interferometry (VLBI) radio mapping and astrometric experiments, precise velocity knowledge is required to enable proper modeling of the data. Several of these future missions require velocity accuracies of 0.5 cm/sec. This level of accuracy can in some cases be provided by the existing DSN, but it is unlikely that the DSN can provide it at all points in the orbit [11]. In general, however, the scheduling of deep space antennas is the overriding concern because of the scarcity of available antenna time.

The use of small ground antennas and GPS receivers for tracking high Earth/elliptical orbiters along with GPS satellites is attractive both because of the potential for freeing up significant amounts of expensive large antenna time at observatories and because of the high accuracy which is potentially achievable with the use of GPS technology. A recent analysis examined tracking for POINTS (Precision Optical Interferometer in Space), an optical astrometric mission for identification and characterization of planetary systems around other stars in the solar neighborhood.[4] This mission requires velocity determination to an accuracy of 0.5 mm/sec.[5,6] Clearly, conventional DSN tracking systems [11] cannot provide this level of accuracy (missing by 1–2 orders of magnitude). However, JPL's daily GPS orbit formal errors are at the level of 0.1 mm/sec. Even allowing for some accuracy degradation due to the higher altitude of POINTS (100,000 km) as compared with the GPS altitude (20,000 km), placing a GPS-compatible or GPS-like beacon on POINTS for orbit determination might be a viable approach.

The analysis assumes a nearly circular orbit at high altitude (100,000 km). A Ku-band beacon was assumed to be on the POINTS spacecraft with effective data noise of 5 cm over 1 min. A larger source of measurement noise results from electronic delays associated with the separate front end, which would be attached to the GPS ground receivers to enable reception of the POINTS signals. We assumed that these delays would be slowly varying, and estimated them in the filter as first-order Gauss–Markov process noise with an amplitude of 30 cm and a 12-hr time constant. Each of the six ground receivers would, therefore, be tracking up to seven GPS satellites at once, plus POINTS. A simulated data set was fit over a 4-day interval (one POINTS orbital period). The estimation and filtering strategy was selected to be nearly identical to that used at JPL for actual GPS data processing, with the exception that POINTS was included as well. Table 2 shows assumptions of the analysis, which includes systematic error contributions from POINTS solar-radiation pressure mismodeling, Earth orientation, relative station locations, and gravity. The differences between the assumptions for the POINTS error analysis and for the TDRS GPS-like beacon analysis reflect that TDRS requires a near-real-time

[4] B. L. Schumaker, D. B. Eldred, R. Ionasescu, J. W. Melody, R. N. Miyake, C. M. Satter, D. Sonnabend, J. S. Ulvestad, and G. Wang, *The POINTS Instrument for TOPS: 1991 Progress Report* (internal document), Jet Propulsion Laboratory, Pasadena, California, November 12, 1991.

[5] J. Ulvestad, *POINTS Orbit Determination Requirements* (internal document), Jet Propulsion Laboratory, Pasadena, California, July 6, 1992.

[6] B. J. Haines and S. M. Lichten, "POINTS Orbit Determination with GPS-Like Beacon," JPL Interoffice Memorandum 335.8-92-036 (internal document), Jet Propulsion Laboratory, Pasadena, California, December 8, 1992.

result and that the POINTS trajectory can be recovered after the fact. In addition, the POINTS analysis was for a year-2000 time frame, so some improved error assumptions were made. Figure 10 shows the expected orbit velocity and position errors for POINTS. Note that ground tracking of POINTS in this analysis was assumed to be available only 25 percent of the time (shown in Fig. 10). This limited tracking scenario accounts for the possibility that the spacecraft could be turned in such a way that the beacon would be pointing away from Earth, depending on where the astronomical sources were located in the sky.[7] Figure 10 shows that the POINTS solution is robust and, based on our assumptions about the capability to model forces on the satellite, can be predicted several days in advance without serious degradation of accuracy.

In summary, a high Earth orbiter (such as POINTS) can in principle be tracked to position accuracy of a few meters and velocity accuracy of 0.1 mm/sec if equipped with a GPS-like beacon at Ku-band and tracked from at least six modified GPS ground receivers which have been retrofitted so that the high Earth orbiter and GPS satellites can be processed simultaneously. For frequencies other than Ku-band, performance may vary slightly since the ionosphere delay calibration from the GPS L1 and L2 signals will have some error, and this will be more important at lower frequencies.

VIII. Future Work

Future work on the GPS-based ground tracking techniques discussed in this article will focus on several different issues. These include the following:

(1) *GPS-like beacon.* For future missions, we are studying desired signal structure and preliminary hardware design for a GPS-like beacon with the characteristics discussed above.[8] The goal is to use commercially available parts and keep power, mass, and cost to a minimum. The trade-offs between an L-band beacon and a Ku-band beacon are being studied as well. The choice of frequency may also depend on frequency allocations for transmitters, and these issues are being investigated as well.

(2) *Altitude range.* The power and antenna patterns (for both the receiver and transmitter) will ultimately limit the distance from which the beacon signal could be detected with small, inexpensive ground antennas. With a properly designed system, missions in interplanetary space could also use the new proposed technology with a GPS-like beacon. Since, ordinarily, deep space missions require significant amounts of tracking time with oversubscribed large DSN antennas, considerable resource conservation and savings could result.

(3) *Geosynchronous analysis.* It is planned to have the TDRS field experiment data collected in late 1993 and analyzed in 1994. The spatial limitations and noise characteristics of the existing TDRS signals will probably determine performance for present-day TDRS tracking. If the experiment is a success, the new tracking techniques could enhance present or future TDRS operations, or provide an alternative method for tracking other geosynchronous satellites. Because even very short (1-km) baselines could provide moderately accurate (50-m) orbit operations (see Fig. 9), secure ground systems could be designed for military applications as well.

(4) *Modified GPS ground receivers.* The performance of the modified GPS ground receivers used to track TDRSs will be monitored closely in the field experiment. Enhanced designs will be considered, including receivers with more channels to serve additional satellite "customers" simultaneously. One key calibration is the bias introduced by using a different front end or antenna to process the high Earth orbiter data as opposed to the GPS data. This "bias" is actually expected to slowly vary in time with an amplitude of approximately 1 nsec (30 cm). If the temporal behavior can be better stabilized and the amplitude of the variation can be reduced, significantly improved performance would result. The minimum number of stations in the ground network with modified GPS receivers will also be considered for different applications.

IX. Summary

This article describes analyses for tracking satellites at high (geosynchronous or higher) altitudes with two techniques: GPS-like tracking (GLT) with a simple onboard beacon, and short-baseline differential carrier phase (SB$\Delta\Phi$) with a distributed clock. The proposed tracking system dispenses with flight GPS receivers on certain high-altitude satellites for which precise orbit determination is

[7] Since that analysis was completed, however, a new spacecraft design has been developed which allows for placement of two beacon transmitter antennas on opposite sides of the spacecraft so that coverage of the Earth is possible more than 75 percent of the time; see footnote 8.

[8] C. Dunn and L. Young, "POINTS Navigation Beacon Strawman Design," JPL Interoffice Memorandum 335.9-003-93 (internal document), Jet Propulsion Laboratory, Pasadena, California, January 29, 1993.

desired, and also requires no DSN antenna time for orbit determination or navigation functions. Instead, a beacon transmits a series of tones which can be detected in modified GPS ground receivers. In the ideal case, the satellite with the beacon can be tracked to an accuracy comparable to that attainable for GPS satellites themselves, even though the beacon may be at a different frequency than the GPS L-band. In practice, we expect some degradation relative to the GPS satellite orbits if the beacon satellite is at a higher altitude or has restricted ground coverage for other reasons. At the present time, orbits for GPS satellites can be determined on a daily basis to about 30–50-cm accuracy. Detailed analysis has been completed for geosynchronous orbiters (TDRSs). With a global beacon at Ku-band and GLT, TDRS orbit determination accuracy of a few meters would be theoretically possible. However, we also considered using TDRS carrier phase restricted to a relatively small footprint in the southwestern United States and a fairly coarse accuracy two-way range observable. For the current TDRS capability, the study initially focused on SB$\Delta\Phi$ with a fiber-optic link providing a distributed clock to three sites in close proximity. This article presents a hybrid tracking scheme which combines elements of the GLT and SB$\Delta\Phi$ techniques for a demonstration experiment planned for late 1993. For this experiment, three GPS ground receivers are being retrofitted to enable simultaneous tracking of TDRS and GPS satellites. Anticipated orbit accuracy for TDRS is in the 20–50-m range, with the caveat that the longitude component could be degraded by a factor of about five due to unknown errors in the presently available TDRS two-way range. The small tracking network to be tested in the demonstration experiment would offer a number of operational advantages for orbit determination of TDRS or any other geosynchronous satellite. Figure 11 summarizes the relationship between present-day GPS orbit determination accuracy and how anticipated performance for tracking TDRS in the demonstration depends on various data restrictions and limitations.

The covariance analysis for a satellite (POINTS) in a 100,000-km near-circular orbit shows that few-meter position (and <0.1 mm/sec velocity) accuracy could be achieved with a minimal tracking network consisting of only six ground sites equipped with the modified GPS receivers. The analysis assumed that due to blockage on the satellite only 25 percent of the data would be available and shows that, even with such a limitation, the GLT-based solution is very robust.

Operationally, the advantages of the GLT and SB$\Delta\Phi$ concepts discussed in this article include the low cost of small antennas and GPS receivers in comparison with larger antennas and systems typically used for Earth orbiter ground-based tracking; high accuracy from cancellation (or calibration) of media, Earth platform, and timing errors with the simultaneous observations of the GPS and the high Earth orbiters; and operational convenience available from existing rapid communication links and automated processing of GPS (and GPS-like) data. Use of the GPS-based ground tracking system for high Earth orbiter tracking could result in substantial operational savings compared to technology presently being used. In the case of the DSN, very scarce large antenna time would remain available for interplanetary tracking and telemetry.

Acknowledgments

Figure 3(b) was provided by J. Zumberge, and Figs. 4 and 5 by S. C. Wu.

References

[1] W. Bertiger, S. C. Wu, T. P. Yunck, R. J. Muellerschoen, Y. Bar-Sever, E. S. Davis, S. M. Lichten, T. N. Munson, and P. Willis, "Early Results from the TOPEX/POSEIDON GPS Precise Orbit Determination Demonstration," paper AAS 93-154 presented at the AAS/AIAA Space Flight Mechanics Meeting, Pasadena, California, February 22–24, 1993.

[2] S. C. Wu, "Differential GPS Approaches to Orbit Determination of High-Altitude Earth Satellites," paper AAS 85-430 presented at the AAS/AIAA Astrodynamics Conference, Vail, Colorado, August 12–15, 1985.

[3] S. C. Wu, T. P. Yunck, S. M. Lichten, B. J. Haines, and R. P. Malla, "GPS-Based Precise Tracking of Earth Satellites from Very Low to Geosynchronous Orbits," *Proceedings of the National Telesystems Conference*, Ashburn, Virginia, pp. 4-1–4-8, May 19–20, 1992.

[4] S. M. Lichten and J. A. Estefan, "High-Precision Orbit Determination for High-Earth Elliptical Orbiters Using the Global Positioning System," paper AIAA-90-2954 presented at the AIAA/AAS Astrodynamics Conference, Portland, Oregon, August 20–22, 1990.

[5] B. J. Haines, S. M. Lichten, R. P. Malla, and S. C. Wu, "Application of GPS Tracking Techniques to Orbit Determination of TDRSS," paper presented at the Flight Mechanics/Estimation Theory Symposium, Greenbelt, Maryland, May 5–7, 1992.

[6] S. Nandi, C. D. Edwards, and S. C. Wu, "TDRSS Orbit Determination Using Short Baseline Differenced Carrier Phase," paper presented at the Flight Mechanics/Estimation Theory Symposium, Greenbelt, Maryland, May 5–7, 1992.

[7] S. M. Lichten and W. I. Bertiger, "Demonstration of Sub-Meter GPS Orbit Determination and 1.5 Parts in 10^8 Three-Dimensional Baseline Accuracy," *Bulletin Geodesique*, vol. 63, pp. 167–189, 1989.

[8] J. F. Zumberge, G. Blewitt, M. B. Heflin, D. C. Jefferson, and F. H. Webb, "Analysis Procedures and Results From the 1992 IGS Campaign," *EOS Trans. AGU (Suppl.)*, vol. 73, no. 43, p. 134, October 27, 1992, and presented at the Fall AGU Meeting, San Francisco, California, December 7–11, 1992.

[9] C. D. Edwards, "Development of Real-Time Connected Element Interferometry at the Goldstone Deep Space Communications Complex," *Proceedings of the AIAA/AAS Astrodynamics Conference*, Portland, Oregon, pp. 324–334, August 20–22, 1990.

[10] C. E. Dunn, S. M. Lichten, D. C. Jefferson, and J. S. Border, "Subnanosecond GPS-Based Clock Synchronization and Precision Deep-Space Tracking," *The Telecommunications and Data Acquisition Progress Report 42-111, vol. July–September 1992*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–10, November 15, 1992.

[11] J. A. Estefan, "Precise Orbit Determination of High-Earth Elliptical Orbiters Using Differenced Doppler and Range Measurements," *The Telecommunications and Data Acquisition Progress Report 42-106, vol. April–June 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–22, August 15, 1991.

Table 1. Error models for TDRS GLT error analysis.

A priori for estimated parameters	
Parameter	**Value**
TDRS position (X, Y, Z)	5 km
TDRS velocity (X, Y, Z)	50 m/sec
TDRS solar radiation pressure	5%
GPS position (X, Y, Z)	100 m
GPS velocity (X, Y, Z)	1 m/sec
GPS solar radiation pressure	25%
GPS Y-bias	10^{-12} km/sec^2
GPS carrier phase biases	0.003 sec
GPS/TDRS/station clocks	0.001 sec white noise
Zenith troposphere	40 cm (a priori) + 12 cm/$\sqrt{\text{day}}$ random walk
Consider parameters[a]	
Parameter	**Value**
DSN station coordinates	3 cm
Non-DSN station coordinates	10 cm
GM Earth	2 parts per billion
Lumped Earth gravity field	25% GEM-10 – GEM-L2
X-, Y-pole motion	25 cm
UT1–UTC	0.6 msec

[a] Not estimated, treated as systematic errors.

Table 2. Estimation strategy for joint GPS/POINTS orbit determination.

Data noise (30-min observations)	
Parameter	Value
GPS carrier phase	1 cm
GPS P-code pseudorange	30 cm
POINTS Ku-band beacon case	5 cm random noise over 1 min
POINTS pseudorange (tones)	+30 cm/12 hr systematic error (estimated)
A priori for estimated parameters	
Parameter	Value
POINTS position (X, Y, Z)	5 km
POINTS velocity (X, Y, Z)	50 m/sec
GPS position (X, Y, Z)	100 m
GPS velocity (X, Y, Z)	1 m/sec
GPS solar radiation pressure	25%
GPS Y-bias	10^{-12} m/sec^2
GPS carrier phase biases	1 sec
GPS/POINTS/station clock errors	1 sec white noise
Zenith troposphere	40 cm a priori
	+5 cm/$\sqrt{\text{day}}$ random walk
Consider parameters[a]	
Parameter	Value
POINTS solar radiation pressure	2%
Polar motion (X, Y)	5 cm
Geocentric location (X, Y, Z)	5 cm
Earth rotation (UT1–UTC)	0.1 msec
Station locations (X, Y, Z)	1 cm
Earth gravitational constant	1 part per billion
Geopotential field (lumped)	25% GEM-10 – GEM-L2

[a] Not estimated, treated as systematic errors.

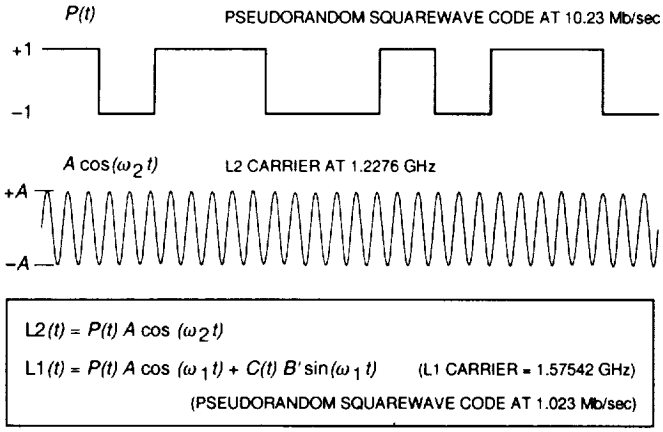


Fig. 1. GPS carrier phase is modulated by the P-code, *P(t)*, and the C/A code, *C(t)*. The codes include data bits (not shown above) with information about GPS orbits and clocks. Carrier phase and high-precision pseudorange obtained from either code or codeless tracking techniques are used in high-accuracy GPS applications. For lower accuracy applications, an inexpensive C/A-code-only GPS receiver may be adequate.

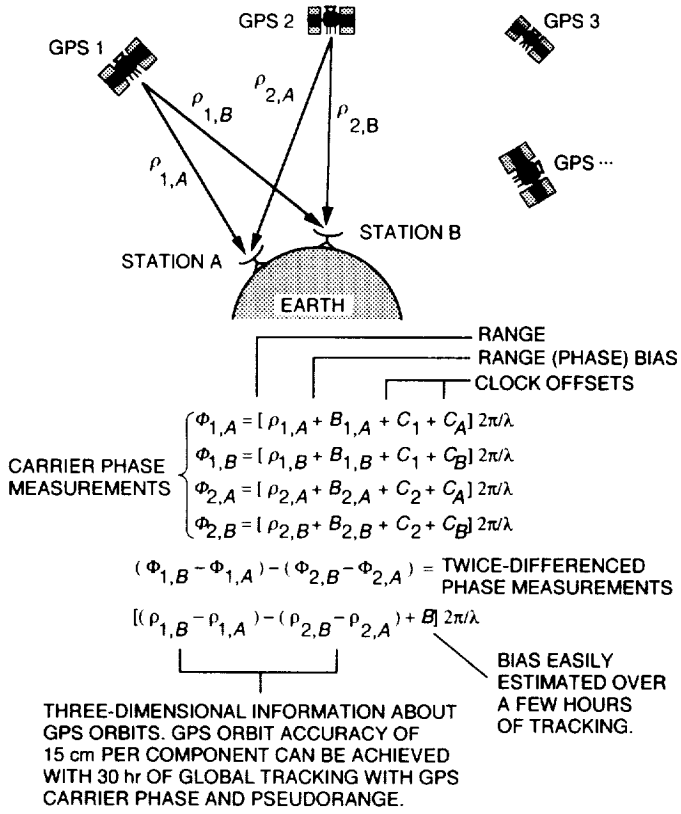


Fig. 2. Differential GPS tracking. Kalman filter simultaneously estimates GPS orbits, biases, *B*, and clock parameters, *C*, instead of explicitly differencing data. Pseudorange data tightly constrain clocks and biases and result in a more accurate solution than is achieved with explicitly differenced data.

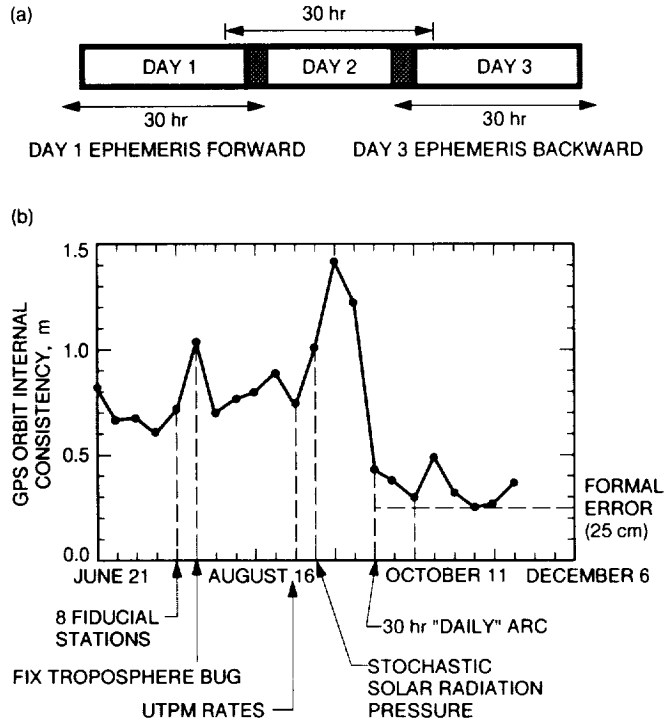


Fig. 3. JPL GPS orbit quality: (a) assessment from daily GPS precise orbit determination, rms computation over 3 hr, and (b) evolution in 1992 using orbit repeatability from day to day similar to that in (a). Present-day results show three-dimensional precision of 30–50 cm, corresponding to 15–25 cm per orbit component.

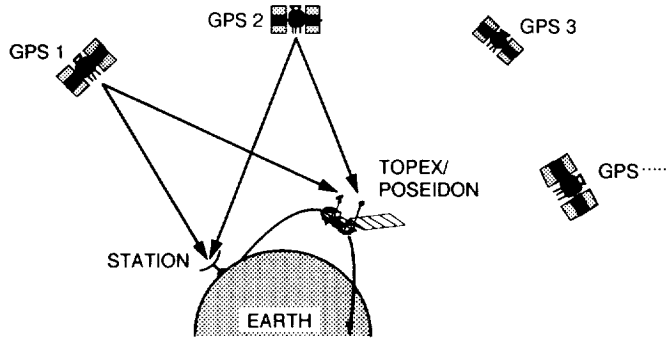


Fig. 4. Geometry for low Earth orbiter tracking with a GPS flight receiver. Simultaneous measurements with common satellites and receivers are necessary for estimation of clock offsets in high-precision applications.

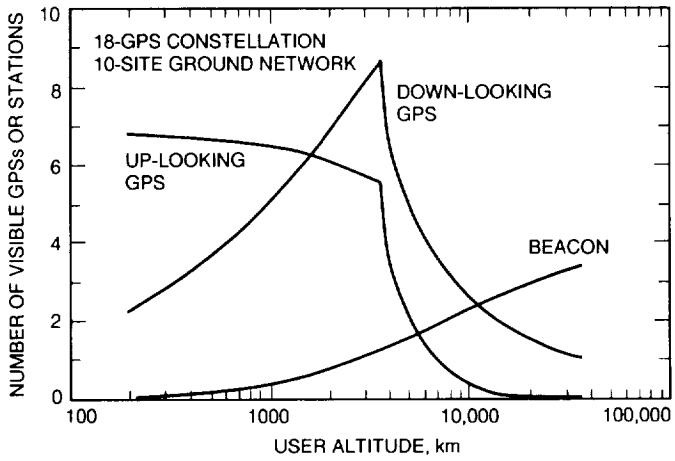


Fig. 5. Average number of GPS satellites in view for a satellite carrying a flight receiver (up-looking and down-looking GPS), and average number of ground stations in view as seen from a user satellite with a GPS-like beacon. The satellite is equipped with a GPS-like beacon that can be tracked simultaneously with the GPS signals by a GPS ground receiver.

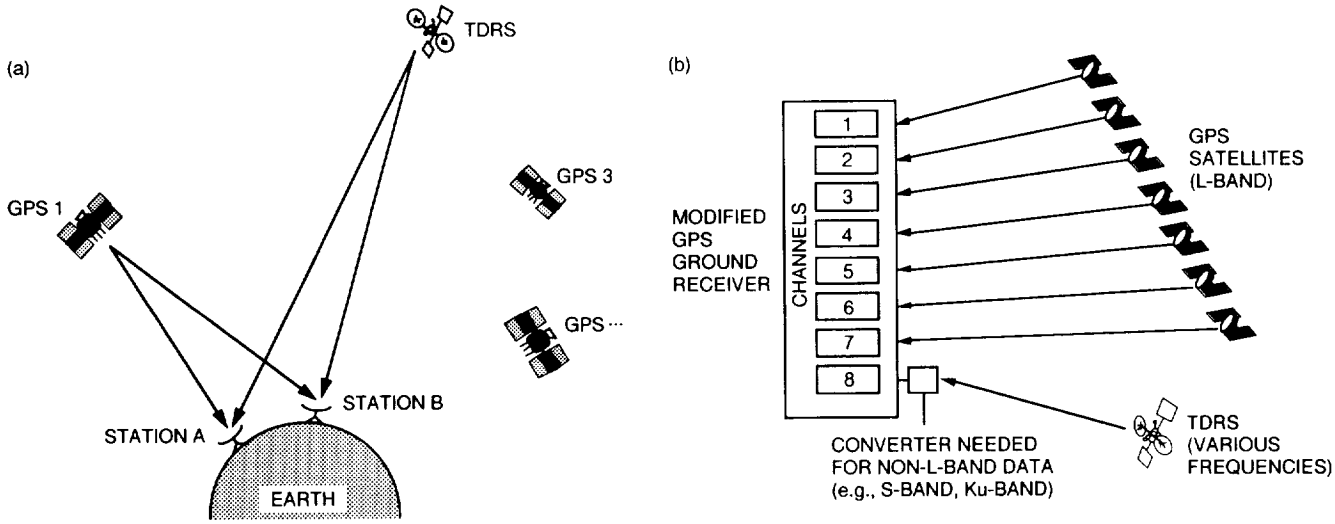


Fig. 6. GLT concept: (a) a high Earth orbiter is equipped with a beacon that can be tracked in GPS ground receivers. The orbit determination problem is similar to a hypothetical GPS orbit determination with one "GPS" satellite at higher altitude, and (b) a GPS ground receiver modified to simultaneously track a high Earth orbiter along with GPS satellites. For a satellite such as TDRS, which broadcasts at frequencies outside the GPS L-band (1.2–1.6 GHz), a small separate antenna with a downconverter would be added to the GPS ground instrument.

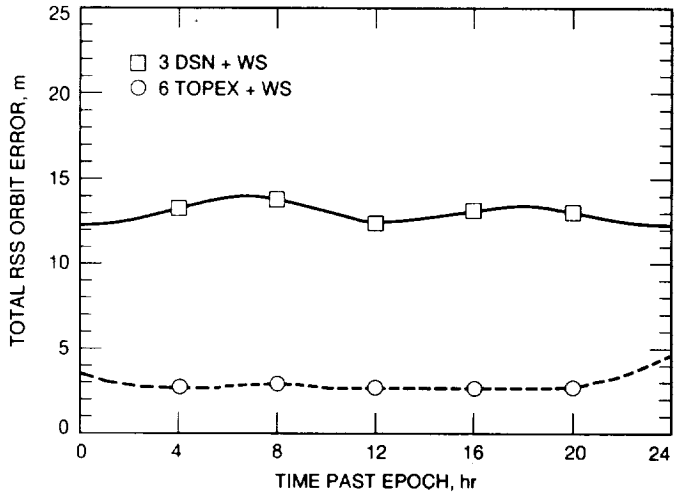


Fig. 7. Expected orbit accuracy for a geosynchronous orbiter (TDRS) equipped with a hypothetical beacon with Ku-band tones to provide a one-way range observable tracked from a global network of four or seven ground sites with modified GPS receivers. The modified GPS receivers would be as shown in Fig. 6(b).

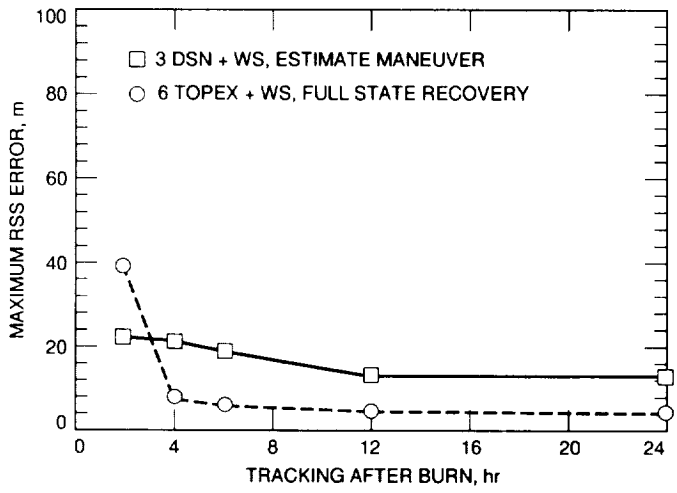
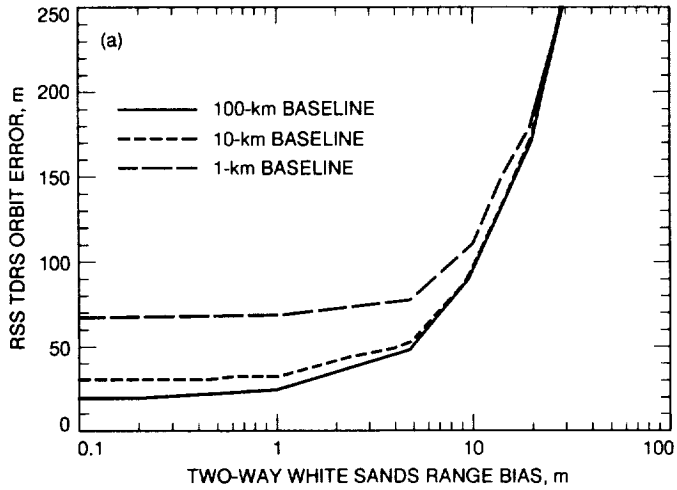


Fig. 8. Expected orbit accuracy corresponding to Fig. 7 after a maneuver for a four-site network where the maneuver is modeled and estimated, or with a seven-site network where an independent new determination of the full state is made.



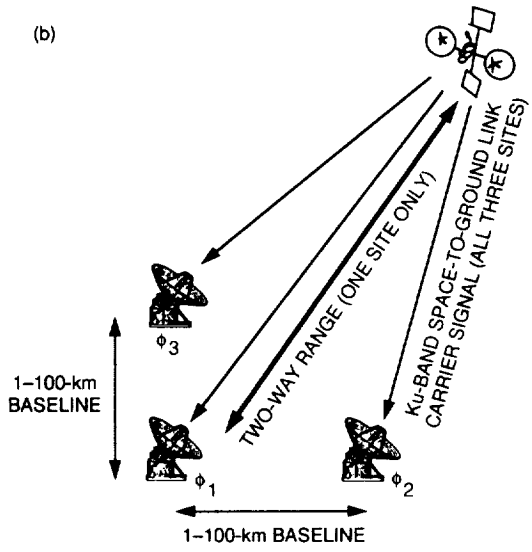


Fig. 9. Hypothetical TDRS tracking with short-baseline differential carrier phase and distributed (connected element) clock plus two-way range to one of the three sites: (a) expected accuracy and (b) schematic of ground network and data links.

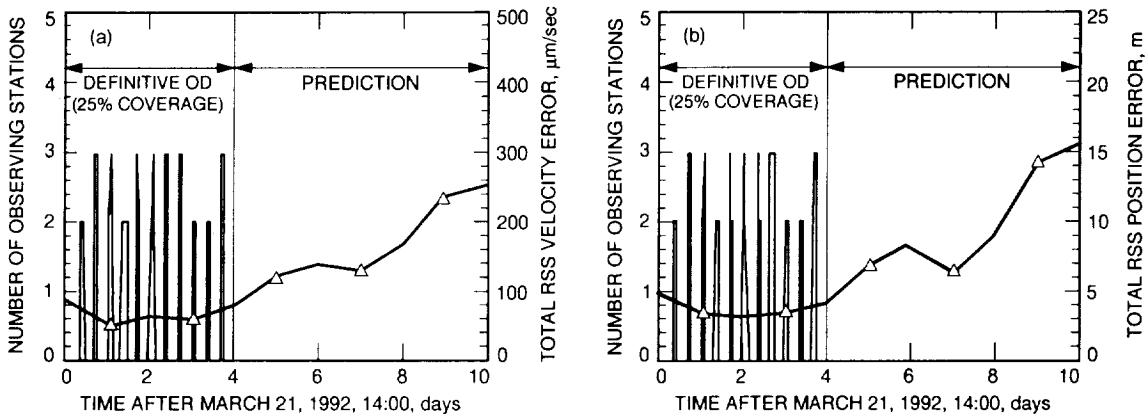


Fig. 10. Orbit determination (OD) for POINTS based on assumptions described in the text and in Table 2: (a) velocity error and (b) position error. "Definitive OD" shows interval over which data are collected with 25-percent coverage, with the number of observations for POINTS also plotted. Triangle symbols show expected accuracy for velocity and position with six ground sites. More than 80 percent of the orbit error is due to data noise and measurement coverage, with only small contributions from the systematic (consider) errors listed in Table 2.

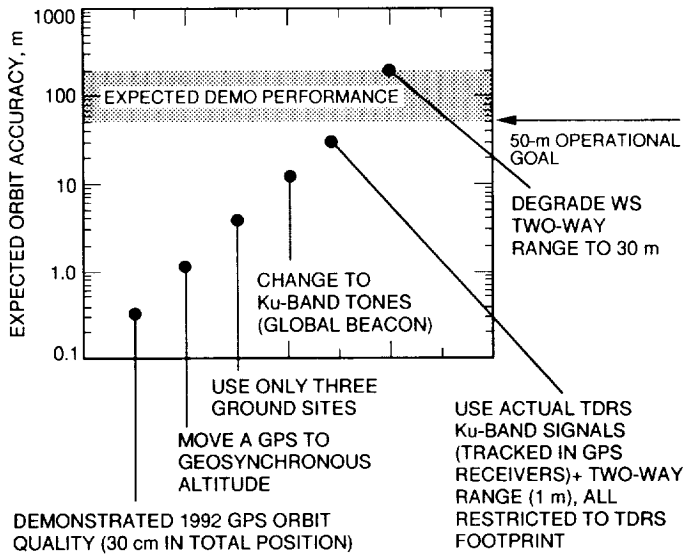


Fig. 11. Comparison of achieved GPS orbit accuracy with expected TDRS orbit accuracy from the demonstration experiment planned for 1993–1994. Geographical limitations of present-day TDRS signals and the data quality will ultimately determine the TDRS orbit quality when tracked along with GPS in GPS ground receivers.

Improved Thermal Force Modeling for GPS Satellites

Y. Vigue
Tracking Systems and Applications Section
R. E. Schutz and P. A. M. Abusali
Center for Space Research, University of Texas, Austin, Texas

Geophysical applications of the Global Positioning System (GPS) require the capability to estimate and propagate satellite orbits with high precision. An accurate model of all the forces acting on a satellite is an essential part of achieving high orbit accuracy. Methods of analyzing the perturbation due to thermal radiation and determining its effects on the long-term orbital behavior of GPS satellites are presented. The thermal imbalance force, a nongravitational orbit perturbation previously considered negligible, is the focus of this article. The Earth's shadowing of a satellite in orbit causes periodic changes in the satellite's thermal environment. Simulations show that neglecting thermal imbalance in the satellite force model gives orbit errors larger than 10 m over several days for eclipsing satellites. This orbit mismodeling can limit accuracy in orbit determination and in estimation of baselines used for geophysical applications.

I. Introduction

Mismodeling of satellite force parameters can have a significant effect on satellite orbits, especially in orbit prediction [1]. Some applications require the capability to estimate and propagate satellite orbits with high precision. TOPEX/POSEIDON precision orbit determination, for example, requires precise modeling of nongravitational forces to fulfill mission requirements [2]. In addition, some of the observed drag and orbit decay on the Laser Geodynamics Satellite (LAGEOS) spacecraft have been attributed to unmodeled thermal forces [3,4]. To achieve a high level of orbit accuracy, an accurate model of all the forces acting on an Earth-orbiting satellite is necessary.

The focus of this analysis was to assess the effects on satellite orbits of neglecting thermal reradiation and of mismodeling nongravitational forces. Radiative heat transfer between a satellite and its environment is the basis for the thermal force model. A satellite in Earth orbit is continuously illuminated by radiation, most of which comes from the sun. The thermal imbalance force is directly related to the temperature distribution of the satellite in its changing environment. An uneven temperature distribution causes surfaces to reradiate energy at different rates. Some studies have shown that most of the thermal gradient forces on the TOPEX/POSEIDON satellite originate within the spacecraft body [2]. Other analyses have shown that the dominant source for thermal reradiation forces on a Global Positioning System (GPS)-like satellite is the solar panels, due to their large exposed area and low heat capacity [5]. The satellite's heated body reradiates energy at a rate proportional to its temperature, losing the energy in the form of photons. By conservation

of momentum, a net momentum flux out of the body creates a reaction force against the radiating surface, and the net thermal force can be observed as a small perturbation that affects long-term orbital behavior of the spacecraft [5]. The partial differential equations and boundary conditions describing the temperature distribution and the heat transfer between surfaces, along with the application of the finite element method, are presented in this article. A brief description of the statistical estimation technique used for studying the effect of the thermal imbalance force on satellite orbits is included.

II. Radiation and Heat Conduction Formulation

Two types of heat transfer that affect a spacecraft's orbit are radiation and heat conduction. The exchange of energy between the spacecraft and its surroundings is described by radiation heat transfer. Conduction is the transfer of heat by molecular motion within a solid medium. Figure 1 shows these types of heat transfer. The rate of radiant energy transfer is given by the Stefan–Boltzmann Law [6]:

$$E_r = \varepsilon \sigma T^4 \tag{1}$$

By conservation of momentum, the thermal force or rate of change of momentum for a radiating surface element, assuming a Lambertian surface, is expressed as [5]

$$\vec{df}_{\text{thermal}} = -\frac{2}{3}\,\frac{\varepsilon \sigma T_a^4}{c}\, dA\hat{\mathbf{n}}_a \tag{2}$$

The unit vector in Eq. (2) is defined as normal to and directed out of the sun-tracking surface of the solar panel. The differential force must be integrated over the entire surface to determine the complete thermal force:

$$\vec{\mathbf{f}}_{\text{thermal}} = -\frac{2\sigma}{3c}\int_{\Omega} \varepsilon T_a^4\, dA\hat{\mathbf{n}}_a \tag{3}$$

Clearly, thermal forces cannot be computed unless spacecraft surface temperatures are known. In general, the temperature at any point within a body satisfies the heat equation [5]

$$K\nabla^2 T = \rho\, C_p \frac{\partial T}{\partial t} \tag{4}$$

The solution to this second-order partial differential equation requires that boundary conditions be specified. The boundary conditions are defined by thermal radiation and heat conduction. As given by the conservation of energy principle, the total amount of energy coming into a surface is equal to the total amount of energy leaving the surface, assuming there is no internally generated or stored energy (no sinks and no sources). The boundary condition for the satellite surface can be obtained by expressing this condition as

$$q_{\text{in}} = q_{\text{out}} \tag{5}$$

where q_{in} is the amount of incoming radiative energy due to external sources and internal conduction, and q_{out} is the amount of radiative energy leaving the boundary due to reradiation and conduction. Figure 2 shows the conservation of energy principle for a satellite solar panel surface. Using this concept, the boundary conditions for each surface were constructed. The incident radiative solar energy received per unit area per unit time by side a and side b of the solar panel are represented by h_a and h_b [5].

$$KA\frac{\partial T_b}{\partial X} = \varepsilon_b \sigma A T_b^4 - h_b A \tag{6a}$$

$$-KA\frac{\partial T_a}{\partial X} = \varepsilon_a \sigma A T_a^4 - h_a A \tag{6b}$$

The actual amount of incident radiative energy received by each side of the solar panel is a function of panel orientation and the orbit of the satellite. The subscript a represents the left boundary in local coordinates (cold side), and the subscript b represents the right boundary, which, for a GPS-type satellite, is assumed to be continuously facing the sun during orbit. The term on the left side of the equal sign in Eqs. (6a) and (6b) is the heat flux, energy per unit time per unit area, in the local x-direction, which is perpendicular to the solar panel face. The values used for some of the parameters described above are listed in Table 1 and are consistent with values used for GPS satellites.

PDE-Protran, a finite element method program, was used to solve the transient heat conduction and radiation problem presented here. PDE-Protran was developed by Granville Sewell and is a general-purpose two-dimensional partial differential equation solver [7]. This software was combined with a program which incorporated material properties, the satellite's orbit orientation, and thermal environment to determine solar panel surface temperatures.

Grid points were chosen to divide the solar panel into small sections or "elements," where the temperature of the solar panel was computed for each grid point in one dimension, across the thickness of the solar panel. These grid points coincide with the boundaries between each layer of the solar panel's "sandwiched" materials (listed in Table 2). Accurate and current knowledge of physical parameters such as surface emissivity, thermal conductivity, heat capacity, and material density is required. For this analysis, the material properties are assumed to remain constant throughout the satellite's orbit, and only the solar radiation environment varies with time as the satellite experiences eclipsing or shadowing from the sun by the Earth. The material parameters directly influence the thermal forces which are calculated and have an effect on the prediction and propagation of the spacecraft trajectory. Also, these material properties may change in time or degrade, due to the harsh environment of space.[1]

III. Orbit Analysis Technique

In this investigation, the equations of motion for an Earth satellite are assumed to include the two-body gravitational effect and the thermal imbalance forces only, and are given in vector form by

$$\ddot{\vec{r}} = -\frac{\mu \vec{r}}{r^3} + \frac{\vec{f}_{\text{thermal}}}{m} \tag{7}$$

and thermal imbalance force perturbing the satellite, $\vec{f}_{\text{thermal}}$, is computed as

$$\vec{f}_{\text{thermal}} = \frac{2\sigma A}{3c} \left(\varepsilon_b T_b^4 - \varepsilon_a T_a^4\right) \hat{n}_a \tag{8}$$

The effect of the thermal imbalance force on a satellite can be observed by comparing the perturbed orbit with the unperturbed two-body orbit in time. Since there is no closed-form analytical solution for the perturbed equations of motion, a numerical integration technique was necessary to solve the ordinary differential equations of motion. Because the perturbed and unperturbed orbits originate with the same initial conditions, the displacement between them at a given time can be observed. A least-squares estimation technique is used to determine the state of the satellite in its orbit at a specified epoch [8]. The initial conditions of one orbit can be adjusted at a given time to eliminate the secular divergence between the perturbed and unperturbed orbits to observe the periodic behavior. In this analysis, two types of GPS satellite orbits were studied. The satellites of the GPS are distributed in six, evenly spaced orbit planes. When completed, the final constellation will consist of 24 satellites at an orbit altitude of approximately 20,000 km, with an orbit period of about 12 hr. In this constellation, most satellites are exposed to full sunlight. As the orbit geometry changes, however, some GPS satellites will experience eclipsing or shadowing from the sun by the Earth. Both eclipsing and noneclipsing satellites are the focus of this study. Throughout its orbit, the GPS solar panel maintains a fixed orientation toward the sun. Nodal motion was not considered, since it is not significant for the short time interval (1 week) used in this study. No internally generated energy was modeled in this study, but the absorbed solar radiation that is converted to electricity was modeled, using the efficiency of the solar panel at 14.1 percent. Although studies have shown that for a TOPEX/POSEIDON satellite the thermal radiation forces originating with the spacecraft body are twice those from the spacecraft solar panels, the major source for thermal reradiation forces on a GPS-like satellite is the spacecraft's thin, large, solar panels [2,5]. Consequently, in this analysis, the GPS satellite's main body was not considered. Other studies are currently considering this problem of modeling thermal reradiation forces for a complete GPS spacecraft.

IV. Discussion of Results

In order to determine the direction and magnitude of the thermal force, the surface temperatures were calculated using the finite element method program, PDE-Protran [7]. Several simulations were tested. The input required for the simulation is shown in Table 2. This table lists the material properties for a Block II GPS satellite solar panel [9,11].[2-5] The initial conditions included a solar panel orientation perpendicular to the sun and an initial temperature of 300 K. The time step used in the analysis was 100 sec (one GPS orbit is approximately 43,200 sec, and the eclipsing period lasts approximately 3200 sec).

GPS satellites experience an eclipsing season for only a few weeks every year. Eclipsing has a strong effect on the solar radiation environment of those satellites. This is

[1] T. T. Lam, personal communication, Aerospace Corporation, El Segundo, California, May 1990.

[2] Ibid.

[3] W. Pence, personal communication, Rockwell International, Seal Beach, California, October 1990.

[4] J. Albeck, personal communication, Spectrolab Corporation, Sylmar, California, August 1990.

[5] D. Marvin, personal communication, Aerospace Corporation, El Segundo, California, January 1991.

evident in the temperature of a GPS satellite solar panel over one orbit, as shown in Fig. 3. The steady-state temperature for the sun-facing side is approximately 317 K, and for the shaded side it is 313 K. These values compare well with the approximate value of 313 K, which has been measured on the cold, shaded side of the solar panel for a GPS satellite.[6,7] The face exposed to the sun has not been directly measured and, therefore, the temperature difference between the surfaces is not well known, but is believed to be approximately 5 K [10].[8–10] During the eclipse period, which lasts approximately 1 hr, panel temperature declines approximately 253 K. After exiting the shadow region, the solar panels slowly return to their steady-state temperatures over approximately 3 hr.

Modeling the cover-glass surface accurately has been difficult during this study since data on the properties of this surface were not readily available. The thermal conductivity of this fused silica layer is very low as compared to that of two other dominant layers, the aluminum core and the solar cell layers.[11] The cover-glass layer, on the sun-facing side of the solar panel, contributes most of the temperature imbalance, primarily because of its low thermal conductivity and greater thickness as compared to other solar panel layers, especially the aluminum core. Although it is believed that the solar panel cover-glass layer is transparent to all incident radiation, the material properties of this specific layer of the solar panel were not deleted from this analysis. It was important to simulate the solar panel as it exists in orbit to observe the long-term orbital effects of the thermal imbalance force on a GPS satellite. This simulation is adequate as long as the correct material properties are used in the analysis.

As an example, two simulations were performed using identical solar panel parameters (values given in Table 2) except for different thermal conductivities for the sun-facing cover-glass layer. These simulations are presented to show the sensitivity of the temperature calculations to the thermal conductivity of the cover glass. The value for the thermal conductivity given in test case 2, shown in Table 3, was used to demonstrate how unrealistic thermal forces can be computed when using incorrect values for the solar panel material properties. Previously, however, this was believed to be the correct value for the thermal conductivity of the fused silica cover-glass layer of a GPS satellite solar panel [9].[12] The results, shown in Table 3, describe the steady-state temperatures and thermal accelerations that were computed using the specified values for the cover-glass thermal conductivity. Again, both test cases shown in Table 3 are identical except for the value of thermal conductivity for the solar panel cover-glass layer.

In this article, the reference frame is defined as spacecraft-centered radial and along-track components. The along-track component is also referred to as the transverse or down-track direction, defined in the direction of the satellite velocity vector. Figure 4 shows radial and along-track components of the acceleration due to thermal reradiation over one orbit for an eclipsing satellite. These components compare well with those of studies which have shown unmodeled nongravitational forces to cause errors of the magnitudes shown in Fig. 4 [10]. Also, these results were computed using the information presented in Table 2 and described as test case 1 in Table 3.

Figure 5 shows the differences between two orbits, one computed using two-body effects only and the other—an eclipsing orbit—computed with two-body effects and the thermal imbalance force during 1 week. The radial rms is 0.5 m, and the along-track rms is 5.2 m. These results were computed using a technique similar to the method used to predict satellite orbits based on a set of initial conditions and a complete force model of the spacecraft, which could include the solar radiation pressure and thermal imbalance force. In the case of the eclipsing satellite after 7 days, the along-track components differ by approximately 13 m.

Figure 6 also shows the differences between two orbits, one computed using two-body effects only and the other computed using two-body effects and the thermal imbalance force for a satellite not in an eclipsing plane. The radial rms is 0.5 m, and the along-track rms is 1.6 m. It can be seen from these results that an eclipsing satellite experiences a larger perturbation in the along-track direction over the span of 1 week than a satellite not in an eclipsing orbit plane. For the noneclipsing satellite after 7 days, the along-track difference is approximately 5 m.

Figures 7 and 8 represent the results computed using a least-squares estimation algorithm in which the simulated observation data contained only the two-body gravitational and thermal imbalance reradiation forces. The force model used in the estimation algorithm contained the two-body gravitational force model and a solar radiation pressure model to observe the ability of the force model to account for thermal imbalance forces which have

[6] Pence, op. cit.

[7] Albeck, op. cit.

[8] Pence, op. cit.

[9] Albeck, op. cit.

[10] Marvin, op. cit.

[11] Ibid.

[12] Lam, op. cit.

been difficult to model but exist in the observations. The calculated best estimate of the satellite epoch state, in the least squares sense, includes the satellite position, velocity, and a solar-radiation pressure scale factor.

Figure 7 shows the orbit fit residuals for a satellite in an orbit plane that is regularly eclipsing. The radial rms is 5 cm, and the along-track rms is 80 cm. After 7 days, the along-track orbit error is almost 2 m. These results show that the solar-radiation pressure scale factor in the estimation scheme is capable of absorbing most of the orbit error due to thermal reradiation, but not all of the orbit error, especially in the along-track direction.

Figure 8 also shows orbit fit residuals for a GPS satellite using the same estimation technique, but for a satellite in a noneclipsing orbit plane. The radial rms is 9 mm, and the along-track rms is 17 cm. After 7 days, the along-track orbit error is approximately 40 cm. Clearly, the eclipsing of the satellites has an influence on the orbit errors when a thermal reradiation force is not included in the estimation force model. Larger orbit errors are calculated when the satellite is in an eclipsing orbit plane. A 1-week prediction can be made using the satellite state computed for the best least squares estimate in Fig. 7 and compared to the best least squares estimate for that predicted week. Studies have shown that, for eclipsing satellites, the quadratic-like growth in the along-track direction can give errors as large as 50 m after a 1-week prediction [12].

V. Concluding Remarks

The current analysis has shown that orbit errors larger than 10 m occur when mismodeling nongravitational forces such as the thermal imbalance force presented here. A finite element method technique has been used to calculate satellite solar panel temperatures which are used to determine the magnitude and direction of the thermal imbalance force. Although this force may not be responsible for all of the force mismodeling, conditions may work in combination with the thermal imbalance force to produce accelerations on the order of 1.0×10^9 m/sec^2. One possible contribution currently being studied is the solar panel misalignment acting together with the thermal imbalance force, a contribution that may account for much of the unmodeled perturbations. If submeter-accurate orbits and centimeter-level accuracy for geophysical applications are desired, a time-dependent model of the thermal imbalance force should be used, especially when satellites are eclipsing and the observed errors are larger than those for satellites in noneclipsing orbits. One study has shown that estimating additional stochastic solar radiation parameters improves GPS orbit accuracy significantly, especially for eclipsing satellites [13]. This technique can be used to absorb the orbit error caused by mismodeling thermal imbalance forces.

Although modeling the spacecraft solar panels alone may be considered insufficient, thermal force modeling of the entire spacecraft is a complicated problem. This has been done for spacecraft such as TOPEX/POSEIDON, where precise orbit determination is critical to mission success [2]. The study presented here, however, focused only on modeling the solar panels, where the material composition is not nearly as complex. Also, the problem of radiation absorbed and conducted through the solar panel and reradiated out is a simple one-dimensional time-dependent heat-transfer problem, with no internal heat generation from scientific instruments or electronics.

Nongravitational perturbations like the thermal imbalance force have been observed for years on satellites like LAGEOS, and are still not completely understood. Thermal forces are dependent on the environment and specifically on such parameters as the satellite mass, cross-sectional area, and material composition. Unfortunately, these parameters can change or degrade with long-term exposure in space. For this reason, it may be more appropriate to estimate stochastic force parameters to represent the thermal reradiation forces since the nature and rate of material degradation of the satellite in orbit is unknown [13]. The results obtained using the finite element model in this study agree with the work of others who have conducted similar studies using the finite difference technique to determine spacecraft thermal gradient forces in an effort to improve the satellite force models.

Acknowledgments

Consulting on PDE-Protran was provided by Granville Sewell of the Center for High Performance Computing at the University of Texas at Austin. The authors would like to thank Lt. Randy White of the United States Air Force Global Positioning System Joint Program Office in Los Angeles, California, for his assistance.

References

[1] Y. Vigue, *Thermal Imbalance Effects on a GPS Satellite*, Technical Memorandum 90-01, University of Texas Center for Space Research, Austin, Texas, May 1990.

[2] P. G. Antreasian and G. W. Rosborough, "Prediction of Radiant Energy Forces on the TOPEX/POSEIDON Spacecraft," *Journal of Spacecraft and Rockets*, vol. 29, no. 1, pp. 81–90, January–February 1992.

[3] V. J. Slabinski, *LAGEOS Acceleration Due to Intermittent Solar Heating During Eclipse Periods*, paper no. 3.9, Gaithersburg, Maryland: American Astronomical Society, July 1988.

[4] D. P. Rubincam, *LAGEOS Orbit Decay Due to Infrared Radiation from Earth*, NASA Technical Memorandum 87804, National Aeronautics and Space Administration, Washington, DC, January 1987.

[5] R. A. Cook, *The Effects of Thermal Imbalance Forces on a Simple Spacecraft*, Technical Memorandum 89-02, University of Texas Center for Space Research, Austin, Texas, May 1989.

[6] A. J. Chapman, *Heat Transfer*, New York: Macmillan Publishing Company, p. 14, 1984.

[7] G. Sewell, *Analysis of a Finite Element Method: PDE/Protran*, New York: Springer–Verlag, 1985.

[8] B. Tapley, "Statistical Orbit Determination Theory," *Recent Advances In Dynamical Astronomy*, edited by B. Tapley and V. Szebehely, August 1972.

[9] Y. Vigue, *Recent Work on the Effects of Thermal Imbalance Forces on a GPS Satellite*, Technical Memorandum 90-02, University of Texas Center for Space Research, Austin, Texas, p. 5, May 1990.

[10] H. F. Fliegel, T. E. Gallini, and E. R. Swift, "Global Positioning System Radiation Force Model for Geodetic Applications," *Journal of Geophysical Research*, vol. 97, no. B1, pp. 559–568, January 1992.

[11] H. S. Rauschenbach, *Solar Cell Array Design Handbook*, New York: Van Nostrand-Reinhold Company, pp. 467–468 and 493, 1980.

[12] B. Schutz, C. S. Ho, P. A. M. Abusali, and B. D. Tapley, "Casa Uno GPS Orbit and Baseline Experiments," *Geophysical Research Letters*, vol. 17, no. 5, pp. 643–646, April 1990.

[13] Y. Vigue, S. M. Lichten, R. J. Muellerschoen, G. Blewitt, and M. B. Heflin, "Improved Treatment of GPS Force Parameters in Precise Orbit Determination Applications," AAS paper no. 93-159, Pasadena, California, February 1993.

Appendix

Glossary

Character	Purpose and/or usage
A	surface area
c	speed of light
C_p	specific heat
E_r	energy emitted by a real body, summed over all wavelengths
$\vec{\mathbf{f}}$	thermal imbalance force per unit area
h	incident solar radiation received by solar panel
K	thermal conductivity
m	satellite mass, kg
$\hat{\mathbf{n}}$	unit vector normal to surface of solar panel
q	radiative energy
$\vec{\mathbf{r}}$	geocentric satellite position vector
r	radial distance from Earth's center of mass to satellite
t	time
T	temperature
α	surface absorptivity
ε	emissivity
ρ	density
Ψ	solar constant
σ	Stefan–Boltzmann constant
μ	Earth's gravitational parameter
Subscript	
a	solar panel sun-tracking front side
b	solar panel back side, no direct sunlight
in	incoming to the spacecraft
out	leaving the body
r	radiative
thermal	thermal imbalance

Table 1. GPS thermal and orbit parameters.

Model parameter	Value
Initial orbit radius, m	26,550,000
Surface emissivity (ε_a)	0.78
Surface emissivity (ε_b)	0.83
Surface absorptivity (α_a), percent	0.77–14.1 (panel efficiency)
Solar panel surface area (A), m^2	10.832
Satellite mass (m), kg	845
Initial panel temperature ($t = 0$), K	300
Stefan–Boltzmann constant (σ), W/m^2 K	5.6699×10^{-8}
Speed of light (c), m/sec	2.998×10^8
Solar constant (Ψ), W/m^2	1368.2
Total panel thickness (eight layers), m (in.)	0.01478 (0.582)

Table 2. GPS Block II solar panel properties.

Panel layer composition	Thickness, m	Density, kg/m^3	Specific heat, J/kgK	Conductivity, W/mK
Cover glass	0.00749	2186.622	753.624	1.417
Adhesive	0.00005	1079.472	1256.04	0.116
Solar cell	0.00025	2684.84	711.756	147.994
Interconnect cell adhesive	0.00018	1051.793	1256.04	0.116
Kapton cocured	0.000076	1162.508	1130.436	0.1506
Graphite epoxy	0.00019	2186.622	1373.27	0.8706
Aluminum core	0.00635	24.91088	1046.7	250.966
Graphite epoxy	0.00019	2186.622	1373.27	0.8706

Table 3. Cover glass thermal simulation.

Test case 1	Test case 2
$K = 1.417$ W/mK	$K = 0.04327$ W/mK
Hot side $T_a = 317.41$ K	Hot side $T_a = 340.30$ K
Cold side $T_b = 313.66$ K	Cold side $T_b = 285.37$ K
Thermal acceleration = 1.88×10^{-10} m/sec^2	Thermal acceleration = -8.01×10^{-9} m/sec^2

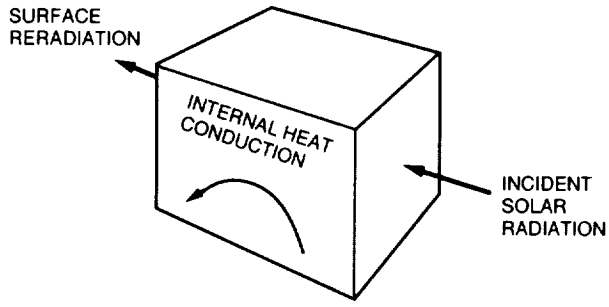


Fig. 1. General heat transfer diagram for a spacecraft.

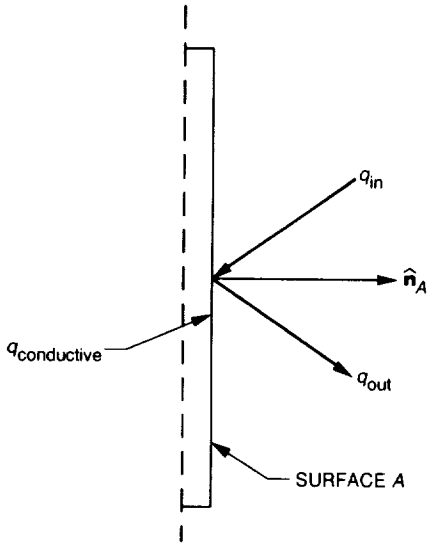


Fig. 2. Conservation of energy diagram for a satellite surface.

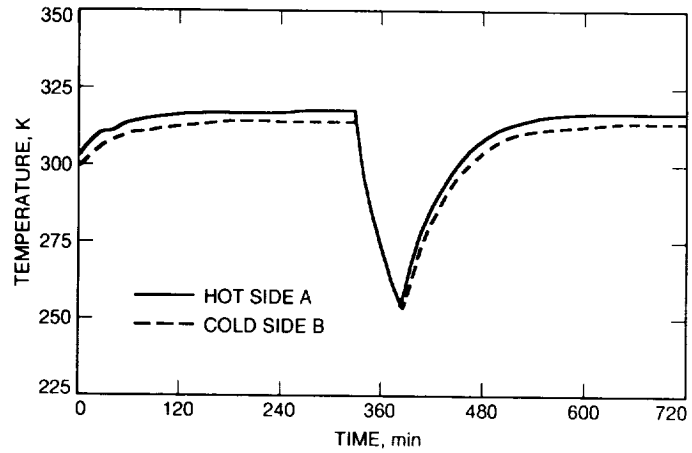


Fig. 3. Temperature history simulation for a GPS solar panel.

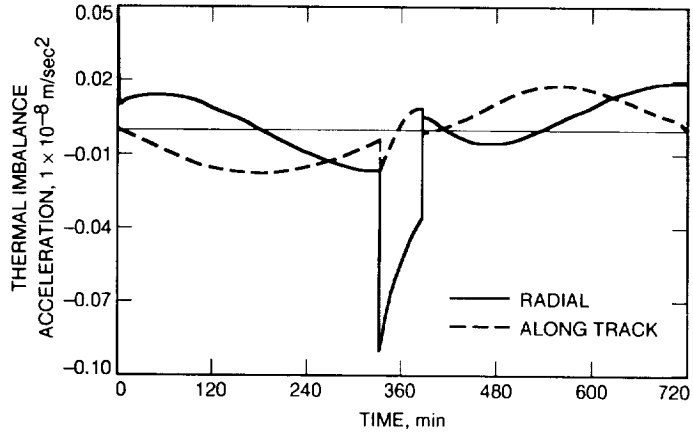


Fig. 4. Radial and along-track components for the thermal force over one orbit.

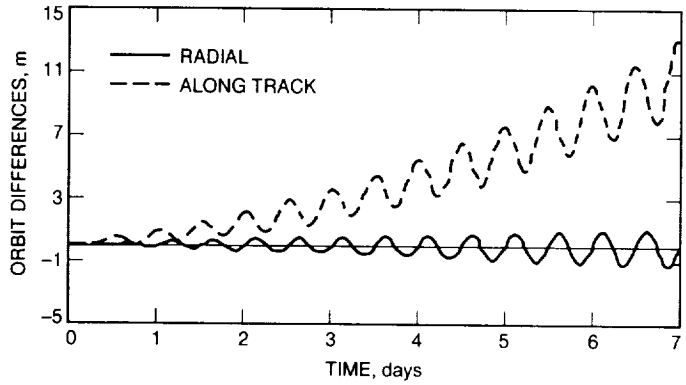


Fig. 5. Radial and along-track orbit differences, eclipsing satellite.

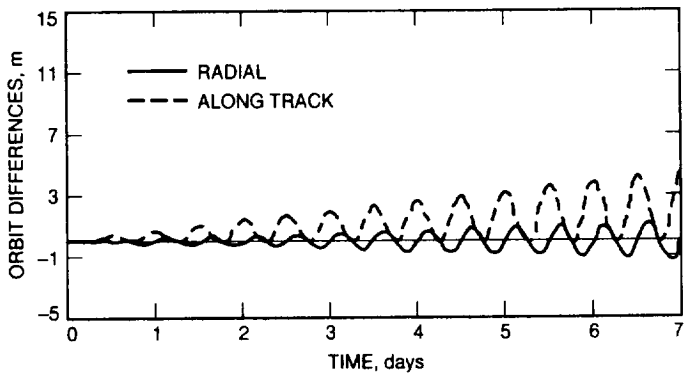


Fig. 6. Radial and along-track orbit differences, noneclipsing satellite.

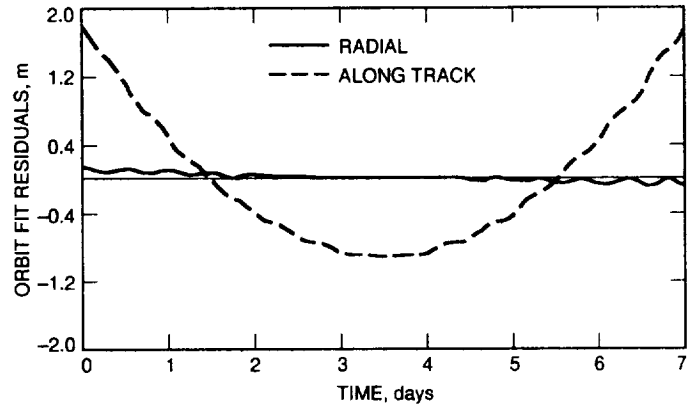


Fig. 7. Orbit fit residuals, with a solar-radiation pressure scale factor, eclipsing satellite.

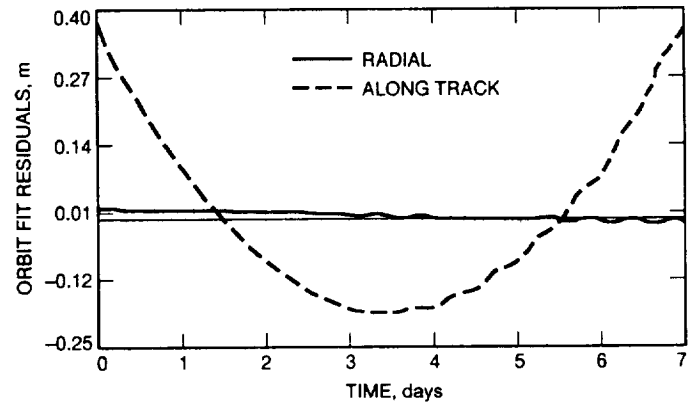


Fig. 8. Orbit fit residuals, with a solar-radiation pressure scale factor, noneclipsing satellite.

Experimental Modification and Identification of the DSS-13 Antenna Control System Model

C. S. Racho and W. K. Gawronski
Ground Antennas and Facilities Engineering Section

This work is the first step toward increasing the bandwidth of the DSS-13 antenna position loop controller. A wider bandwidth of the controller allows for faster tracking rates and better pointing performance under windy conditions. To achieve this goal, the antenna control system model has to be improved, such that it will accurately reflect the dynamic properties of the antenna. The existing analytical model, due to its many uncertainties, could not be used in the design of the controller. However, by using experimental data, the analytical model is modified and improved, and a new model is obtained through system identification techniques.

I. Introduction

The position controller at the DSS-13 beam-waveguide antenna does not employ model-based control law. In the interest of implementing model-based control law at the DSS-13 antenna, an analytical model of this antenna was developed [1]. The experiment described in this article was then designed to validate this analytical model and to identify a new antenna model from the experimental data (the latter model will be referred to as the experimental model). The data collected by this experiment at the DSS-13 antenna were collected for each axis separately. The data were processed and used to modify the analytical model and to obtain an experimental model through system identification techniques. Since the latter approach requires only a portion of the data, a sample was taken from the remaining data and used to validate the experimental model.

II. Description of the Experimental Software and Hardware

The purpose of this experiment is to gather time series data on the input to the antenna drives and the output of the antenna position. These data are processed to determine the frequency responses, or transfer functions, of the DSS-13 antenna. These transfer functions represent an open-loop system. They are used as a reference to determine the correction of the parameters of the analytical model, and also to identify the transfer function of the system using identification software.

A. Experimental Software

The LabView 2 application software generates an analog random (white noise) signal. This signal is input into the rate loop via the test input lead in the servo interface chassis, and the encoder output is sampled from the data converter. The input and the output signals are buffered until the designated number of samples is taken. The sampled data are then saved in floating point MATLAB[1] loadable format files.

B. Experimental Hardware

The data acquisition system consists of the LabView 2 software running on a Macintosh IIfx computer. The com-

[1] Copyright 1985–91, Inc. All rights reserved.

puter contains an analog board which can output analog signals through two channels to the antenna rate loop. It also contains two digital boards which sample the azimuth and the elevation position encoder output (Fig. 1).

The National Instrument NB-MIO-16 multifunction analog board for the Macintosh is used to produce the analog rate signals injected into the antenna rate loop. The board contains two 12-bit digital-to-analog converters with voltage outputs. These two analog outputs are hooked up to the test input leads of either the azimuth or the elevation rate input, or both. The test inputs are located in the servo's interface chassis. The servo interface is switched to manual mode so that a rate signal is accepted from the test input lead instead of the antenna position controller. The brakes in both axes are released manually before the experiment begins.

The National Instrument's NB-DIO-32F 32-bit, parallel, digital, input/output interface board for the Macintosh II is used to sample the binary output of the data converter. The data converter buffers the 24 bits of encoder data available in each axis and provides the data in latched binary form.

A Hewlett Packard frequency counter is used to measure the sample period for each run of the experiment, since the sample period depends on the execution speed of the LabView 2 software, and it cannot be determined in advance.

C. Data Acquisition

The experimental data were collected separately for each axis, namely, for elevation rate input to elevation position output, elevation rate input to azimuth position output, azimuth rate input to azimuth position output, and azimuth rate input to elevation position output. The maximum rate analog input voltage was 1.3 V, and the minimum was 0.1 V.

For each axis and each of its cross couplings, the experimental data were taken in separate sequences. Each sequence contained 8192 data points, and there were anywhere between 10 and 40 sequences per run of the experiment per axis. The available travel in the elevation axis, in conjunction with the rate input and the sample interval, restricted the number of sequences that could be taken per run of the experiment. The azimuth axis, with 450 deg of travel from one limit to the other, never proved to be a limiting factor given the same rate input and sample period.

The data were gathered on several occasions. On each occasion, the wind speeds at DSS 13 were low (between 0 and 3.6 m/sec). When no input was applied to the elevation axis, data were taken for the elevation fixed at both 90 and 60 deg. When no input was applied to the azimuth axis, it was left fixed at 140 deg.

III. Analysis

A model-based control is currently not employed at the DSS-13 antenna. In order to design a model-based controller, a suitably accurate antenna model is required. There were two approaches taken to obtain such a model. First, the model presented in [1] was adjusted to fit the data collected in this experiment. In a second approach, a model was obtained by using a system identification technique based on the data.

A. Model Description

There are two analytical models for the DSS-13 antenna [1,2]: a model for the antenna at 90 deg elevation, and the other for the antenna at 60 deg elevation. These models are described in the state space form

$$\dot{x} = Ax + Bu, \qquad y = Cx \tag{1}$$

In this model, the state vector x is of dimension n, the input u is of dimension p, the output y is of dimension q, and the matrices A, B, and C are of dimensions $n \times n$, $n \times p$, and $q \times n$, respectively. The full order of the analytical model in [1] is $n = 90$, and the reduced order model is $n = 27$. For the system whose input is azimuth and elevation rate and whose output is azimuth and elevation position, $p = 2$ and $q = 2$, and

$$B = [B_{az} B_{el}], \qquad C = \begin{bmatrix} C_{az} \\ C_{el} \end{bmatrix} \tag{2}$$

where B_{az} and B_{el} represent azimuth and elevation input rate, and C_{az} and C_{el} represent azimuth and elevation encoder position.

The four transfer functions (elevation rate to elevation encoder position, elevation rate to azimuth encoder position, azimuth rate to azimuth encoder position, and azimuth rate to elevation encoder position) are used to compare analytical and experimental data. Let $T_{ij}(s)$ be the transfer function defined by

$$T_{ij}(s) = C_i(sI - A)^{-1}B_j, \qquad i, j = az, el \tag{3}$$

By judiciously varying a set of A, B_i, and C_j matrices, one changes the transfer functions such that they better fit the empirical transfer functions. A "good" fit is required only over the frequency range where high coherence between input and output is observed, in most cases for frequencies from 0.01 to 10 Hz. In this case, the good fit means that the peak frequencies and magnitudes of the model transfer function line up with those of the empirical transfer function.

B. Frequency Responses of the DSS-13 Antenna

Using frequency domain analysis, the data were used to modify the analytical model. The input, $u(t)$, and output, $y(t)$, time series data have been detrended and passed through a Hanning filter to prevent spectral leakage [3]. Then, a fast Fourier transform (FFT) was performed on the resulting time series data. The magnitude of the transfer function, $T_{uy}(f)$, and the coherence, $\gamma_{uy}(f)$, were estimated from the time series data (c.f., [3]) using $u(t)$ and $y(t)$, filtered and detrended input and output vectors of 8192 samples each.

$$T_{uy}(f) = \frac{P_{uy}(f)}{P_{uu}(f)} \tag{4}$$

$$\gamma_{uy} = \frac{|\, P_{uy}(f)\,|^2}{P_{uu}(f)P_{yy}(f)} \tag{5}$$

where $P_{uu}(f)$ is the power spectral density estimate of $u(t)$, $P_{yy}(f)$ is the power spectral density estimate of $y(t)$, and $P_{uy}(f)$ is the cross spectral density estimate of $u(t)$ and $y(t)$. The phase is estimated from the time series data by using the following formula:

$$\phi(f) = \arctan\frac{Im(T_{uy})}{Re(T_{uy})} \tag{6}$$

where $Re(.)$ and $Im(.)$ denote real and imaginary parts.

The transfer function estimates were obtained by averaging the magnitude and phase of each sequence. The magnitude, $T_{uy}(f)$, and the phase, $\phi_{uy}(f)$, are plotted versus frequency in Figs. 2(a)–(f) (dashed line), where $f = \omega/2\pi$ is frequency in Hz, and ω is frequency in rad/sec. In Figs. 3(c), 4(c), 5(c), and 6(c), the coherence was also plotted to determine the range of frequencies over which the data were valid.

C. Modification of the Analytical Model

The analytical model as described in [1] is a combination of the antenna structural model and models of the elevation and azimuth drives. It is derived from the best available knowledge of the antenna structure and servos. The antenna structural model is obtained from its finite element model. Although complex, as the finite element model usually is, its accuracy is still limited. For example, damping ratios or "non-structural" masses, such as counterweight mass, are usually only roughly estimated. The available field data allow one, to some extent, to correct some parameters of the analytical model so that its properties fit more closely the properties derived from the experiments. In addition, the drive models have some uncertainties. The gearbox stiffness is not known precisely, since it depends on the countertorque value, which is itself a fuzzy number. Also, the gains of the drive amplifiers are not set precisely, or can change. These and other less known factors impact the model accuracy.

D. System Identification Model

A second model is identified directly from the experimental data using the system identification software System/Observer/Controller Identification Toolbox (SOCIT), written in the MATLAB language [4]. The "okid" function of SOCIT identifies a state space model, i.e., the A, B, and C matrices, given the input–output data, sample period, and the number of observer Markov parameters. For more detailed information about SOCIT software and this function, the reader is referred to [4] and [5].

The SOCIT software was used to identify a model for each axis and cross-axis, so that the A, B, and C state space matrices were obtained for each subsystem. The order of the system was chosen based on the system's Hankel singular values plotted as part of the okid function output. Typically, the order chosen was between 25 and 30. The resulting state space matrices were transformed into balanced coordinates, so that a matrix A was in diagonally dominant form. The 2 × 2 diagonal blocks represent the system modes in decreasing order of importance. The diagonal elements of a 2 × 2 block represent the system damping, and the off-diagonal elements represent natural frequencies at those modes.

The order of each subsystem was reduced to the smallest acceptable order, i.e., such that all the modes visible in the data up to 10 Hz were preserved. The order, n, of the T_{azaz}, T_{azel}, T_{elel}, and T_{elaz} subsystems was n = 14, 13, 10, and 11, respectively. The identified model had some discrepancies (with respect to the experimental data) which apparently could be removed. Namely, the damping

ratios were adjusted for the system modes which appear either underdamped or, more commonly, overdamped.

These four identified subsystems are then used to simulate time series responses, given the actual input data, u_{az} or u_{el}, and small bias signals, u_{azb} or u_{elb},

$$y^s_{azaz} = T_{azaz} u_{az} \tag{7a}$$

$$y^s_{azel} = T_{azel} u_{az} + T_{elel} u_{elb} \tag{7b}$$

$$y^s_{elel} = T_{elel} u_{el} \tag{7c}$$

$$y^s_{elaz} = T_{azaz} u_{azb} + T_{elaz} u_{el} \tag{7d}$$

The use of small bias in the cross-coupling responses was necessary because the straight gains $| T_{azaz} |$ and $| T_{elel} |$ are much larger than the cross gains $| T_{azel} |$ and $| T_{elaz} |$. For the same reason, the bias can be neglected in the straight responses (azimuth-to-azimuth and elevation-to-elevation). The simulated responses, y^s_{azaz}, y^s_{azel}, y^s_{elel}, and y^s_{elaz}, were compared to the actual SISO output data, y_{azaz}, y_{azel}, y_{elel}, and y_{elaz}, and the approximation errors in azimuth and elevation, ϵ_{azaz}, ϵ_{azel}, ϵ_{elel}, and ϵ_{elaz}, were computed.

$$\epsilon_{azaz} = \frac{\| y_{azaz} - y^s_{azaz} \|}{\| y_{azaz} \|} \times 100\% \tag{8a}$$

$$\epsilon_{azel} = \frac{\| y_{azel} - y^s_{azel} \|}{\| y_{azel} \|} \times 100\% \tag{8b}$$

$$\epsilon_{elel} = \frac{\| y_{elel} - y^s_{elel} \|}{\| y_{elel} \|} \times 100\% \tag{8c}$$

$$\epsilon_{elaz} = \frac{\| y_{elaz} - y^s_{elaz} \|}{\| y_{elaz} \|} \times 100\% \tag{8d}$$

These formulas estimate the relative discrepancy between the measured and simulated signals.

IV. Discussion of the Results

A. Results of Analytical Model Adjustments

The magnitude plots of the four transfer functions obtained from the analytical model (solid line) and from the experimental data (dashed line) are shown in Figs. 7(a)–(d). The discrepancies between the analytical and experimental transfer functions are immediately obvious. In particular, there is a mismatch in the first resonance frequency in the azimuth-to-azimuth and elevation-to-elevation transfer functions, and there are higher resonance peaks in the analytical model for almost all resonance frequencies. The mismatch in the first fundamental frequencies is due to the underestimation of nonstructural masses in the structural analysis. The high peaks at the resonance frequencies are caused by an assumption of very low structural damping (0.5 percent). Modifying the rigid-body structural masses (an increase of 45 percent in the elevation rigid body modal mass and an increase of 70 percent in the azimuth rigid-body modal mass) and increasing modal damping for the higher frequency modes to 5 percent improves the fit between the analytical model and experimental transfer function curves. Further improvement is obtained by adjusting the drive parameters. In the elevation and azimuth drives, the amplifier gain, denoted k_r in [1], is reduced by 30 percent (from 80 to 56), and in the azimuth drive, the gear box stiffness is reduced by 15 percent (from 2×10^7 to 1.7×10^7). The results of the modifications of the analytical model show a better fit to the experimental results, as seen in the solid-line plots of Figs. 2(a)–(d). The worst fit between the experimental and model transfer function curves occurs in the elevation input to azimuth output. This is due to the very small value of this function. It is at least 100 times smaller, in magnitude, than the elevation-to-elevation transfer function, or at least 10 times smaller than the azimuth-to-elevation transfer function. The complexity of the analytical model made it difficult to determine what other parameters were responsible for the remaining differences.

The significant modal frequencies as estimated from the data are shown in Table 1.

B. Results of System Identification

In Figs. 3, 4, 5, and 6, the identified transfer function plots are presented with the empirical transfer function plots. Both plots show low frequency anomalies. The cross-coupling transfer functions T_{elaz} and T_{azel} are expected to approach zero value at zero frequency, since no static coupling between elevation and azimuth or between azimuth and elevation is observed. However, in Figs. 4(a) and 5(a), such a tendency is not present. The magnitude of the transfer function from azimuth to azimuth or from elevation to elevation should roll off at a 20-dB/dec rate for low frequencies, since the system contains two integrators (or two poles at zero). This can be observed in the measured frequency range of these plots. However, for fre-

quencies lower than measured ($f < 0.01$ Hz), the identified transfer function does not rise.

The departure of the empirical transfer function plots from the expected zero value in the low frequencies is caused by the presence of a small bias at the input port where no signal is applied. For example, when an elevation rate signal is applied and the azimuth position is sampled, a small azimuth rate input still exists, and it shows up in the output data at low frequencies. Its presence is explained with the comparatively high gain for straight connection (azimuth-to-azimuth or elevation-to-elevation) when compared with the cross-connection (elevation-to-azimuth or azimuth-to-elevation). The unmeasured bias can be explained by the lack of causality between input and output in cross-coupling for low frequencies, as visible in the coherence plots of Figs. 4(c) and 5(c), where the coherence is almost zero. In order to correct this problem, the data were filtered with a high-pass filter before being used in the system identification software. As a result, the cross-coupling transfer function plots are obtained as seen in Figs. 4(b) and 5(b).

For the straight transfer functions (azimuth-to-azimuth or elevation-to-elevation), the discrepancies between the identified model and the empirical transfer function in the very low frequency range are due to the presence of two very small eigenvalues of the identified A. By setting them to zero, the discrepancies were eliminated.

The measurements were used to obtain both the adjusted model and the identified model. However, the identification software requires a relatively small portion of the data gathered. Hence, the remaining time series data, y_{azaz}, y_{azel}, y_{elel}, and y_{elaz}, taken at the DSS-13 antenna are compared to the simulated time series data, y^s_{azaz}, y^s_{azel}, y^s_{elel}, and y^s_{elaz}, of the identified model (Figs. 8–11). The discrepancy between the two signals, relative to the scale of the original signal, are computed using Eqs. (8a)–(8d).

An additional comparison was performed as follows. Define a positive function $r_i(k)$

$$r_i(k)\frac{\| y_i^m - k_i y_i^s \|}{\| y_i^m \|}, \quad i = el,\ az \tag{9}$$

where superscripts m and s denote measured and simulated output, respectively. For $k_i = 1$, r_i shows the distance (or an approximation error) between the experimental and simulated data. One can further improve the fit between the experimental and simulated results by varying the parameter k_i. This additional fit is possible because the system was identified from a detrended set of data, but is compared to a non-detrended one. Let $r_i(k_i)$ achieve the minimal value for $k_i = k_{io}$, i.e., let $r_i(k_i) \geq r(k_{io})$. If $k_{io} = 1$, the simulated series is the best-fitted one and no modification is necessary. Thus, the gain adjustment factor δ_i, defined as

$$\delta_i = \mid k_{io} - 1 \mid \times 100\%, \qquad i = az, el \tag{10}$$

is the measure of good fit. The gain factor δ_i is the percentage that the system gain is adjusted to improve the fit of the simulated time series y_{azazs} and y_{elels} to the measured time series y_{azazs} and y_{elels}.

The approximation errors and the gain factors are presented in Tables 2–5. Notice that the errors and factors are relatively close within each run of the experiment. The percentages for each run are of the same order and are small.

V. Conclusions

This article presents the use of experimental data collected at the DSS-13 beam waveguide antenna to adjust an existing analytical model and to identify a set of single-input, single-output models. Four models (azimuth-to-azimuth, azimuth- to-elevation, elevation-to-elevation, and elevation-to-azimuth) were obtained from the experiment. The measurement and analytical techniques used to arrive at both of these models are described. The experiment designed to gather the field data is also described in detail.

The next step is to obtain a viable and reasonably compact model of the DSS-13 antenna control system. The reduced-order models of the azimuth and elevation axes can and will be developed separately. Both models will then be used in the design of a model-based antenna position loop controller.

Acknowledgments

The authors would like to thank Ben Parvin for his support; Jeff Mellstrom for his valuable contributions; and Leon Alvarez, Jer-Nan Juang (of NASA Langley Research Center), and Michael Thorburn for their helpful discussions.

References

[1] W. Gawronski and J. A. Mellstrom, "Modeling and Simulations of the DSS-13 Antenna Control System," *The Telecommunications and Data Acquisition Progress Report 42-106, vol. April–June 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 205–248, August 15, 1991.

[2] W. Gawronski and J. A. Mellstrom, "Control and Dynamics of the Deep Space Network Antennas," *Control and Dynamic System*, vol. 57, edited by C. T. Leondes, San Diego, California: Academic Press, 1993.

[3] J. S. Bendat and A. G. Piersol, *Random Data, Analysis and Measurement Procedure*, 2nd ed., New York: Wiley, 1986.

[4] J.-N. Juang, L. G. Horta, and M. Phan, "System/Observer/Controller Identification Toolbox (SOCIT)," NASA Technical Memorandum 107566, Langley Research Center, Hampton, Virginia, February 1992.

[5] J.-N. Juang, M. Phan, L. G. Horta, and R. W. Longman, "Identification of Observer and Kalman Filter Markov Parameters: Theory and Experiment," *Proc. of the AIAA Guidance, Navigation, and Control Conf.*, New Orleans, Louisiana, pp. 1172–1179, August 1991.

Table 1. Natural frequencies, f Hz, of the four transfer functions.

f_{azaz}	f_{azel}	f_{elel}	f_{elaz}
1.65	1.62	2.01	1.98
3.24	3.26	3.15	3.21
4.24	4.21	5.22	5.30
–	5.76	–	–

Table 2. Approximation errors for the azimuth-to-azimuth time series.

Data set	ϵ_{azaz}, percent	δ_{az}, percent
1	0.31	0.69
2	0.21	1.00
3	0.27	1.01
4	0.27	1.32

Table 3. Approximation errors for the elevation-to-elevation time series.

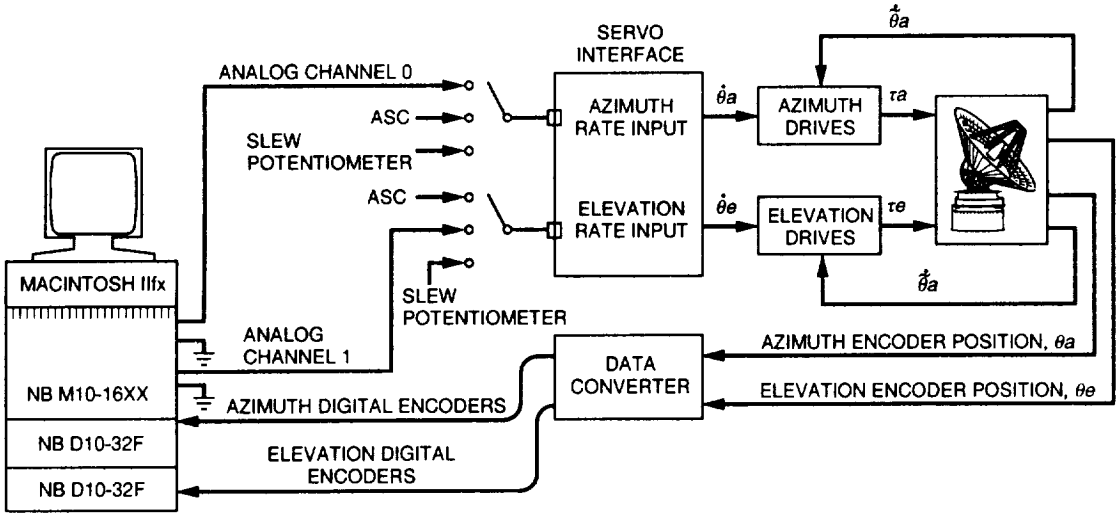
Data set	ϵ_{elel}, percent	δ_{el}, percent
1	0.17	1.07
2	0.17	1.22
3	0.23	2.18
4	0.17	2.05

Table 4. Approximation errors for the azimuth-to-elevation time series.

Data set	ϵ_{azel}, percent	u_{elb}, V
1	6.81	0.0014
2	6.75	0.0013
3	6.36	0.0014
4	5.95	0.0014

Table 5. Approximation errors for the elevation-to-azimuth time series.

Data set	ϵ_{elaz}, percent	u_{azb}, V
1	4.12	0.0014
2	4.57	0.0014
3	4.66	0.0015
4	4.19	0.0014



ASC = ANTENNA SERVO
$\dot{\theta}a$ = AZIMUTH RATE INPUT
$\dot{\theta}e$ = ELEVATION RATE INPUT
$\overset{*}{\theta}a$ = AZIMUTH TACHOMETER FEEDBACK
$\overset{*}{\theta}e$ = ELEVATION TACHOMETER FEEDBACK

Fig. 1. The experimental setup at the DSS-13 beam waveguide antenna.

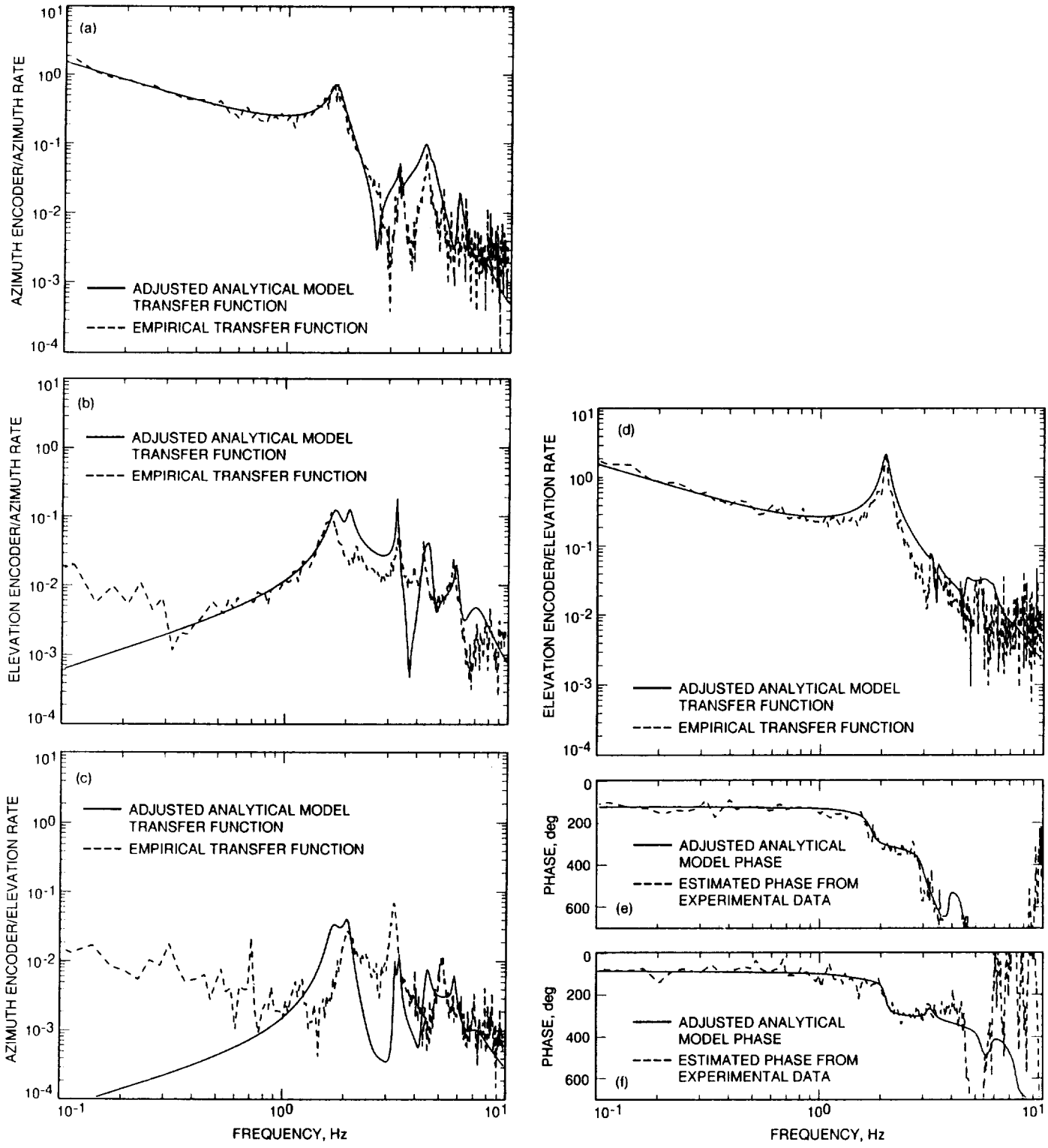


Fig. 2. Magnitude and phase averages: (a) azimuth rate input to azimuth encoder output, (b) azimuth rate input to elevation encoder output, (c) elevation rate input to azimuth encoder output, (d) elevation rate input to elevation encoder output, (e) azimuth rate input to azimuth position output phase, and (f) elevation rate input to elevation position output phase.

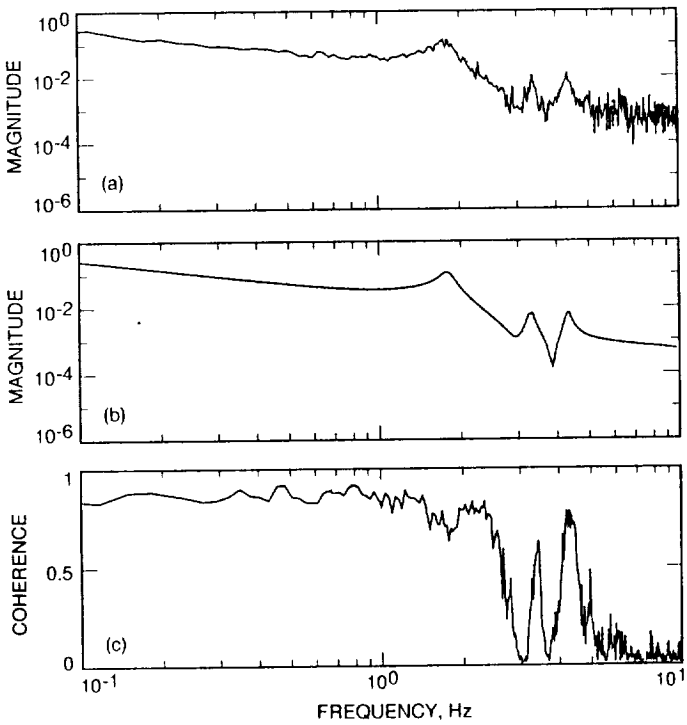


Fig. 3. Azimuth rate input to azimuth position output: (a) estimated transfer function from SISO data, (b) system identification model transfer function, and (c) coherence of SISO data.

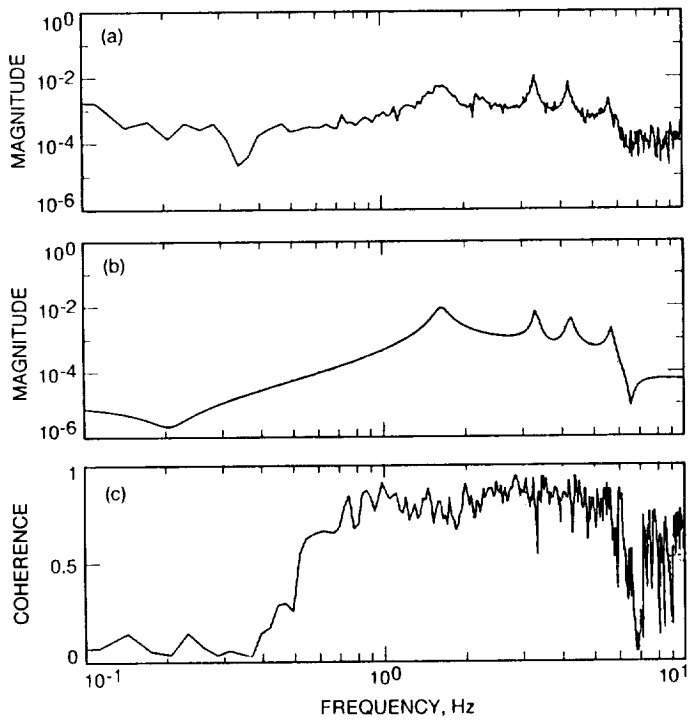


Fig. 4. Azimuth rate input to elevation position output: (a) estimated transfer function from SISO data, (b) system identification model transfer function, and (c) coherence of SISO data.

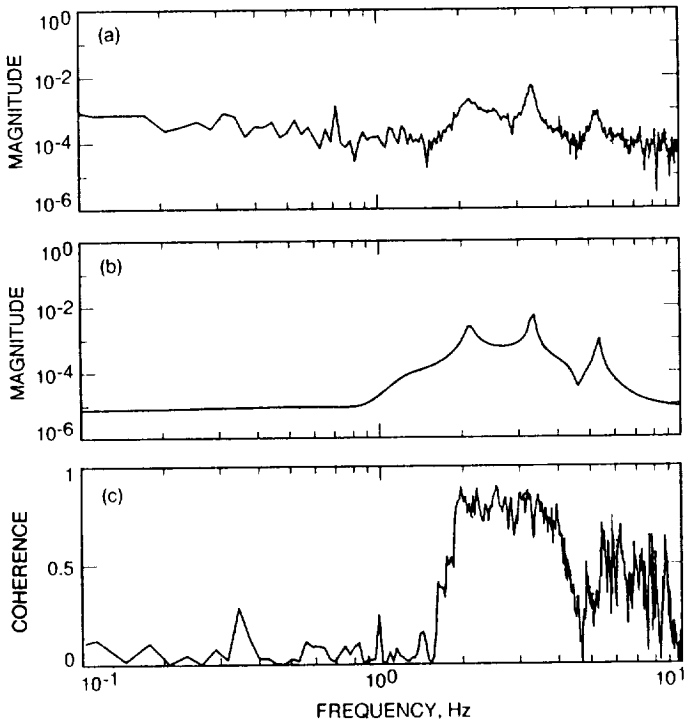


Fig. 5. Elevation rate input to azimuth position output: (a) estimated transfer function from SISO data, (b) system identification model transfer function, and (c) coherence of SISO data.

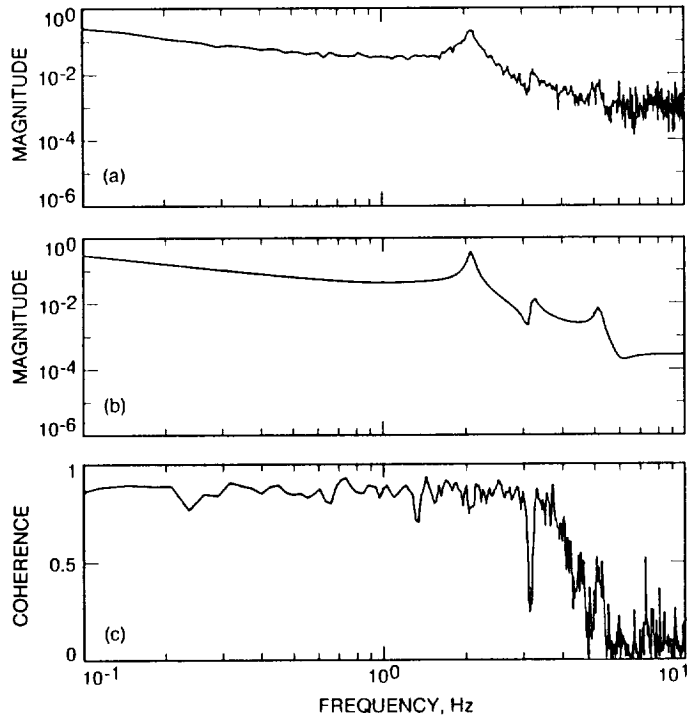


Fig. 6. Elevation rate input to elevation position output: (a) estimated transfer function from SISO data, (b) system identification model transfer function, and (c) coherence of SISO data.

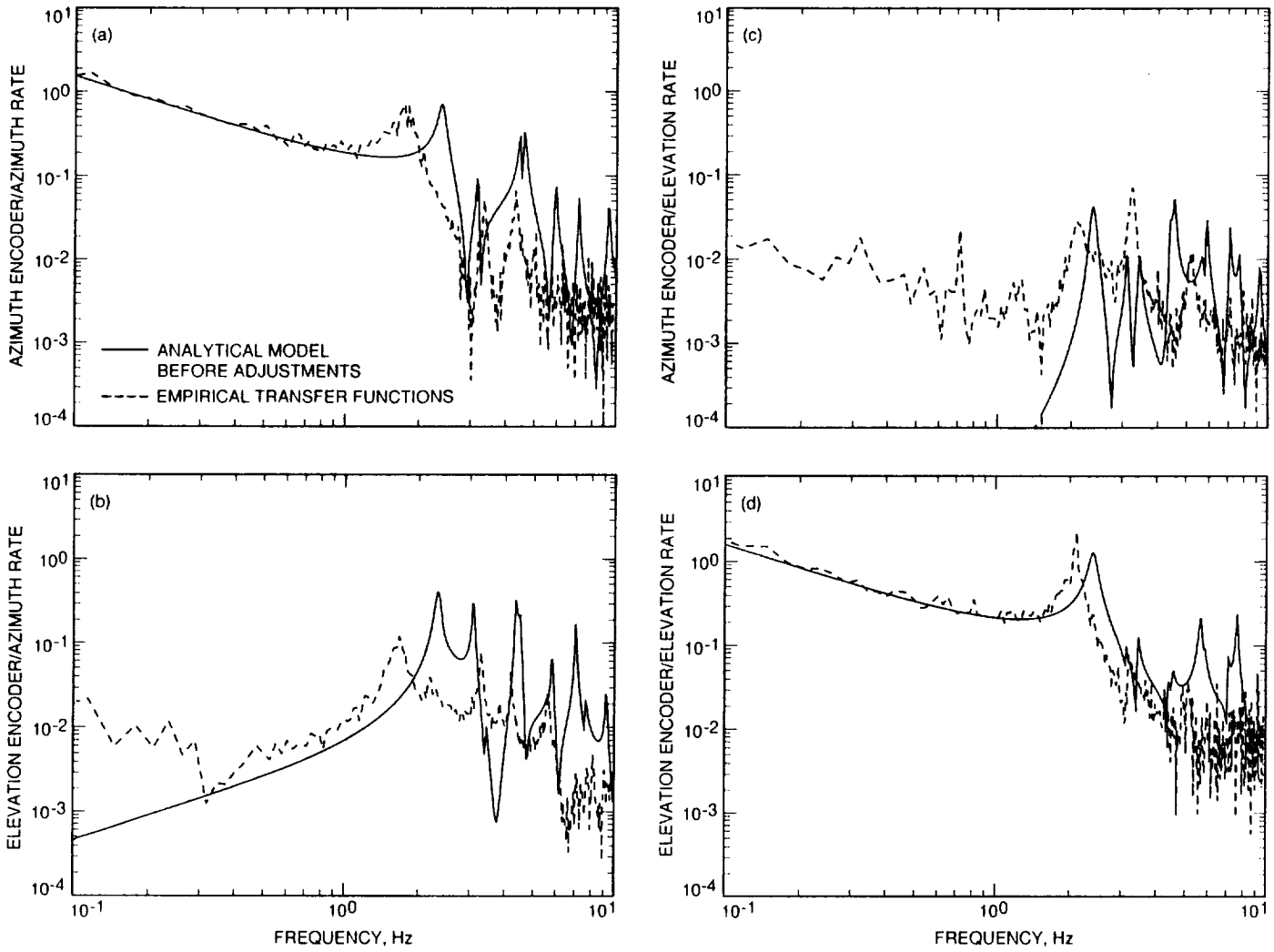


Fig. 7. Magnitude plots of the transfer functions from the analytical model and the experimental data: (a) azimuth rate input to azimuth position output, (b) azimuth rate input to elevation position output, (c) elevation rate input to azimuth encoder output, and (d) elevation rate input to elevation encoder output.

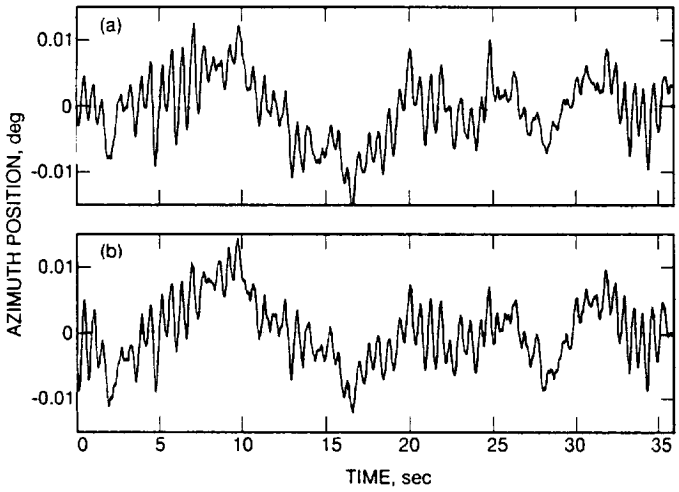


Fig. 8. Detrended azimuth-to-elevation time series output data: (a) measured and (b) simulated.

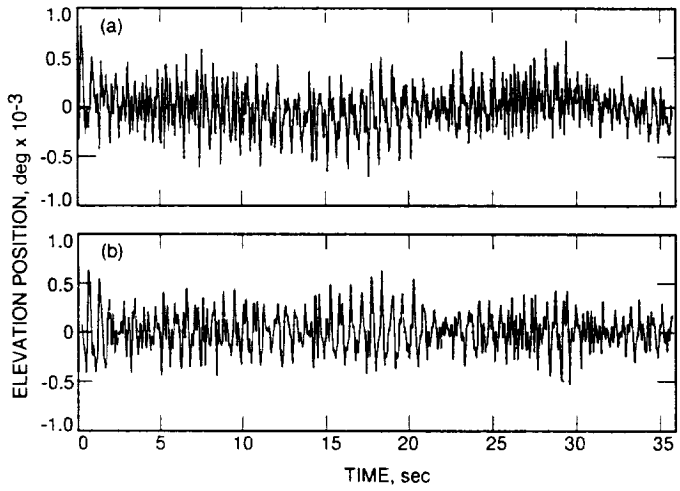


Fig. 9. Detrended azimuth-to-azimuth cross-coupling time series output data: (a) measured and (b) simulated.

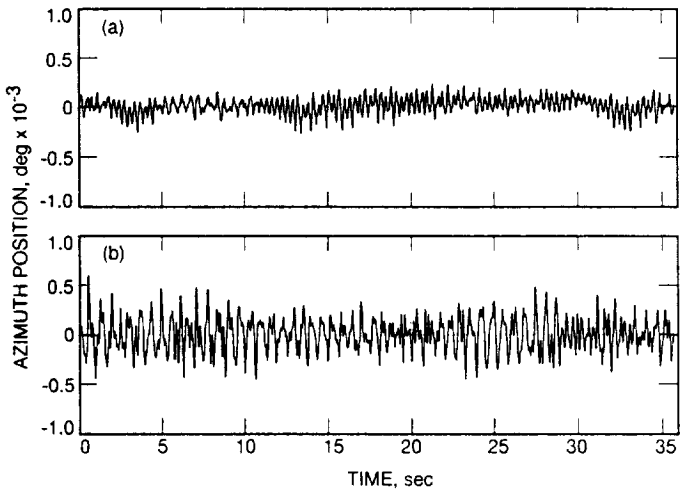


Fig. 10. Detrended elevation-to-azimuth cross-coupling time series output data: (a) measured and (b) simulated.

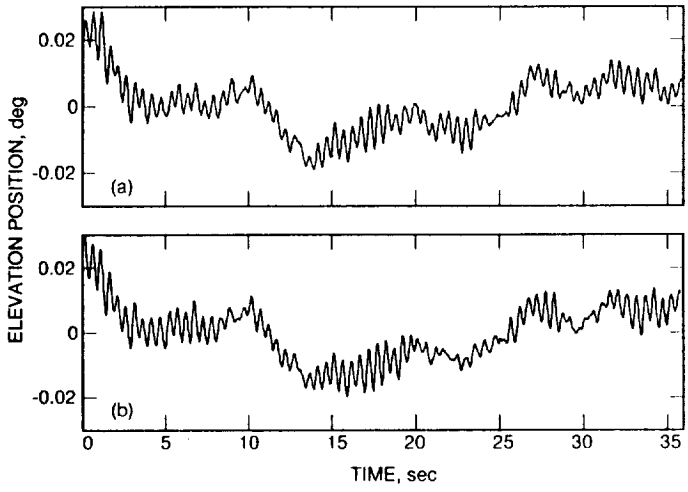


Fig. 11. Detrended elevation-to-elevation time series output data: (a) measured and (b) simulated.

Performance of the X-/Ka-/KABLE-Band Dichroic Plate in the DSS-13 Beam Waveguide Antenna

J. C. Chen, P. H. Stanton, and H. F. Reilly
Ground Antennas and Facilities Engineering Section

The first Ka-band downlink demonstration was recently carried out by the Ka-Band Link Experiment (KABLE) in association with the Mars Observer spacecraft. In order to support the mission, a dichroic plate was required in the DSS-13 beam waveguide antenna to allow simultaneous X- and Ka-band operation. An X-/Ka-/KABLE-band dichroic plate was designed to transmit Ka-band downlink (31.8–32.3 GHz), Ka-band uplink (34.2–34.7 GHz), and KABLE (33.6–33.8 GHz) frequencies, while reflecting X-band (8.4–8.5 GHz). A computer program was developed for the analysis of a dichroic plate with rectangular apertures by using the mode-matching method. The plate was then fabricated and tested. The reflection, group delay, and noise temperature in the antenna system due to the dichroic plate were measured. The experimental results show good agreement with theoretical prediction.

I. Introduction

The DSN has a need for a dichroic plate that will simultaneously receive X- and Ka-bands on a beam waveguide antenna where an ultralow-noise receiver is used at Ka-band. The plate is required to pass a circularly polarized wave at (1) the Ka-band downlink (31.8–32.3 GHz) with a low insertion loss (as low as 0.04 dB), (2) a high-power Ka-band uplink (34.2–34.7 GHz), and (3) the Mars Observer spacecraft Ka-Band Link Experiment (KABLE) frequency (33.6–33.8 GHz), while at the same time reflecting the X-band downlink (8.4–8.5 GHz). A thin dichroic plate with apertures or patches is not mechanically suitable for these requirements; therefore, a thick metallic plate with rectangular apertures was designed and fabricated (Fig. 1). Mechanical constraints also require an oblique angle of incidence, while bandwidth considerations dictate the use of a skew grid.

II. X-/Ka-/KABLE-Band Dichroic Plate Design

The X-/Ka-/KABLE-band dichroic plate was designed using the thick, frequency-selective surface program based on the mode-matching method [1]. The cell size and pattern (D_x, D_y, and Ω), the aperture size (A_x, A_y), and the thickness of the plate (t) are adjusted to meet the requirements. The angle of incidence is $\theta = 30.0$ deg, and $\phi = 0.0$ deg (Fig. 2). The aperture wall thickness is limited to a 0.203-mm (0.008-in.) minimum due to mechanical constraints. The dichroic plate design was optimized with the following priority: (1) Ka-band downlink, (2) Ka-band uplink, and (3) KABLE frequencies.

The optimized design of the X-/Ka-/KABLE-band dichroic plate employs a rectangular aperture of size 5.080-mm (0.200-in.) A_x by 5.156-mm (0.203-in.) A_y, cell size of 6.198-mm (0.244-in.) D_x by 5.359-mm (0.211-in.) D_y

with a 60-deg skew angle (Ω), and plate thickness of 9.271-mm (0.365-in.) t; see Fig. 3(a). The computed reflection loss is about 0.007 dB at 32 GHz, 0.078 dB at 34.5 GHz, and 0.103 dB at 33.7 GHz [2]. The phase difference between the two linear polarizations is 3.4 to 4.0 deg at the Ka-band downlink, 1.8 to 3.8 deg at uplink, and −0.2 to 0.5 deg at KABLE. The phase shift can be corrected by adjusting the polarizer [2]. The overall size of the plate is 765.048 mm by 647.700 mm (30.12 in. by 25.50 in.), with the elliptical perforated area 561.848 mm by 444.50 mm (22.12 in. by 17.5 in.).[1] See Fig. 3(b).

III. Fabrication

The dichroic plate is made of copper for low conductivity loss. A tolerance study indicated that a ±0.0254-mm (±0.001-in.) tolerance on the dimensions of the apertures was acceptable in order to meet the RF requirements [2]. The plate was first drilled with an array of circular holes, then each hole was cut into a rectangular hole with a wire electrical discharge machine, producing higher accuracy and sharp rectangular corners. The walls of the rectangular apertures (waveguides) were then surface-treated with an extrude hone process. In the extrude hone process, two vertically opposed cylinders extrude abrasive media back and forth through the waveguide passages. The surface finish of the dichroic plate is 250μ mm (10μ in.). The average size of the finished apertures is 5.0607 mm by 5.1265 mm (0.19924 in. by 0.20183 in.) with 0.127-mm-radius (0.005-in.-radius) corners according to JPL fabrication. The discrepancy in one dimension of the aperture is thus 0.0193 mm (0.00076 in.), and in the other dimension it is 0.0297 mm (0.00117 in.); these are within the allowable tolerance. Figure 1 is a photograph of the X-/Ka-/KABLE-band dichroic plate.

IV. Reflection and Group Delay

In the experimental measurement of the reflection and group delay, the alignment between the transmit horn, dichroic plate, and receiving horn is very important. There are two kinds of experimental setups: the reflection setup for determining the reflection coefficients, especially the resonant frequencies, which is shown in Fig. 4(a), and the transmission setup for measuring the group delay, which is shown in Fig. 4(b). Both experiments were performed using an HP 8510C network analyzer. The reflection of the dichroic plate at 30 deg from the normal direction was measured at frequencies from 31 to 36 GHz. The measured and calculated reflection coefficients for transverse electric (TE) and transverse magnetic (TM) polarizations are shown in Figs. 5 and 6. There are two resonant frequencies between 31 and 36 GHz for both the TE and TM polarizations. The resonant frequencies of the dichroic plate are within 0.04 to 0.36 percent of the predicted values (Table 1). The combined transmitted power for TE and TM polarizations is shown in Fig. 7.

The group delay of the dichroic plate was measured using the transmission experiment setup. The calculated and the measured group delay versus frequency is shown in Figs. 8 and 9. The group delay is 0.074 to 0.115 nsec from 31.8 to 34.7 GHz for TE and TM polarizations. The group delay is about 0.002 nsec higher than predicted (Table 2).

V. Noise Temperature

In the antenna system, the dichroic plate is illuminated by a horn whose radiation pattern is considered to be a group of plane waves traveling with different amplitudes at different angles. These plane waves will not all strike the dichroic plate at the same angle. Since the dichroic plate is optimized for a 30-deg incident angle, the reflection is minimized at that angle. Therefore, the power reflected by the dichroic plate for horn illumination is larger than the reflection for pure plane wave incidence at 30 deg.

The conductivity loss (I^2R loss) for a rough surface, in addition to the reflected energy from the horn pattern, contributes to the overall noise temperature in the beam waveguide system. The noise temperature, which includes the reflection for plane wave incidence and the conductivity loss for a smooth surface, is calculated to be 1.34, 5.83, and 7.57 K minimum at 32.0 (downlink), 34.5 (uplink), and 33.7 GHz (KABLE), respectively. Measurements were made at the KABLE frequency of 33.7 GHz in the DSS-13 beam waveguide antenna using a corrugated horn. The measured noise temperature at 33.7 GHz was 11 K, which is 3.43 K higher than the calculation, due to the effects described above.

VI. Conclusion

The design of an X-/Ka-/KABLE-band dichroic plate has been presented. The theoretical and experimental results show good agreement in predicting the performance of the dichroic plate (Table 3). The X-/Ka-/KABLE-band dichroic plate is installed in the DSS-13 beam waveguide antenna at Goldstone, California (Fig. 10), for use in the KABLE experiment and future Ka-band operation.

[1] W. Veruttipong and J. C. Chen, "New Optics Design of the X/Ka Basement Feed System for the DSS-13 Phase 2," JPL Interoffice Memorandum 3328-91-0103 (internal document), Jet Propulsion Laboratory, Pasadena, California, August 14, 1991.

References

[1] J. C. Chen, "Analysis of a Thick Dichroic Plate With Rectangular Holes at Arbitrary Angles of Incidence," *The Telecommunications and Data Acquisition Progress Report 42-104, vol. October–December 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 9–16, February 15, 1991.

[2] J. C. Chen, "X-/Ka-Band Dichroic Plate Design and Grating Lobe Study," *The Telecommunications and Data Acquisition Progress Report 42-105, vol. January–March 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 21–30, May 15, 1991.

Table 1. The measured and calculated resonant frequencies of the X-/Ka-/KABLE-band dichroic plate. The calculations are based on 40 waveguide modes.

Polarization	Resonant frequency	Measurement, GHz	Calculation, GHz	Error, percent
TE	First	32.03125	32.05	0.06
TE	Second	34.00625	34.08	0.22
TM	First	32.15625	32.17	0.04
TM	Second	35.95000	35.82	0.36

Table 2. Group delay of the X-/Ka-/KABLE-band dichroic plate for TE and TM polarizations.

Ka-band	Frequency, GHz	Group delay, nsec					
		TE polarization			TM polarization		
		Measured	Calculated	Δ	Measured	Calculated	Δ
Downlink	31.8	0.115	0.114	0.001	0.111	0.111	0.000
	32.0	0.108	0.107	0.001	0.105	0.105	0.000
	32.3	0.100	0.101	0.001	0.099	0.099	0.000
KABLE	33.6	0.085	0.083	0.002	0.077	0.074	0.003
	33.7	0.085	0.083	0.002	0.076	0.073	0.003
	33.8	0.085	0.083	0.002	0.075	0.073	0.002
Uplink	34.2	0.085	0.083	0.002	0.075	0.072	0.003
	34.5	0.085	0.083	0.002	0.074	0.072	0.002
	34.7	0.084	0.082	0.002	0.075	0.072	0.003

Table 3. Performance of the X-/Ka-/KABLE-band dichroic plate installed at the DSS-13 beam waveguide antenna.

Frequency characteristics	Downlink, 32 GHz		KABLE, 33.7 GHz		Uplink, 34.5 GHz	
	Calculated[a]	Measured	Calculated[a]	Measured	Calculated[a]	Measured
Reflection loss, dB	0.007	0.006	0.103	0.096	0.078	0.089
Conductivity loss,[b] dB	0.013 (approximate)	—	0.010 (approximate)	—	0.009 (approximate)	—
Ellipticity, dB	0.294	—	0.220	—	0.603	—
Group delay, nsec	0.106	0.102	0.078	0.080	0.078	0.080
Noise temperature, K	1.34 minimum[c]	TBD[d]	7.57 minimum[c]	11	5.83 minimum[c]	N/A

[a] Calculation is based on perfect plane-wave incidence.
[b] Conductivity loss is calculated for a dominant mode propagating in a rectangular waveguide 5.16 mm by 5.08 mm with a length of 9.27 mm. It does not include the effect of surface roughness.
[c] Calculated noise temperature only includes the reflection loss and conductivity loss.
[d] To be determined.

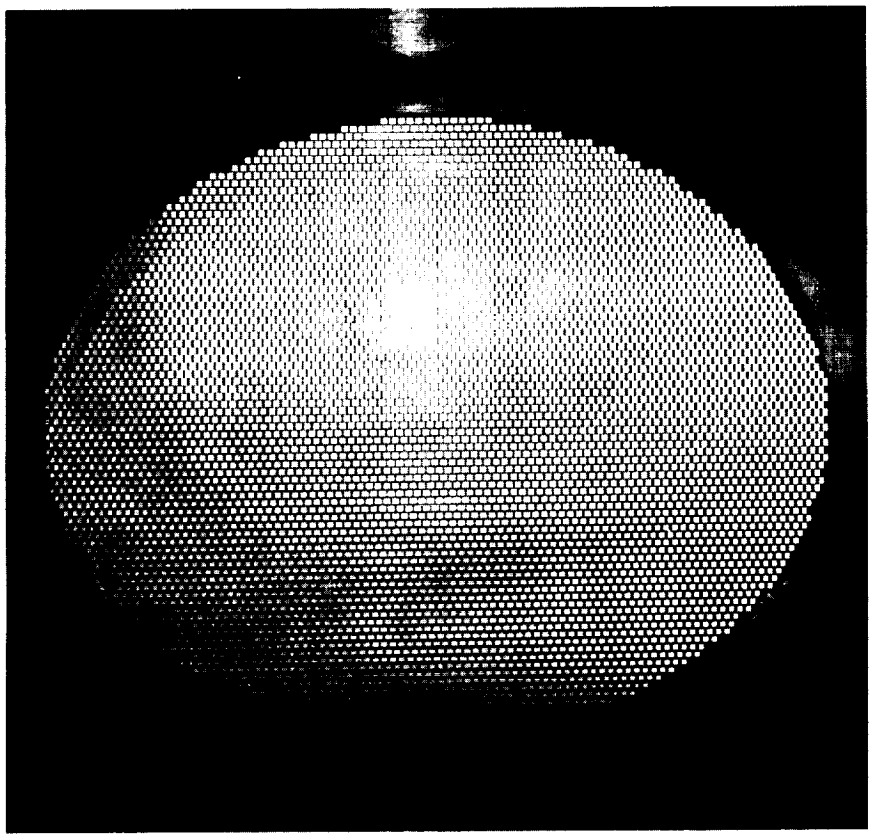


Fig. 1. The X-/Ka-/KABLE-band dichroic plate.

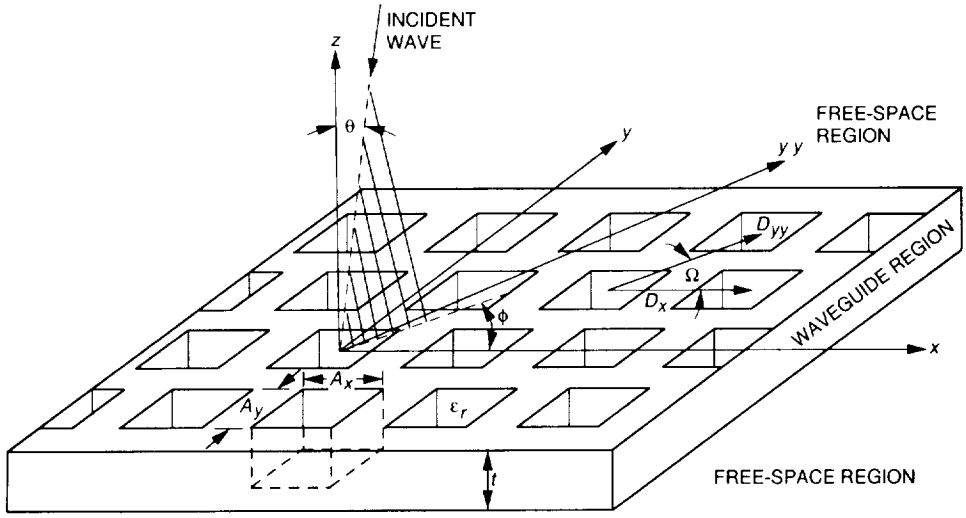


Fig. 2. Geometry of a dichroic plate with rectangular apertures.

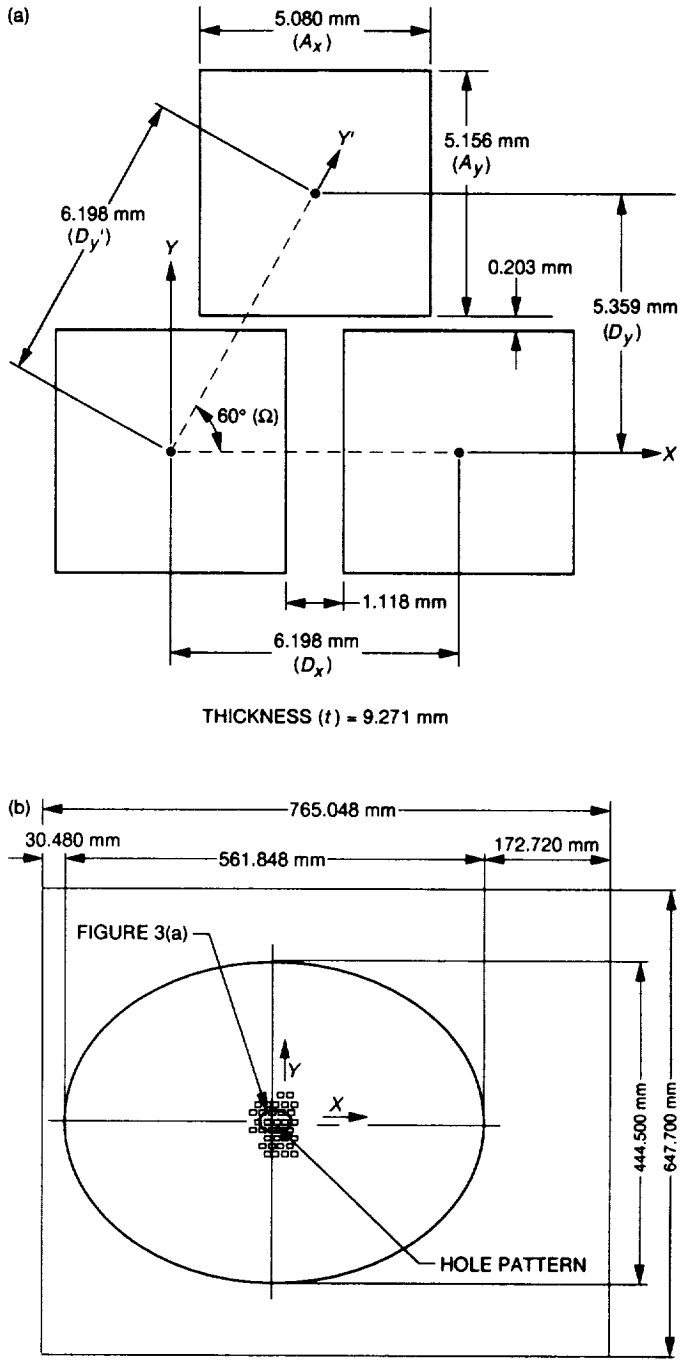


Fig. 3. Diagram of the X-/Ka-/KABLE-band dichroic plate: (a) dimensions and (b) overall size.

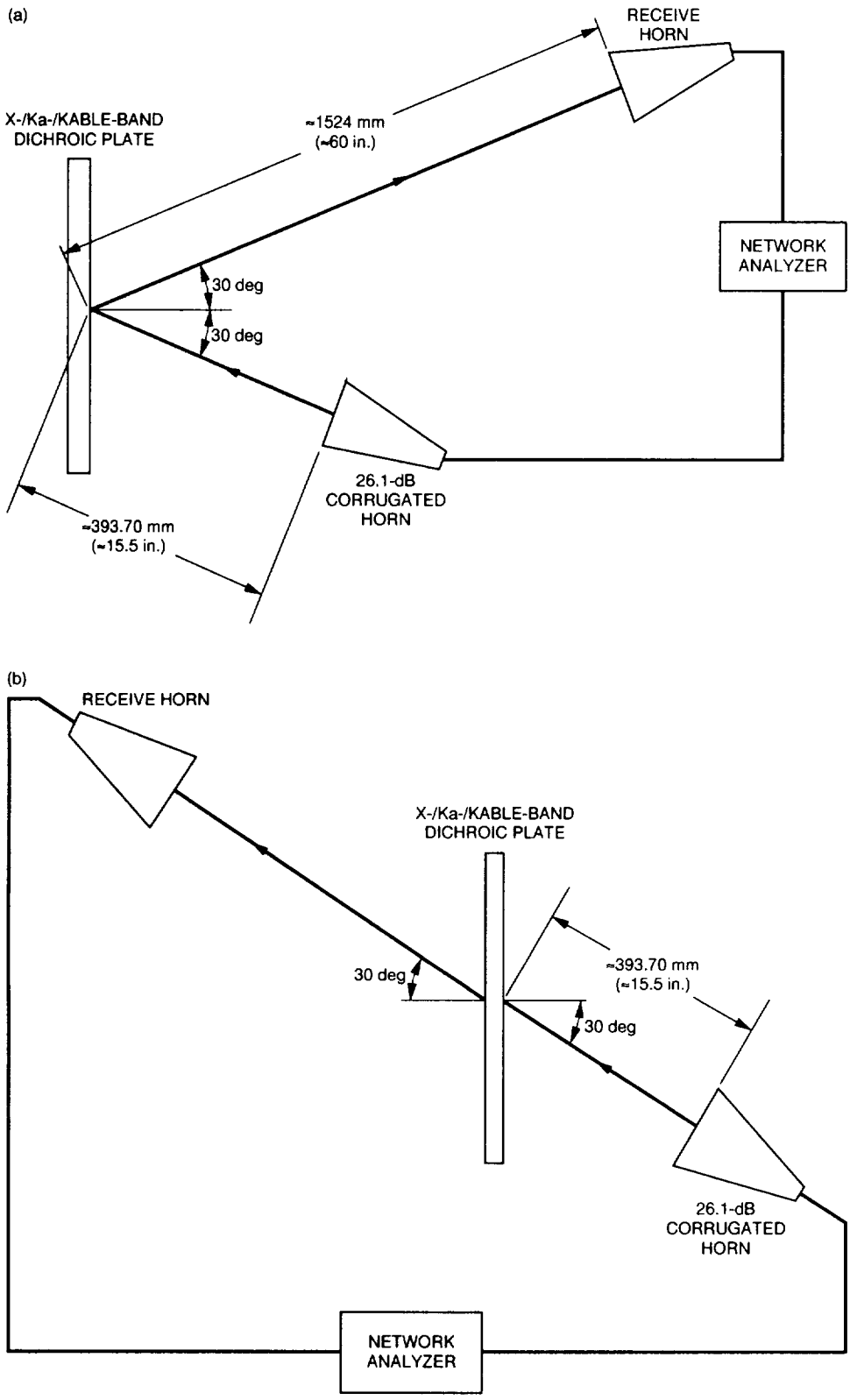


Fig. 4. Experiment setup for (a) reflection measurements and (b) group delay measurements.

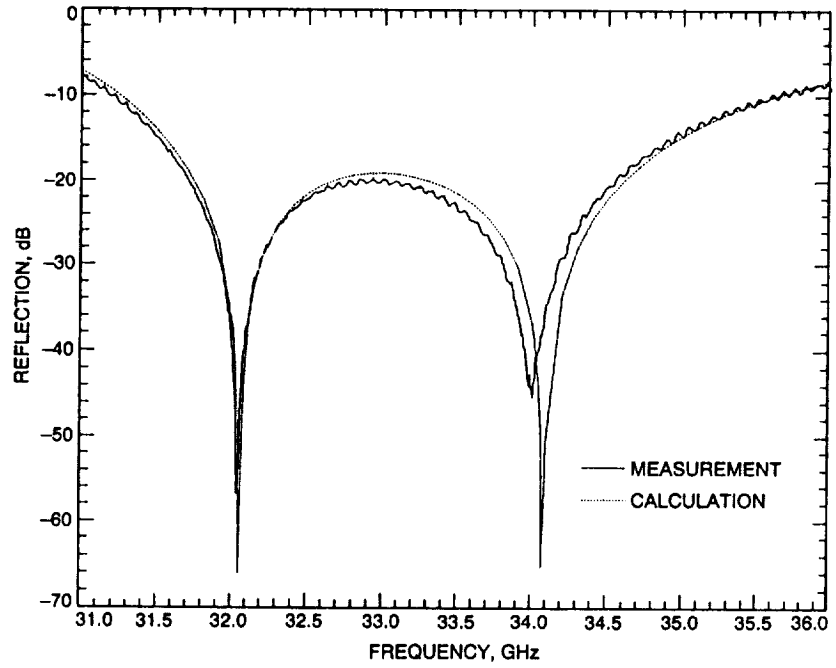


Fig. 5. Calculated and measured reflection coefficient versus frequency for the X-/Ka-/KABLE-band dichroic plate for TE polarizations.

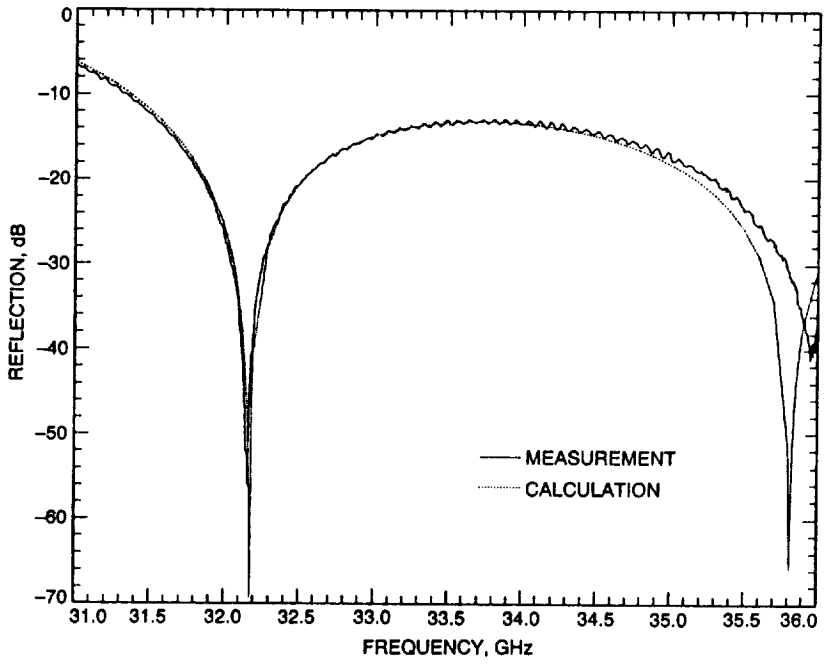


Fig. 6. Calculated and measured reflection coefficient versus frequency for the X-/Ka-/KABLE-band dichroic plate for TM polarizations.

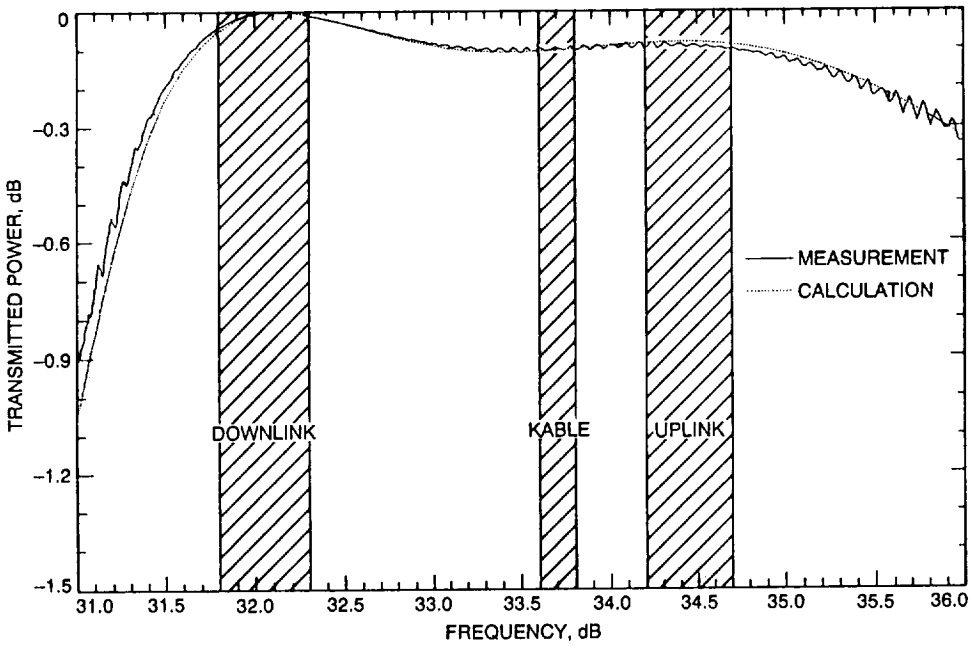


Fig. 7. Transmitted power for the X-/Ka-/KABLE-band dichroic plate.

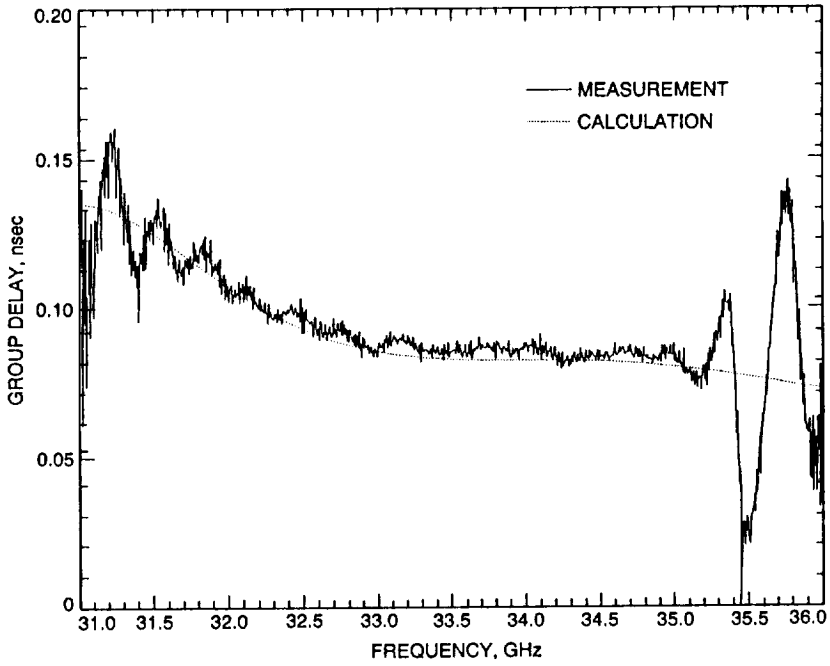


Fig. 8. Calculated and measured group delay versus frequency for the X-/Ka-/KABLE-band dichroic plate for TE polarizations.

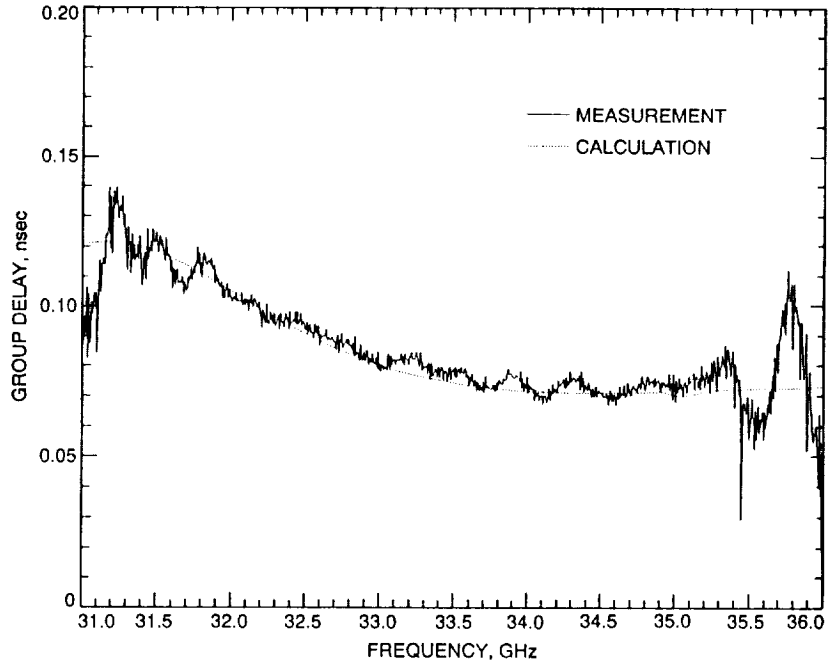


Fig. 9. Calculated and measured group delay versus frequency for the X-/Ka-/KABLE-band dichroic plate for TM polarizations.

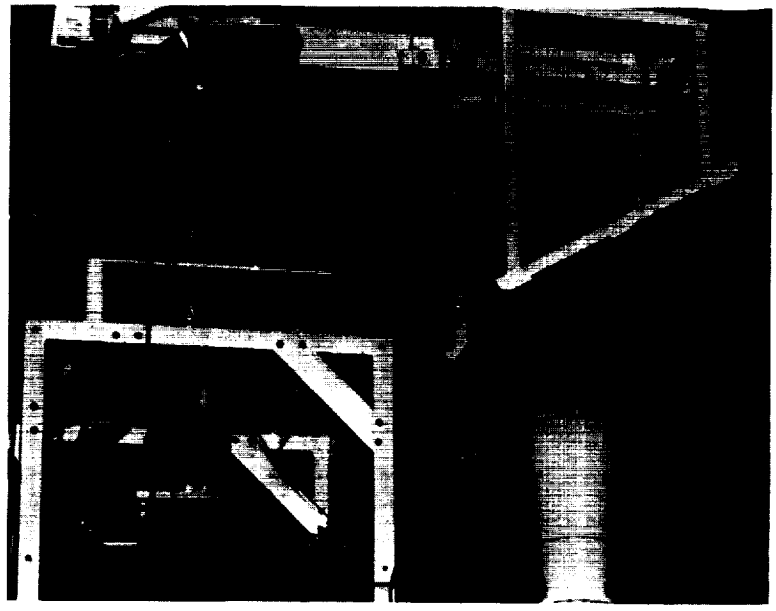


Fig. 10. The X-/Ka-/KABLE-band dichroic plate installed at the DSS-13 beam waveguide antenna at Goldstone, California.

Cryogenic, Low-Noise High Electron Mobility Transistor Amplifiers for the Deep Space Network

J. J. Bautista
Radio Frequency and Microwave Subsystems Section

The rapid advances recently achieved by cryogenically cooled high electron mobility transistor (HEMT) low-noise amplifiers (LNAs) in the 1- to 10-GHz range are making them extremely competitive with maser amplifiers. In order to address future spacecraft navigation, telemetry, radar, and radio science needs, the Deep Space Network is investigating both maser and HEMT amplifiers for its Ka-band (32-GHz) downlink capability. This article describes the current state of cryogenic HEMT LNA development at Ka-band for the DSN. Noise performance results at S-band (2.3 GHz) and X-band (8.5 GHz) for HEMTs and masers are included for completeness.

I. Introduction

A key figure of merit in the specification of the communications link to a deep space mission is the ratio of the gain of the ground-based antenna divided by the system noise temperature (G/T_{sys}). The low-noise amplifier addresses the need to keep the system noise temperature as low as technology permits. Of secondary concern is the provision of broad bandwidth and high gain. Cryogenically cooled amplifiers using masers and high-electron mobility transistors (HEMTs) are employed by the DSN to meet these needs.

Historically, the extraordinarily sensitive receiver systems operated by the DSN have employed ruby masers as the low-noise front-end amplifier [1]. The rapid advances recently achieved by cryogenically cooled HEMT low-noise amplifiers (LNAs) in the 1- to 10-GHz range are making them extremely competitive with maser amplifiers [2,3]. In order to address its future navigation, telemetry, radar, and radio science needs, the DSN is investigating both maser [4] and HEMT amplifiers for its 32-GHz downlink capability.

The telemetry needs at 32 GHz are best met with the lowest noise devices. For bandwidths of less than 400 MHz, maser noise temperatures at this frequency are expected to continue to outperform HEMT noise temperatures. On the other hand, the maser's instantaneous bandwidth is considerably smaller than the HEMT's. Thus, future requirements for large bandwidths, such as interferometric techniques for navigation or for radio astronomy, are more likely to be met with HEMT LNAs. In addition, the use of follow-up HEMT amplifiers reduces the gain requirements for the maser, permitting wider maser bandwidths.

II. Cryogenic Cooling

Cryogenic cooling is applied to a variety of low-noise microwave receivers, such as field-effect transistor (FET) and HEMT amplifiers, mixers, upconverters, parametric

amplifiers, and masers, to reduce thermal noise. To provide physical temperatures down to 4.5 K, commercially available helium closed-cycle refrigerators (CCRs) are usually employed. For temperatures below 4.5 K, a pumped, open cycle liquid helium bath, or hybrid CCR with liquid bath, is used. In the hybrid system, the CCR is used to cool radiation heat shields in order to conserve liquid helium. To obtain the lowest possible noise temperatures, as many as are physically possible of the input and output rf components are cooled below ambient. Currently, the DSN relies on helium CCRs to provide the needed cryogenic temperatures [5].

Maser amplifiers provide the best possible telemetry support for deep space missions. However, these systems operate at a physical temperature of 4.5 K, requiring complex and expensive cryogenic systems. HEMT LNAs, on the other hand, require less cooling power and operate at a higher physical temperature of 12 K, where more cooling power is available. At this higher temperature, a less complex (less expensive) and more reliable refrigeration system can be used. The lower cost of HEMT LNAs will lead to greater frequency coverage and the economic realization of multiple-element room-temperature and/or cryogenic array feed systems. For example, in preparation for Voyager's encounter with Neptune in 1989, JPL planned to array the DSN Goldstone antenna complex with the National Radio Astronomy Observatory (NRAO) Very Large Array (VLA) in Socorro, New Mexico. It cost eight million dollars to equip the 27 VLA antennas with 8.4-GHz HEMT/CCR receivers. To equip the VLA with maser/CCR receivers would have cost approximately 25 million dollars.

Cooling below 4.5 K can result in significant performance improvements for masers, but not for HEMTs. Immersing a maser LNA in a bath of superfluid helium (2.2 K) achieves more than just an improved thermal contact. The gain of a maser in decibels increases in inverse proportion to the physical temperature, while the noise temperature decreases in direct proportion to it. For example, on cooling an 8.4-GHz maser from 4.5 to 1.6 K, the gain in decibels increases threefold, while the noise temperature decreases by a factor of three. Figure 1 shows the noise temperature for physical temperatures from 4.6 to 1.6 K [6].

The noise temperature and gain of a HEMT, on the other hand, are relatively independent of temperature below 12 K.[1] Noise temperature measurements at 8.4 GHz of a three-stage HEMT amplifier in liquid helium from 4.2 to 1.6 K are displayed in Figs. 2 and 3. The results show that the amplifier noise temperature of 5.45 K remains constant on cooling from 4.2 K to the lambda point (2.2 K), decreases abruptly to 5.25 K, and remains constant on further cooling from 2.2 to 1.6 K (the accuracy for these measurements was ±0.1 K). The gain was observed to remain constant throughout the temperature range. Measurements of a two-stage 32-GHz HEMT LNA exhibited similar behavior under the same physical conditions.

Sub-4-K temperature operation is easier to attain in an open helium bath system than in a closed-cycle one. Although sub-4-K physical temperatures are very difficult to implement in a Cassegrain antenna, the laboratory-like environment of a beam-waveguide antenna will allow the implementation of a liquid helium open bath system. In a tipping environment, large (more than 20 deg from zenith) changes in orientation cause significant thermal load increases, resulting in noise temperature increases and reduced operating times. Although the maser will clearly benefit from the advent of beam-waveguide antennas, further understanding of the noise limitations in HEMT devices will result in the development of devices that continue to improve, decreasing to liquid helium temperatures.

III. Noise Fundamentals: Parameters and Models

For purposes of circuit modeling and device characterization, any noisy linear two-port device can be represented as a noiseless linear two-port device with the noise sources at the input and/or output [7,8]. Depending upon the utility of the representation, the internal (voltage and/or current) noise sources can be placed at the input or output port of the noiseless network. Figure 3 shows a convenient representation that leads to four noise parameters (T_{min}, R_{opt}, X_{opt}, and R_n) that can be determined from the measurement of noise temperature as a function of input match. It consists of a series noise voltage (e_n) and shunt noise current (i_n) source at the input [9]. In this representation, the two-port device's noise parameters are given by the equivalent noise resistance

$$R_n = \frac{< |e_n|^2 >}{4kT_oB}$$

the noise conductance

$$g_n = \frac{< |i_n|^2 >}{4kT_oB}$$

[1] J. J. Bautista and G. G. Ortiz, "HEMT Noise at 1.6 K," submitted to the *Journal of the Electrochemical Society*.

and the correlation coefficient

$$r = \frac{< e_n\, i_n >}{\sqrt{< e_n^2 >< i_n^2 >}}$$

where T_o = 290 K, k is Boltzmann's constant, and B is the noise bandwidth.

The noise temperature (T_n) of the two-port device driven by a generator impedance Z_g is given by the expression

$$T_n = T_{min} + \frac{T_o g_n}{R_g |Z_g - Z_{opt}|^2}$$

where Z_{opt} is the optimal generator impedance that yields a minimum noise temperature and $Z_g = R_g + jX_g$ is the generator impedance. The relationship between the first set of noise parameters and those in the above expression is given by the following equations

$$X_{opt} = \frac{Im(C)}{g_n}$$

$$R_{opt} = \sqrt{R_n/g_n - X_{opt}^2}$$

$$T_{min} = 2T_o[g_n R_{opt} + Re(C)]$$

where

$$C = r\sqrt{R_n g_n}$$

In principle, the above noise parameters (Z_{opt}, T_{min}, and R_n) for FET and HEMT devices can be determined by measuring the noise temperature for four or more different known source impedances at a given frequency. However, since there are errors associated with the source impedance and the noise temperature measurements, additional measurements are usually taken to improve the statistics. The noise parameters, along with the scattering parameters, can then be utilized for the optimum design of an amplifier circuit.

The most often used model for device and circuit optimization is the semi-empirical one developed by Fukui [10]. In this model, the noise parameters are simple functions of the equivalent small-signal circuit elements (transconductance, g_m, gate-to-source capacitance, C_{gs}, and source and gate resistances, r_s and r_g) at a given frequency. These circuit elements are in turn analytic functions of the device's geometrical and material parameters. The semiempirical approach of Fukui yields the following expressions for the noise parameters

$$T_{min} = \frac{k_1 \omega T_o C_{gs} \sqrt{(R_g + R_s)/g_m}}{2\pi}$$

$$R_n = \frac{k_2}{g_m}$$

$$R_{opt} = k_3(1/4g_m + R_g + R_s)$$

$$X_{opt} = \frac{2\pi k_4}{\omega C_{gs}}$$

where k_1, k_2, k_3, and k_4 are fitting factors that are determined experimentally. Although this model is widely used by device designers and served to guide the development of the first cryogenic HEMT devices for the Voyager 2 encounter with Neptune, the model provides very little insight into the physics of noise in HEMTs and FETs.

Considerable research has been conducted in the area of noise performance of field-effect transistors over the last two decades. Currently, however, a noise model which is useful for both device optimization and circuit design is not available.

A good noise model must agree with measurements and accurately predict noise parameters. The more recent analytical models that consider fundamental semiconductor steady-state transport properties only treat thermal noise within the channel. These are progressively more complex treatments of van der Ziel's original work [11,12]. The numerical noise model approach taken by Cappy et al. [13] takes into account electron dynamics and appears to explain noise temperature results. However, the dependence of the measured noise temperature on device parasitics and input circuit impedance complicates the full evaluation of numerical models.

A potentially powerful approach for submicron gatewidth cryogenic device development would be a synthesis of M. Pospieszalski's frequency- and temperature-dependent circuit model [14] and Joshin's one-dimensional electron transport noise model [15].

Pospieszalski's model uses simple (small signal) circuit concepts that yield closed-form expressions for the noise

parameters. This model introduces frequency-independent equivalent temperatures for the intrinsic gate resistance (T_g) and drain conductance (T_d). The equivalent noise model for an intrinsic chip device is shown in Fig. 4. For low frequencies, that is for $\omega/\omega_t << \sqrt{(T_g/T_d)(1/r_{gs}g_{ds})}$ and $R_{opt} >> r_{gs}$, the noise parameters are given by the following expressions

$$R_{opt} = \frac{\omega_t}{\omega}\sqrt{\frac{r_{gs}T_g}{g_{ds}T_d}}$$

where

$$\omega_t = \frac{g_m}{C_{gs}}$$

$$X_{opt} = \frac{1}{\omega C_{gs}}$$

$$T_{min} = \frac{2\omega}{\omega_t}\sqrt{g_{ds}T_d r_{gs}T_g}$$

and

$$g_n = \frac{T_{min}}{2R_{opt}T_o}$$

The utility of this model is that it allows the prediction of the noise parameters for a broad frequency range from a single frequency noise-parameter measurement at a given temperature.

In Joshin's microscopic model [15], the intrinsic gate and drain noise generators are derived from the physical properties of the GaAs semiconductor at ambient. By adding a physical temperature dependence to Joshin's model, it may be possible to predict and calculate the noise parameters directly from the physical properties of the semiconductor materials. Most models are not applicable at cryogenic temperatures and have shown limited agreement with room-temperature measurements. In any case, it is clear that much experimental and theoretical work remains before the HEMT device is optimized and exploited for cryogenic LNA applications.

IV. HEMT Devices

The HEMT is essentially a high-performance GaAs FET with a more detailed and layered active material structure. Although the HEMT structure has many similarities to a conventional GaAs FET, different physical mechanisms control or limit the carrier transport properties in the region between the source and the drain. The conventional device structure shown in schematic cross-section in Fig. 5. is grown in layers on a GaAs substrate. First, to inhibit impurity diffusion into the active region, a 1-μm-thick buffer of GaAs is grown on a semi-insulating GaAs wafer. A spacer layer of AlGaAs from 20 to 60 Å is grown to reduce dopant donor ion/conduction electron interactions. Next a 300- to 400-Å AlGaAs layer doped with silicon atoms is grown. Finally, a heavily doped AlGaAs layer 450-Å thick is grown to provide ohmic contacts. This layered structure produces a conduction band discontinuity that forms a triangular one-dimensional quantum well at the AlGaAs heterojunction. Electrons from the AlGaAs layer are attracted to and collect at the one-dimensional conduction band minimum in the vertical direction on the GaAs side of the heterojunction, forming a two-dimensional electron gas in a plane normal to the vertical direction.

The primary advantage of this heterojunction structure is that, unlike the heavily doped channel of a conventional GaAs FET, there are significantly fewer impurities in the undoped GaAs where the two-dimensional electron gas resides. The result is that electrons experience fewer scattering events and thus travel at higher saturated velocities than in conventional FETs. The additional spatial separation provided by the spacer layer of the channel electrons from their parent ions results in enhanced electron mobility. At room temperature, HEMT mobility is more than a factor of two greater than the FET's, while at cryogenic temperatures its mobility is more than two orders of magnitude greater than the FET's.

Additional device enhancements currently being investigated are new materials with higher intrinsic mobility (e.g., InGaAs and InP), shorter gate lengths (<0.10 μm) to further reduce scattering events, and alternate doping strategies such as planar doping (i.e., confining the dopants to an atomic plane) for stronger carrier confinement.

The three-terminal device is completed with the fabrication of a Schottky-barrier gate which controls the number of electrons in the two-dimensional electron gas. In addition to a lower noise figure than for a GaAs FET, the HEMT also has several other inherent characteristics that make it more attractive for low-noise amplifiers. The scattering parameters (S-parameters) of a HEMT in a 50-ohm impedance circuit exhibit lower output return loss (S_{22}) and higher gain (S_{21}) than a GaAs FET of the same dimensions. This results in an inherently

better output match and larger gain-bandwidth product. The HEMT also has much lower noise conductance, g_n, and X_{opt} (where $Z_{opt} = R_{opt} + jX_{opt}$ is the optimum source impedance) than the comparable metal semiconductor FET (MESFET), resulting in a lower noise temperature over a broader bandwidth. In addition, its performance improves more rapidly with cooling than that of the GaAs FET. Figure 6 shows the noise temperature response comparison for a FET and HEMT at 8.4 GHz, as a function of physical temperature.

V. Device Characterization

Since one of the primary functions of the LNA is to minimize the receiver system noise temperature, the characterization and selection of HEMT devices is critical to LNA performance. The selection of the 0.25-μm gate length conventional AlGaAs/GaAs HEMTs for the 32-GHz LNAs was based on its previously demonstrated reliability and exceptionally high-gain and low-noise characteristics [16,17] at 8.4 GHz. The selected device gate width of 75 μm was determined by the trade-offs associated with optimum impedance matching, circuit bandwidth, intermodulation distortion, power handling capability, and power dissipation.

The devices were fabricated on selectively doped AlGaAs/GaAs heterostructures grown by molecular beam epitaxy (MBE) with a Varian GEN II system on a 3-in.-diameter GaAs substrate. The details of the material growth conditions are discussed elsewhere [18]. The HEMT wafer exhibited a sheet carrier density of $8.1 \times 10^{12}/\text{cm}^2$ with a mobility at 77 K of more than 75,000 cm^2/Vsec. All levels were defined by electron beam lithography, and the T-shape gates were fabricated using a tri-layer resistance technique [19] to achieve a low series gate resistance.

For low-noise performance at cryogenic temperatures, the HEMT device must exhibit good pinch-off characteristics and high transconductance, g_m. Good pinch-off characteristics are achieved by strong confinement of the charge carriers to the channel region, with a sharp interface of high quality and a large conduction band discontinuity. An enhanced g_m at the operating bias is obtained by judicious choices for doping concentration and space layer thickness [20]. An Al mole fraction of approximately 30 percent is required for a large conduction band discontinuity, while the high g_m is achieved with a 40-Å space layer and a doping concentration of approximately $2 \times 10^{18}/\text{cm}^3$. Although these values result in a high-performance room-temperature device, at physical temperatures below 150 K the device suffers from IV (current voltage) collapse [21] and exhibits the persistent photoconductivity effect associated with the presence of deep donor traps (so called DX centers). In order to obtain excellent device performance at cryogenic temperatures and to eliminate light sensitivity, previous work [2,20] has demonstrated that the Al composition must not exceed 23 percent, and the doping concentration must equal approximately $1.0 \times 10^{18}/\text{cm}^3$.

The data shown in Table 1, comparing two HEMTs with the same Al mole fraction (23 percent) but different doping concentrations in the n–AlGaAs layer, serve to illustrate the difference between low temperature and room-temperature device optimization. Device A has an n–AlGaAs doping concentration of $1.0 \times 10^{18}/\text{cm}^3$, while that of B is two times higher. As expected, at ambient, device B exhibited a higher g_m and associated gain than device A, with approximately the same noise figure for both devices. However, at a physical temperature of 13 K and 8.5 GHz, device B exhibited a minimum noise temperature of 13.1 K, while device A yielded a value of 5.3 K. These results illustrate that cryogenic measurements must be carried out in order to verify the device's noise performance.

VI. LNA Design Approach

A semiempirical method was utilized to obtain minimum noise temperature performance of the 0.25- by 75-μm devices [22]. A two-stage LNA fixture was used to perform the noise characterization of the devices at cryogenic temperatures. The two-stage LNA fixture was designed to achieve the best room temperature low-noise performance based on the measured room-temperature device parameters. Following construction and room-temperature optimization, the LNAs are biased for lowest noise performance at cryogenic temperatures. This approach was chosen for two important reasons. First, a cryogenic noise-parameter measurement system was still under development for frequencies from 26 to 40 GHz and was not available for these measurements. Second, since the device processing was uniform across the wafer and device gains varied from 5 to 8 dB, two cascaded identical devices were required to reduce the noise temperature contribution of the room-temperature noise measurement receiver.

The LNA fixture oxygen-free high-conductivity (OFHC) copper and dc bias circuits [23] are designed for operation at cryogenic temperatures. Diode protection was included in both the gate and drain bias circuits. Light-emitting diodes (LEDs) were mounted on the cover of the fixture above each of the HEMTs for the purpose

of examining their light sensitivity at cryogenic temperatures. The input and output ports utilize a broadband-WR28 to stepped-ridge-waveguide to microstrip transition. The dimensions of the transition are detailed in Fig. 7. Figure 8 shows the insertion loss and return loss of a stepped-ridge fixture that consists of two stepped-ridge transitions connected back-to-back with a half-inch long microstrip 50-ohm line. The input and output matching networks were designed based on the device equivalent circuit values obtained from fitting measured S-parameters at the low-noise bias condition to the model from 2 to 20 GHz at room temperature. The noise parameters are then calculated from the small signal circuit elements using the previously noted Fukui expressions. Input, output, and interstage matching circuits were designed on 10-mil quartz substrate with TaN thin film resistors and TiWAu metallization. A schematic diagram of the two-stage hybrid HEMT LNA is shown in Fig. 9. The edge-coupled symmetric microstrip dc-blocking transmission line also served as a bandpass filter, improving the out-of-band stability. The three- and four-stage LNAs are constructed from the two-stage by the insertion of additional interstage matching circuits.

VII. LNA and Radiometer Performance

The two- and three-stage LNAs were first measured at room temperature with the devices biased for lowest noise at room temperature and then biased for lowest noise performance at cryogenic temperatures. Both LNAs exhibited an average noise figure of approximately 2 dB from 28 to 36 GHz. From 29 to 34 GHz, the gains measured approximately 17 and 23 dB for the two-stage and three-stage LNAs, respectively. The addition of an external isolator only slightly (0.3 dB) degraded the gain and noise figure.

With the devices biased for lowest noise at cryogenic temperature (12 K), the noise temperature (referenced at the room-temperature input waveguide flange) of both LNAs was observed to decrease nearly quadratically as a function of physical temperature as they cooled from 300 to 12 K. The noise temperature of the two-stage LNA decreased from 350 K at ambient to 35 K at 14.5 K, while that of the three-stage LNA decreased from 400 to 41 K at 12.5 K (see Fig. 10). Figure 11 shows the cryogenic noise temperature and gain response from 31 to 33 GHz, along with bias settings for the three-stage LNA. At 32 GHz, the two-stage LNA noise temperature measured 35 K, with an associated gain of 16.5 dB, at a physical temperature of 14 K, while the three-stage LNA yielded a value of 41 K with 26.0-dB associated gain. It is also noted that the three-stage LNA displayed an almost flat noise temperature response across the measurement band with a minimum noise temperature of 39 K at 32 GHz, while the two-stage LNA displayed a noise temperature response monotonically decreasing from 31 to 33 GHz with a minimum noise temperature of 31 K at 33 GHz.

It was further observed that neither amplifier showed a persistent photoconductivity effect. That is, it was found that these devices can be cooled with or without illumination and/or dc bias, without any observable effect on the cryogenic low-noise performance.

Two 32-GHz cryogenic HEMT radiometers were developed employing four-stage LNAs based on the two- and three-stage LNA results. Figure 12 shows a picture of the four-stage LNA; its cryogenic gain and noise performance are plotted in Fig. 13. The first system, pictured in Fig. 14, is a total power radiometer that exhibited an input noise temperature referenced to the input room-temperature flange of 45 K. Its noise temperature and gain response as functions of frequency are plotted in Fig. 15. The other system, shown in Fig. 16 without its vacuum jacket and radiation shield, is a Dicke-switching radiometer which demonstrated an input noise temperature referenced to the horn aperture of 59 K. The noise temperature response as a function of frequency is plotted in Fig. 17.

VIII. Conclusion

Since the invention of the HEMT device in 1983, noise temperatures and device yields have steadily improved. Using devices with 0.25-μm gates produced by electron beam lithography, key parameters of the device structure, such as the thicknesses and doping semiconductor layers, have been systematically optimized for cryogenic operation. Figure 18 summarizes the evolution of performance improvement for both maser and HEMT amplifier systems over the past 25 years. When cooled to 15 K, the best device noise temperatures at 8.4 GHz were improved steadily from 8.5 K in 1985 to 5.3 K in 1987.

The results of the cryogenic coolable state-of-the-art 32-GHz HEMT LNAs and radiometers clearly demonstrate their potential to meet the future space science needs of the DSN. Currently, JPL is investigating 0.1-μm gate lengths and promising alternate structures (planar doped and single and double heterojunction) and materials (InGaAs and InP). Further advances in HEMT technology [24,25] promise to lead to improved performance at all frequencies and to make possible the development of amplifiers above 100 GHz.

References

[1] S. M. Petty and D. L. Trowbridge, *Low Noise Amplifiers in the Deep Space Network—A Radio Communications Instrument for Deep Space Exploration*, JPL Publication 82-104, Jet Propulsion Laboratory, Pasadena, California, pp. 4-1–4-32, July 1983.

[2] M. W. Pospieszalski, S. Weinreb, P. C. Chao, U. K. Mishra, S. C. Palmateer, P. M. Smith, and J. C. M. Hwang, "Noise Parameters and Light Sensitivity of Low-Noise High-Electron-Mobility Transistors," *IEEE Trans. Electron Devices*, vol. ED-33, pp. 218–223, 1986.

[3] S. Weinreb, M. W. Pospieszalski, and R. Norrod, "Cryogenic, HEMT, Low-Noise Receivers for 1.3 to 43 GHz Range," *1988 IEEE MTT-S Digest*, pp. 945–948, 1988.

[4] J. Shell and D. Neff, "A 32-GHz Reflected Wave Maser Amplifier with Wide Instantaneous Bandwidth," *1988 IEEE MTT-S Digest*, pp. 789–792, 1988.

[5] W. H. Higa and E. Wiebe, "One Million Hours at 4.5 Kelvin," *Proc. App. of Closed-Cycle Cryocoolers to Small Superconducting Devices*, Boulder, Colorado, pp. 99–108, October 1978.

[6] D. L. Johnson, S. M. Petty, J. J. Kovatch, and G. W. Glass, "Ultralow Noise Performance of an 8.4-GHz Maser-Feedhorn System," *The Telecommunications and Data Acquisition Progress Report 42-100, vol. October–December 1989*, Jet Propulsion Laboratory, Pasadena, California, pp. 100–110, February 15, 1990.

[7] IRE Subcommittee 7.9 on Noise, "Noise Representation of Noise in Linear Two Ports," *Proc. IRE*, vol. 48, pp. 69–74, January 1960.

[8] H. Rothe and W. Dahlke, "Theory of Noisy Fourpoles," *Proc. IRE*, vol. 44, pp. 811–818, June 1956.

[9] P. Penfield, "Wave Representation of Amplifier Noise," *IRE Circuit Theory*, vol. CT-9, pp. 84–86, March 1962.

[10] H. Fukui, "Design of Microwave GaAs MESFETs for Broadband, Low-Noise Amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-27, pp. 643–650, July 1979.

[11] A. van der Ziel, "Thermal Noise in Field-Effect Transistors," *Proc. IRE*, vol. 50, pp. 1808–1812, 1986.

[12] A. van der Ziel, "Gate Noise in Field-Effect Transistors at Moderately High Frequencies," *Proc. IRE*, vol. 51, pp. 461–467, 1963.

[13] A. Cappy, A. Vanoverschelde, M. Schortgen, C. Versnaeyen, and G. Salmer, "Noise Modeling Submicrometer-Gate, Two-Dimensional, Electron-Gas, Field-Effect Transistors," *IEEE Trans. Electron Devices*, vol. Ed-32, p. 2787, December 1985.

[14] M. Pospieszalski, "Modeling of Noise Parameters of MESFETs and MODFETs and Their Frequency and Temperature Dependence," *IEEE Trans. Microwave Theory Tech.*, vol. 37, no. 9, pp. 1340–1350, September 1989.

[15] K. Joshin, S. Asai, Y. Hirachi, and M. Abe, "Experimental and Theoretical Noise Analysis of Microwave HEMTs," *IEEE Trans. Electron Devices*, vol. 36, no. 10, pp. 2274–2280, October 1989.

[16] K. H. G. Duh, P. C. Chao, P. M. Smith, L. F. Lester, B. R. Lee, J. M. Ballingall, and M. Y. Kao, "Millimeter-Wave Low-Noise HEMT Amplifiers," *1988 IEEE MTT-S Digest*, pp. 923–926, 1988.

[17] P. M. Smith, P. C. Chao, K. H. G. Duh, L. F. Lester, B. R. Lee, J. M. Ballingall, and M. Y. Kao, "Advances in HEMT Technology and Applications," *1987 IEEE MTT-S Digest*, pp. 749–752, 1987.

[18] S. C. Palmateer, P. A. Maki, W. Katz, A. R. Calawa, J. C. M. Hwang, and L. F. Eastman, "The influence of V:III flux ratio on unintentional impurity incorporation during molecular beam epitaxial growth," *Proc. Gallium Arsenide and Related Compounds Conference 1984* (Inst. Phys. Conf. Series 74), pp. 217–222, 1985.

[19] P. C. Chao, P. M. Smith, S. C. Palmeteer, and J. C. M. Hwang, "Electron-Beam Fabrication of GaAs Low-Noise MESFETs Using a New Tri-Layer Resist Technique," *IEEE Trans. Electron Devices*, vol. ED-22, pp. 1042–1046, 1985.

[20] K. H. G. Duh, M. W. Pospieszalski, W. F. Kopp, P. Ho, A. A. Jabra, P. C. Chao, P. M. Smith, L. F. Lester, J. M. Ballingall, and S. Weinreb, "Ultra-Low-Noise Cryogenic High-Electron Mobility Transistors," *IEEE Trans. Electron Devices*, vol. ED-35, pp. 249–256, 1988.

[21] A. Kastarsky and R. A. Klein, "On the low temperature degradation of AlGaAs/GaAs modulation doped field-effect transistors," *IEEE Trans. Electron Devices*, vol. ED-33, pp. 414–423, March 1986.

[22] K. H. G. Duh, W. F. Kopp, P. Ho, P.-C. Chao, M.-Y. Ko, P. M. Smith, J. M. Ballingall, J. J. Bautista, and G. G. Ortiz, "32-GHz Cryogenically Cooled HEMT Low-Noise Amplifiers," *IEEE Trans. Electron Devices*, vol. ED-36, no. 8, pp. 1528–1535, 1989.

[23] S. Weinreb and R. Harris, *A 23-GHz Coolable FET Amplifier*, NRAO Internal Report, NRAO, Charlottesville, Virginia.

[24] P. C. Chao, P. M. Smith, K. H. G. Duh, J. M. Ballingall, L. F. Lester, B. R. Lee, A. A. Jabra, and R. C. Tiberio, "High Performance 0.1 Micron Gate-Length Planar-Doped HEMTs," *1987 IEDM*, paper 17.1, 1987.

[25] P. Ho, P. C Chao, K. H. G. Duh, A. A. Jabra, J. M. Ballingall, and P. M. Smith, "Extremely High Gain, Low Noise InAlAs/InGaAs HEMTs Grown by Molecular Beam Epitaxy," *1988 IEDM*, 1988.

Table 1. Performance comparison of conventional AlGaAs/GaAs HEMTs.

	Ambient temperature, 300 K				Ambient temperature, 13 K
Frequency, GHz	—	8	18	32	8.5
Performance (NF[a], dB/GA[b], dB)	gm, mS/mm	NF/GA	NF/GA	NF/GA	T[c], K/GA
Type A ($1 \times 10^{18}/cm^3$)	380	0.4/15.2	0.7/11.5	1.3/7.5	5.3/13.9
Type B ($2 \times 10^{18}/cm^3$)	450	—	0.7/15	1.2/10.0	13.1/14.5

[a]NF = noise figure
[b]GA = associated gain
[c]T = noise temperature

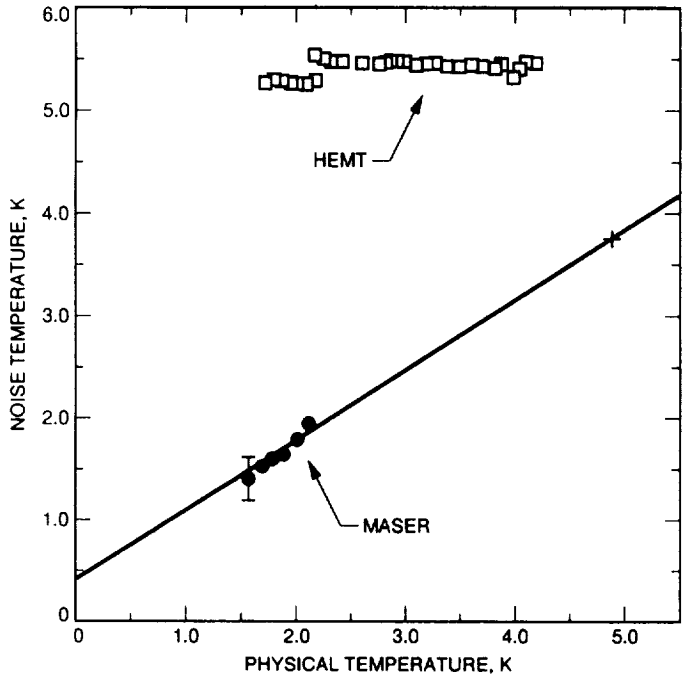


Fig. 1. Noise temperature of an 8.4-GHz ultralow noise maser and HEMT in liquid helium.

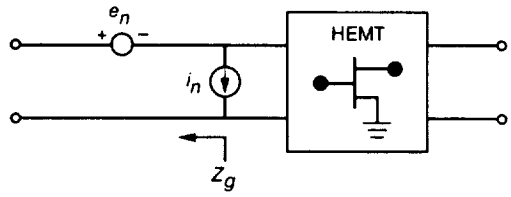


Fig. 3. Noise source representation of a noisy linear two-port circuit.

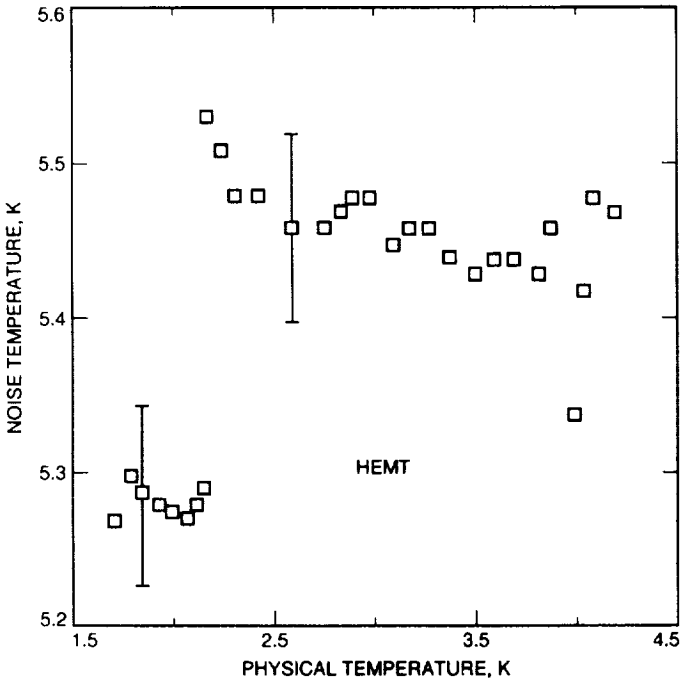


Fig. 2. Noise temperature of an 8.4-GHz HEMT LNA in liquid helium.

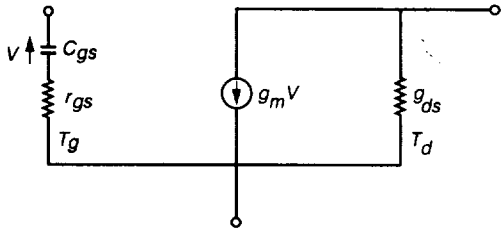


Fig. 4. Pospieszalski HEMT/FET noise model for an intrinsic chip.

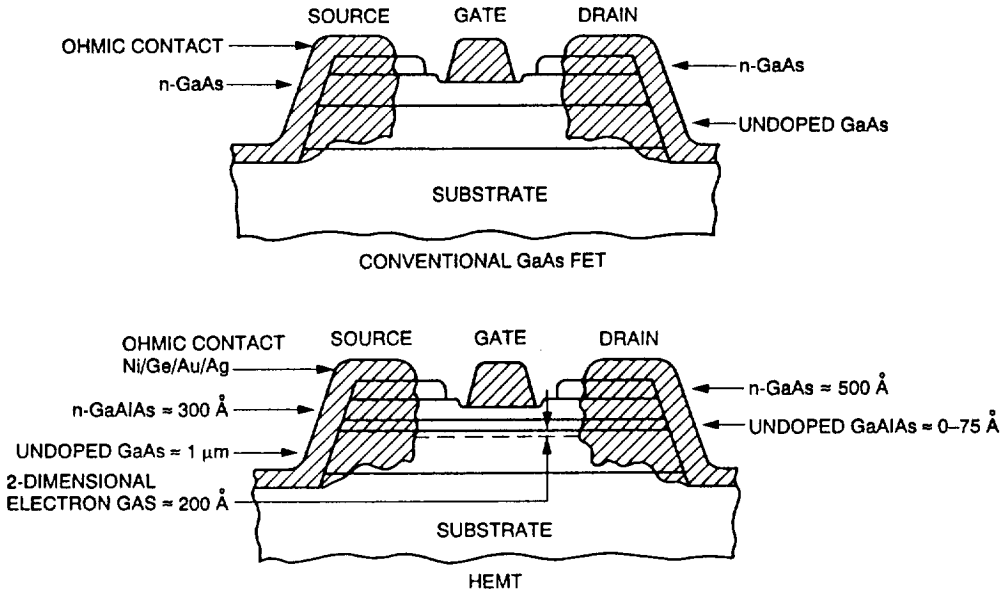


Fig. 5. Representation of MESFET and HEMT device cross-sections.

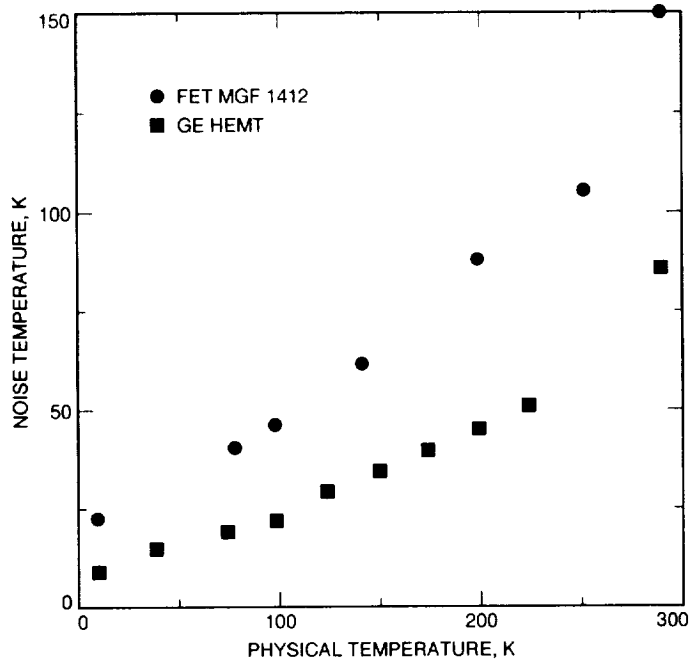


Fig. 6. MESFET versus HEMT noise temperature performance as a function of physical temperature at 8.4 GHz.

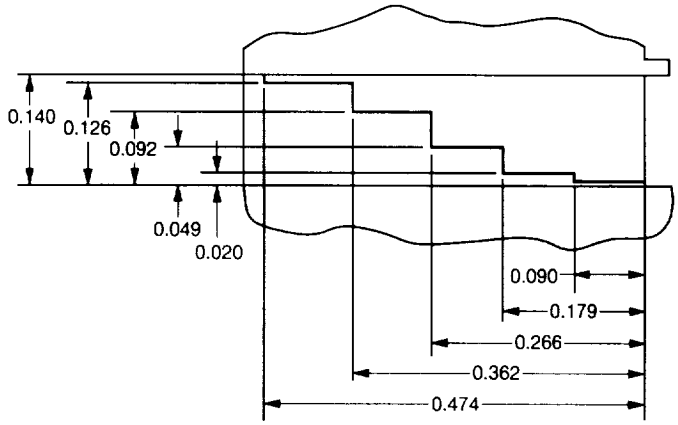


Fig. 7. Ka-band stepped ridge-to-microstrip transition, units in inches.

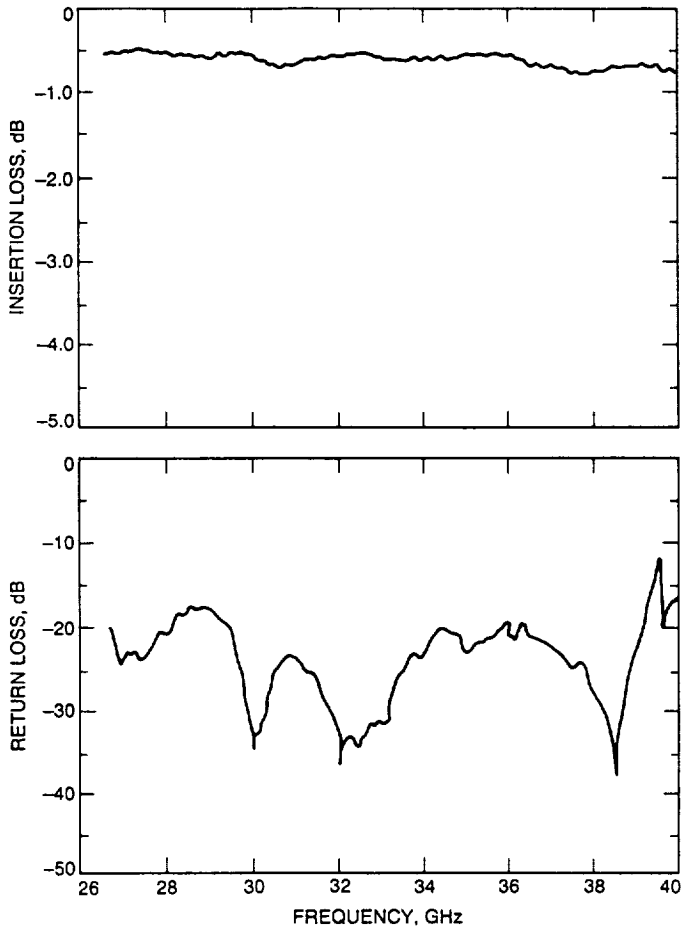


Fig. 8. Measured performance of Ka-band stepped-ridge fixture.

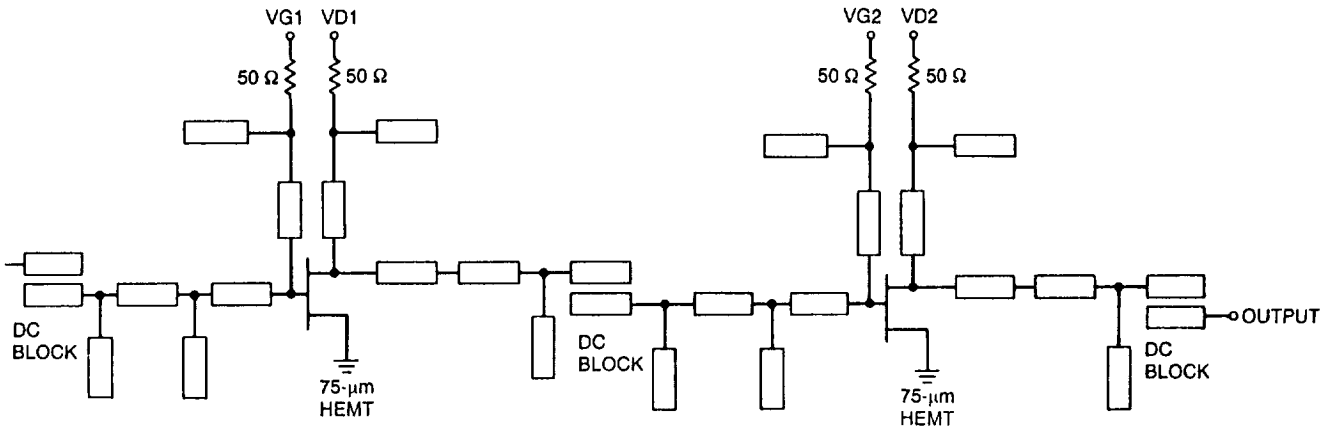


Fig. 9. Two-stage hybrid LNA using 0.25- by 75-μm HEMTs.

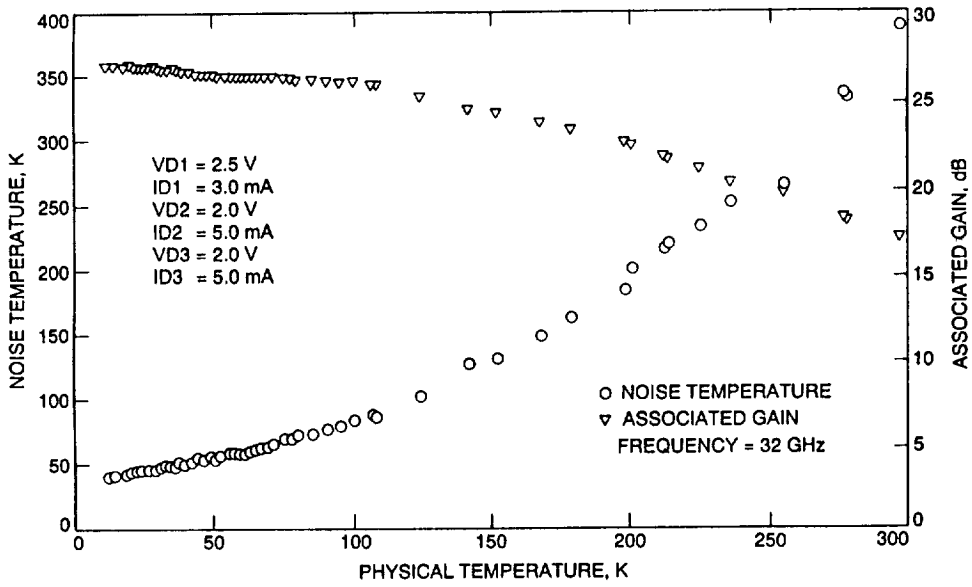


Fig. 10. Three-stage HEMT LNA noise temperature and gain cooling curve.

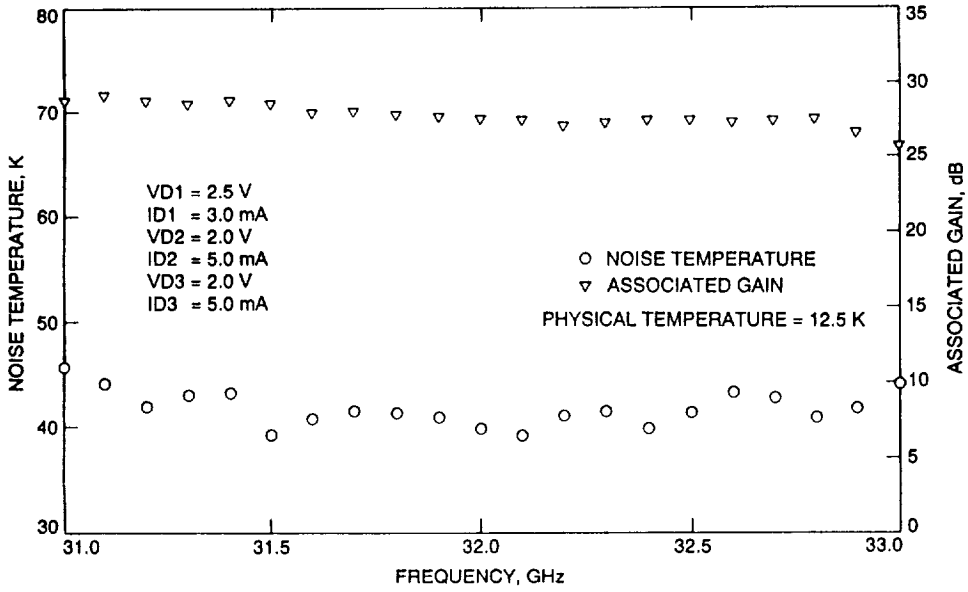


Fig. 11. Three-stage HEMT LNA 12.5-K noise temperature and gain.

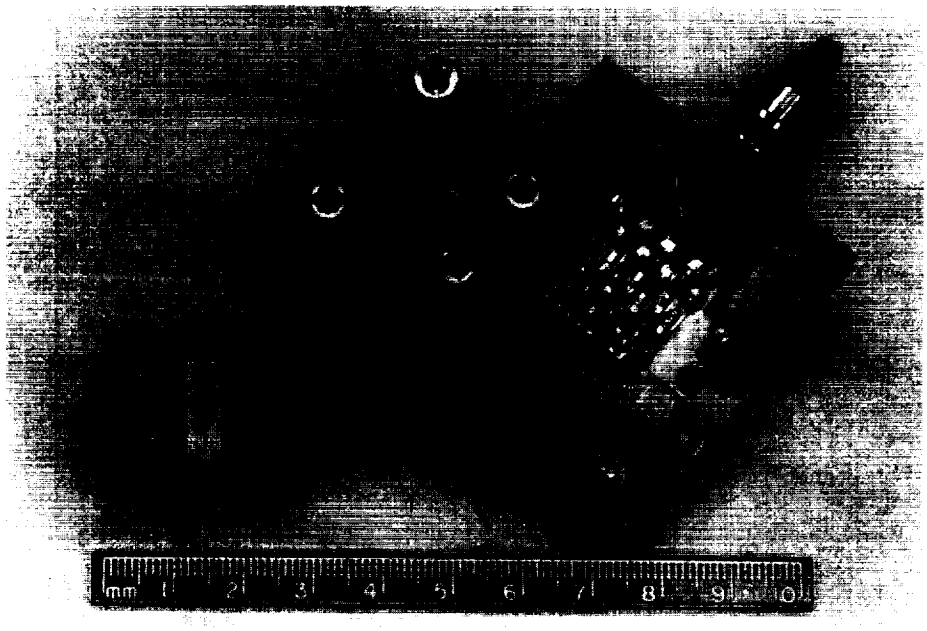


Fig. 12. Four-stage HEMT LNA.

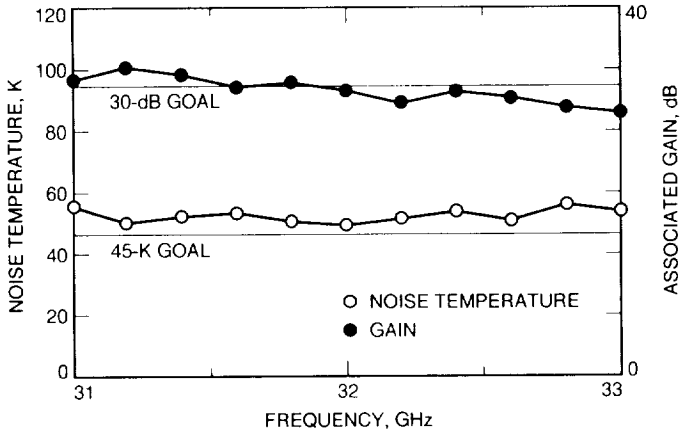


Fig. 13. Cryogenic noise versus gain response of four-stage HEMT LNA.

Fig. 14. A 32-GHz HEMT/CCR total power radiometer.

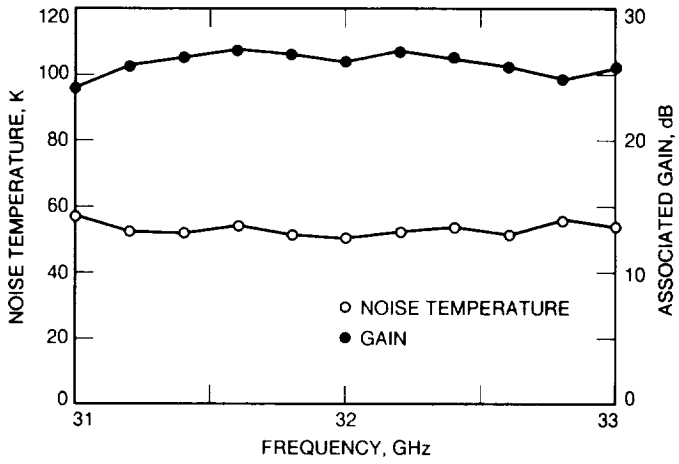


Fig. 15. Noise temperature and gain of 32-GHz HEMT/CCR total power radiometer.

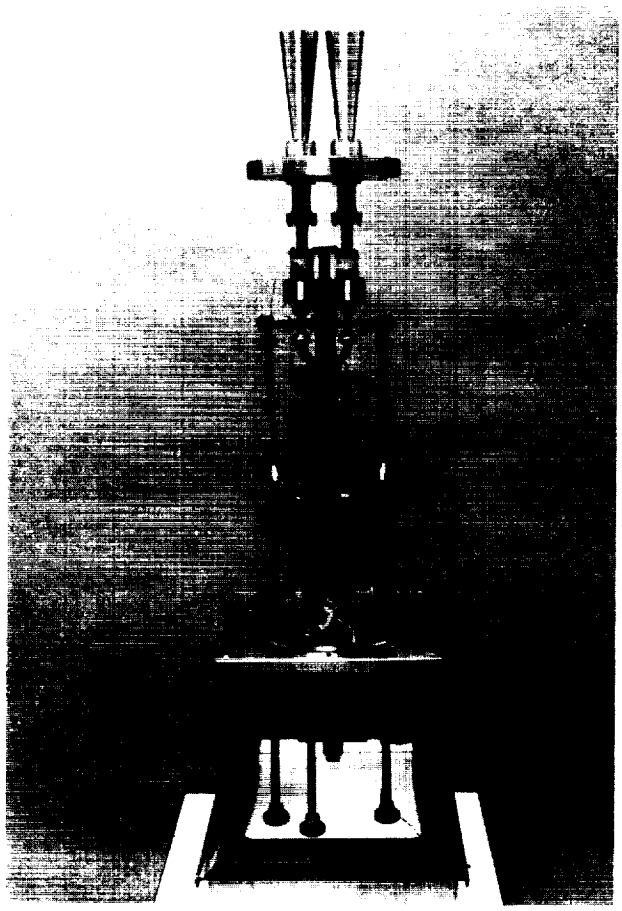


Fig. 16. A 32-GHz HEMT/CCR Dicke-switching radiometer without radiation shields and vacuum jacket.

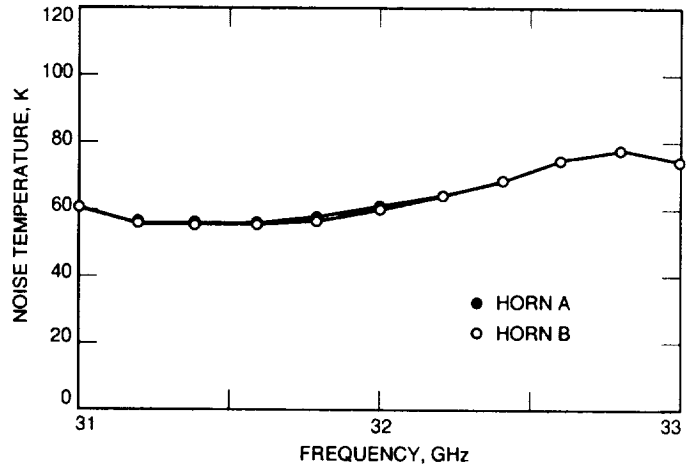


Fig. 17. Noise temperature at horn aperture of 32-GHz HEMT/CCR switching radiometer.

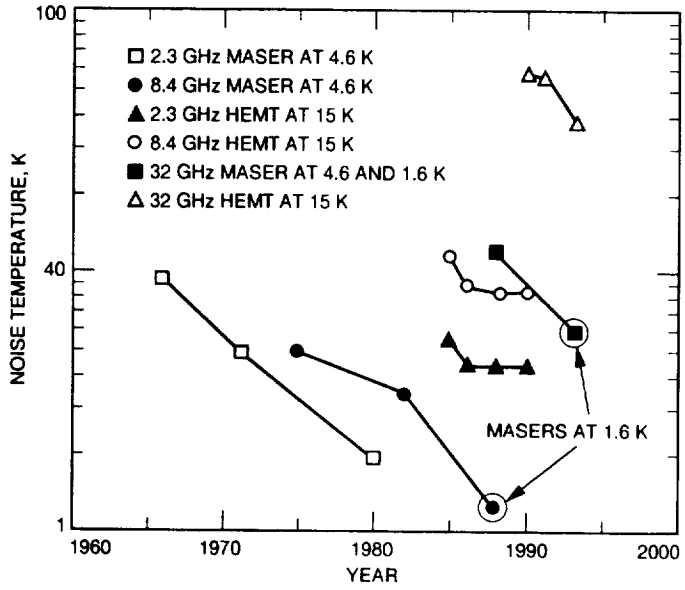


Fig. 18. Evolution of low noise performance of maser and HEMT amplifier systems.

Analysis of Open-Loop Conical Scan Pointing Error and Variance Estimators

L. S. Alvarez
Ground Antennas and Facilities Engineering Section

General pointing error and variance estimators for an open-loop conical scan (conscan) system are derived and analyzed. The conscan algorithm is modeled as a weighted least-squares estimator whose inputs are samples of receiver carrier power and its associated measurement uncertainty. When the assumptions of constant measurement noise and zero pointing error estimation are applied, the variance equation is then strictly a function of the carrier power-to-uncertainty ratio and the operator-selectable radius and period input to the algorithm. The performance equation is applied to a 34-m mirror-based beam-waveguide conscan system interfaced with the Block V Receiver Subsystem tracking a Ka-band (32-GHz) downlink. It is shown that for a carrier-to-noise power ratio ≥ 30 dB-Hz, the conscan period for Ka-band operation may be chosen well below the current DSN minimum of 32 sec. The analysis presented forms the basis of future conscan work in both research and development as well as for the upcoming DSN antenna controller upgrade for the new DSS-24 34-m beam-waveguide antenna.

I. Introduction

An analysis of open-loop conical scan (conscan) pointing error and variance estimators is presented. The analysis models conscan as a beam-pointing error sensor whose input consists of samples of receiver carrier power and uncertainty. This choice of input is consistent with the upcoming DSN upgrade of conscan that will involve the interfacing of the Antenna Pointing Controller and the Block V Receiver Subsystem. With this input, the conscan algorithm is modeled as a weighted least-squares estimator whose variance can be derived as a function of the uncertainty on the receiver input and the operator-selectable inputs to the algorithm. A general variance equation that is applicable to either conscan axis is derived and then simplified when assumptions of constant measurement uncertainty and zero pointing error are applied.

Estimation of the uncertainty on the carrier power samples from the Block V Receiver Subsystem is briefly reviewed, and the results are used to rewrite the conscan pointing-error-variance equation in terms of the carrier-to-noise power ratio. The final equation can then be used to easily quantify the pointing error uncertainty as a function of conscan radius and period for any given antenna's half-power beamwidth and receiver carrier-to-noise power operating condition. The article concludes with an application of the performance equation to a beam-waveguide mirror conscan implementation operating at the Ka-band (32-GHz) frequency.

II. Received Carrier Power Model For Conscan

The conical scan received signal model presented below is similar to the one described in [1]. The input to the conscan algorithm in this analysis will be estimates of received carrier power P_c with measurement uncertainty σ_{P_c}. The assumption here is that the digital receiver subsystem has derived both of these quantities from its estimate of signal-to-noise ratio P_c/N_o. The main results of this analysis will be general with respect to the conscan algorithm input and can be modified slightly to accommodate other received signal level input (e.g., voltages or noise temperatures).

The ratio of the received carrier power $P_{(t)}$ at time t to nominal carrier power P_{cnom} when the antenna boresight is pointing at the target is

$$\frac{P_c(t)}{P_{cnom}} = \exp\left(-\frac{\mu}{h^2}\varepsilon(t)^2\right) \tag{1}$$

where h is the antenna half-power beamwidth, $\mu = 4\ln(2)$, and $\varepsilon(t)$ is the angular displacement of the target to the center of the beam. During conscan, $\varepsilon(t)^2$ in [1] is shown to be

$$\varepsilon(t)^2 = r^2 + \varepsilon_s^2 - 2r\varepsilon_{xel}\cos(wt) - 2r\varepsilon_{el}\sin(wt) \tag{2}$$

where r is the conscan radius, w is the conscan frequency, and ε_{xel} and ε_{el} are cross-elevation and elevation components of the beam pointing error ε_s, defined as

$$\varepsilon_s^2 = \varepsilon_{xel}^2 + \varepsilon_{el}^2 \tag{3}$$

Let $L_1 = \exp\left((-\mu/h^2)r^2\right)$ be the loss factor when the target is at the center of the scan pattern and $L_2 = \exp\left((-\mu/h^2)\varepsilon_s^2\right)$ be the loss factor due to the beam-pointing error being estimated; then inserting Eq. (2) into Eq. (1) and simplifying yields

$$P_c(t) = P_{cnom}L_1L_2 \times \exp\left(\frac{2r\mu}{h^2}\left[\varepsilon_{xel}\cos(wt) + \varepsilon_{el}\sin(wt)\right]\right) \tag{4}$$

It can be shown that

$$P_o = P_{cnom}L_1L_2 \tag{5}$$

and

$$k_s = \frac{2r\mu}{h} \tag{6}$$

where P_o is the average carrier power received over the conical scan period and k_s is the conscan slope. Now for small target errors, the approximation $\exp(x) \approx 1 + x$ can be applied to Eq. (4), and then inserting Eqs. (5) and (6) gives the following:

$$P_c(t) = P_o\left(1 + \frac{k_s}{h}\varepsilon_{xel}\cos(wt) + \frac{k_s}{h}\varepsilon_{el}\sin(wt)\right) \tag{7}$$

or rewriting in vector notation,

$$P_c(t) = \begin{bmatrix} 1 & \cos(wt) & \sin(wt) \end{bmatrix}\begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} \tag{8}$$

where $C_1 = P_o$, $C_2 = ((k_s/h)\varepsilon_{xel})P_o$, and $C_3 = ((k_s/h)\varepsilon_{xel})P_o$. Equation (8) is the measurement equation used along with P_c and σ_{P_c} in the conscan pointing error estimator.

III. General Pointing Error and Variance Estimators

A. Pointing Error Estimator

The conscan estimation period can be any interval of time and, in general, does not have to be a complete period of the sinusoid in Eq. (8); nor does receiver power have to be measured uniformly over this interval. However, the present analysis will assume received signal power $P_c(t)$ is uniformly sampled n times over a conscan period of T sec with a receiver integration time of t_{rec} sec, where $T = nt_{rec}$. Substituting $t = it_{rec}$ and $w = 2\pi/nt_{rec}$ in Eq. (8) yields the conscan carrier power measurement equation at the sampling instants i as

$$P_c(i) = \begin{bmatrix} 1 & \cos\left(\frac{2\pi}{n}i\right) & \sin\left(\frac{2\pi}{n}i\right) \end{bmatrix}\begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} \tag{9}$$

Accumulating n input samples and storing them in matrix form gives

$$\begin{bmatrix} P_c(1) \\ P_c(2) \\ \vdots \\ P_c(n) \end{bmatrix} = \begin{bmatrix} 1 & \cos(\frac{2\pi}{n}) & \sin(\frac{2\pi}{n}) \\ 1 & \cos(2\frac{2\pi}{n}) & \sin(2\frac{2\pi}{n}) \\ & \vdots & \\ 1 & \cos(2\pi) & \sin(2\pi) \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} \quad (10)$$

or

$$\mathbf{Y} = \mathbf{AC} \quad (11)$$

where the measurement vector $\mathbf{Y}$ is $n \times 1$, the measurement distribution matrix $\mathbf{A}$ is $n \times 3$, and the parameter vector $\mathbf{C}$ is 3×1. It is assumed that the uncertainties on the carrier power estimates at the sampling times, denoted as $\sigma_{P_c}(i)$, are also known and available. Assuming that the uncertainties are random and independent over the period T, they are accumulated in a weighting matrix $\mathbf{R}$ as follows:

$$\mathbf{R} = \text{diag}\left(\frac{1}{\sigma^2_{P_c}(1)}, \frac{1}{\sigma^2_{P_c}(2)}, \cdots, \frac{1}{\sigma^2_{P_c}(n)}\right) \quad (12)$$

Given Eqs. (10) and (12), a weighted least-squares estimate $\hat{\mathbf{C}}$ is used for the estimator of $\mathbf{C}$ and is given by (see [2])

$$\hat{\mathbf{C}} = (\mathbf{A}^t\mathbf{RA})^{-1}\mathbf{A}^t\mathbf{RY} \quad (13)$$

where the superscript $\mathbf{t}$ is the transpose operator. With $\hat{\mathbf{C}}$ computed and the elements of $\mathbf{C}$ defined in the conscan received signal model, Eq. (8), the cross-elevation and elevation pointing error estimators are chosen as

$$\hat{\varepsilon}_{xel} = \frac{h\hat{\mathbf{C}}_2}{k_s\hat{\mathbf{C}}_1} \quad (14)$$

and

$$\hat{\varepsilon}_{el} = \frac{h\hat{\mathbf{C}}_3}{k_s\hat{\mathbf{C}}_1} \quad (15)$$

The general conscan algorithm also derives the direction of the error $\hat{\varepsilon}_s$ from the relative magnitudes of $\hat{\mathbf{C}}_2$ and $\hat{\mathbf{C}}_3$, together with an operator input phase term, which compensates for time delays in the antenna system. This will not be pursued here, for the present analysis will only focus on the accuracy of the magnitude of the open-loop pointing error estimates.

B. Pointing Error Variance Estimator

In [2] it is shown that the covariance matrix of the error in the estimate $\hat{\mathbf{C}}$ is given by

$$\mathbf{V} = (\mathbf{A}^t\mathbf{RA})^{-1} \quad (16)$$

where $\mathbf{A}$ is defined in Eq. (10) and $\mathbf{R}$ by Eq. (12). From Eqs. (14) and (15), the uncertainty in the estimates of $\hat{\varepsilon}_{xel}$ and $\hat{\varepsilon}_{xel}$ must be expressed in terms of the errors of the estimator $\hat{\mathbf{C}}$ (i.e., in terms of 3×3 error covariance $\mathbf{V}$). The calculations are carried out in Appendix A, and the results are presented below:

$$\sigma^2_{\hat{\varepsilon}_{xel}} = \left(\frac{h}{k_s}\right)^2\left(\frac{1}{P_o}\right)^2\left[\left(\frac{k_s}{h}\varepsilon_{xel}\right)^2\mathbf{V}(1,1) + \mathbf{V}(2,2) - 2\left(\frac{k_s}{h}\varepsilon_{xel}\right)\mathbf{V}(1,2)\right] \quad (17)$$

and

$$\sigma^2_{\hat{\varepsilon}_{el}} = \left(\frac{h}{k_s}\right)^2\left(\frac{1}{P_o}\right)^2\left[\left(\frac{k_s}{h}\varepsilon_{el}\right)^2\mathbf{V}(1,1) + \mathbf{V}(3,3) - 2\left(\frac{k_s}{h}\varepsilon_{el}\right)\mathbf{V}(1,3)\right] \quad (18)$$

As can be seen, the axial pointing error variance equations, Eqs. (17) and (18), for conscan are a function of many variables: the average received carrier power P_o, the uncertainties on the carrier power samples (embedded in $\mathbf{V}$), the antenna half-power beamwidth h, the conscan radius r and slope k_s, and the magnitude of the assumed static pointing error ε_s being estimated. Recalling that $\varepsilon_s^2 = \varepsilon_{xel}^2 + \varepsilon_{el}^2$ and $P_o = P_{cnom}L_1L_2$ and noting that the loss factor L_2 appears in both variance equations illustrates the axial cross-coupling of the algorithm (i.e., estimating a large pointing error in one axis will increase the estimation uncertainty in the other). The variance is also a function of the number of input samples (and hence conscan estimation period) that form the measurement distribution matrix $\mathbf{A}$, which in turn influences the covariance matrix $\mathbf{V}$ through Eq. (16). Equations (17) and (18) are, in fact, general for any number of receiver power samples measured uniformly or nonuniformly over the scan.

C-2

One important point to note is the dependence of the estimation performance on the operating frequency, which is inversely proportional to the antenna half-power beamwidth h. The radius r can be chosen so that the conscan slope remains constant with respect to operating frequency (e.g., in the DSN, r is typically selected to have a conscan loss of 0.1 dB from the factor L_1 which establishes $k_s = 0.5$ for all frequencies). With k_s constant, the variance can then be seen to be directly proportional to h, or inversely proportional to frequency, indicating a factor of four performance increase at Ka-band (32.0 GHz) over X-band (8.45-GHz) operation. This gain in performance is only realized provided the pointing error ε_s is small, for it will be shown later that loss from the factor L_2 increases dramatically with respect to ε_s at Ka-band.

IV. Analysis of the Conscan Pointing Error Variance Equation

A. Constant Measurement Error Assumption

Now, assuming the uncertainties $\sigma_{P_{c(i)}}$ on the carrier power samples are constant and equal to σ_{P_c} over the scan period, then the covariance matrix $\mathbf{V}$ defined by Eq. (16) reduces to

$$\mathbf{V} = \sigma_{P_c}^2 (\mathbf{A}^t \mathbf{A})^{-1} \tag{19}$$

This is a valid assumption under ideal, closed-loop conscan tracking conditions (benign wind, no spacecraft-induced variations on the downlink signal, etc.) and if the beam-pointing error ε_s is small ($\varepsilon_s \sim r$). Further simplification of $\mathbf{V}$ in Appendix A shows that when the carrier power input is sampled uniformly over the scan period, the pointing error variance Eqs. (17) and (18) each reduce to

$$\sigma_{\hat{\varepsilon}}^2 = \left(\frac{h}{k_s}\right)^2 \left(\frac{1}{n}\right) \left(\frac{\sigma_{P_c}}{P_o}\right)^2 \left(\left(\frac{k_s}{h}\varepsilon\right)^2 + 2\right) \tag{20}$$

where ε represents the actual pointing error in either the elevation or cross-elevation axis.

B. Effects of the Pointing Error Magnitude

1. Performance Degradation Due to Increasing Pointing Errors. Inserting $L_2 = \exp\left((-\mu/h^2)\varepsilon_s^2\right)$ into Eq. (20) and taking the square root gives the axial pointing error standard deviation as

$$\sigma_{\hat{\varepsilon}} = \left(\frac{h}{k_s}\right) \left(\frac{1}{n}\right)^{\frac{1}{2}} \left(\frac{\sigma_{P_c}}{P_{cnom} L_1}\right) \left(\frac{1}{\exp(-\frac{\mu}{h^2}\varepsilon_s^2)}\right) \times \left(\left(\frac{k_s}{h}\varepsilon\right)^2 + 2\right)^{\frac{1}{2}} \tag{21}$$

where ε_s is the beam-pointing error and ε is the axial-pointing error. By simple inspection, the last factor can be accurately approximated by

$$\left(\left(\frac{k_s}{h}\varepsilon\right)^2 + 2\right)^{\frac{1}{2}} \sim \sqrt{2} \tag{22}$$

when $\varepsilon \leq h$ and the conscan slope is chosen as $k_s = 0.5$ for the DSN application. For reference, the half-power beamwidths (assuming a Gaussian beam) for a 34-m antenna operating at the X- and Ka-band frequencies are approximately 65 and 17 mdeg, respectively. Thus, as the magnitude of the beam-pointing error $\varepsilon_s = (\varepsilon_{xel}^2 + \varepsilon_{el}^2)^{1/2}$ increases, both estimates $\hat{\varepsilon}_{xel}$ and $\hat{\varepsilon}_{xel}$ are degraded equally by the inverse of the pointing-error loss factor $\exp\left((-\mu/h^2)\varepsilon_s^2\right)$. This effect is obviously more dramatic at Ka-band due to the narrow beamwidth. Figure 1 illustrates the percentage of increase of the axial pointing error standard deviation $\hat{\varepsilon}$ against the magnitude of ε_s for a 34-m antenna operating at the X- and Ka-band frequencies. In addition to dropping received carrier power by 3 dB at Ka-band, a beam-pointing error of 8.5 mdeg (one-half of the full half-power beamwidth h) is seen in Fig. 1 to double the Ka-band conscan estimation uncertainty. Without considering actual receiver signal-to-noise operating conditions, this Ka-band performance degradation will primarily affect conscan during acquisition, when blind pointing errors may be large. For the closed-loop system, this degradation will effectively increase the beam-pointing error response time.

2. Zero Pointing Error Assumption. As noted before, the magnitude of beam-pointing error ε_s is typically very small during closed-loop conscan tracking. This fact can be used to further simplify Eq. (21) in order to quantify the performance of the conscan pointing error estimator as a function of the carrier power-to-uncertainty ratio and the selectable input variables. Assuming ε_s is essentially zero, then Eq. (21) reduces to

$$\sigma_{\hat{\varepsilon}} = \left(\frac{h}{k_s}\right) \left(\frac{2}{n}\right)^{\frac{1}{2}} \left(\frac{1}{L_1}\right) \left(\frac{P_{cnom}}{\sigma_{P_c}}\right)^{-1} \tag{23}$$

where it is recalled that P_{cnom} is the nominal carrier power received when the antenna is pointing directly at the target and the uncertainty σ_{P_c} is assumed to be known.

V. Carrier Power Uncertainty for the Block V Receiver Subsystem

Estimation of the uncertainties $\sigma_{P_c}(i)$ on the carrier power samples $P_c(i)$ in the Block V Receiver Subsystem is briefly summarized in Appendix B. In general, calculation and analysis of this statistic for spacecraft tracking in a DSN antenna environment is a complicated matter. For this reason, it will not be rigorously pursued here; for a more thorough analysis, the details of the Block V receiver calculation can be found in [3,4], Aung, et al.,[1] and Scheid.[2] Of more interest in the present analysis is the approximation of the expression for σ_{P_c} given in the references, so that Eq. (23) can be written in terms of the nominal receiver carrier-to-noise power ratio $CNR = P_{cnom}/N_o$ instead of P_{cnom}/σ_{P_c}. The quantity CNR is measured when the antenna is pointed directly at the spacecraft. The simplification is carried out in Appendix B with the following result

$$\sigma_{P_c} \approx (2P_{cnom}N_o)^{\frac{1}{2}} \tag{24}$$

which is a valid approximation for a 1-sec receiver estimation period when the loop tracking error is very small and $P_{cnom}/N_o < 40$ dB-Hz. Now, inserting this expression into the conscan pointing-error estimation Eq. (23) and simplifying yields

$$\sigma_{\hat{\epsilon}} = \left(\frac{h}{k_s}\right)\left(\frac{1}{n}\right)^{\frac{1}{2}}\left(\frac{2}{L_1}\right)(CNR^{-1})^{\frac{1}{2}} \tag{25}$$

Equation (25) is a very useful equation as it allows quick evaluation of the pointing error estimation accuracy as a function of the selectable conscan algorithm input (radius r and number of input samples n) for any given operating CNR and antenna half-power beamwidth h.

[1] M. Aung and S. Stephens, "Statistics of the P_c/N_o Estimator in the Block V Receiver," JPL Interoffice Memorandum 3338-92-089 (internal document), Jet Propulsion Laboratory, Pasadena, California, April 29, 1992.

[2] R. E. Scheid, "Statistical Analysis of Antenna Carrier Power," JPL Interoffice Memorandum 343-92-1291 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 9, 1992.

VI. Application to a 34-m Antenna Beam-Waveguide Mirror Conscan System at Ka-Band

All of the conscan equations thus far have been general, but they will now be applied to a 34-m beam-waveguide mirror-based conscan system operating at Ka-band. For this scenario, $h = 17$ mdeg and the radius $r = 1.55$ mdeg for a scan loss of 0.1 dB ($L_1 = 0.977$) and $k_s = 0.5$. These values are inserted into Eq. (25), and the axial pointing error standard deviation $\sigma_{\hat{\epsilon}}$ is then plotted in Fig. 2 as a function of conscan period for various CNR. Also plotted is the line corresponding to the magnitude of the chosen radius r. In Fig. 2, it is assumed that the receiver integration time per carrier power sample is 1 sec and the conscan period is then just equal to n. This performance plot implies that for $CNR \geq 30$ dB-Hz, the conscan estimation period for Ka-band operation may be chosen well below the current DSN minimum of 32 sec and still maintain estimation accuracy of less than 1.55 mdeg.

Advantages of scanning a beam-waveguide mirror instead of the entire antenna dish structure include the ease of obtaining precision pointing of a drastically smaller and stiffer mirror gimbal assembly and also the higher bandwidths achievable by the small-scale axis servos. A conceptual sketch of such a system is presented in Fig. 3, in which it is proposed that either the first or last beam waveguide mirror be actuated in such a conscan scheme. Because high-rate, accurate mirror tracking is available, it was assumed in the previous Ka-band performance plot that the move times between measurement points over the scan are negligible. In fact, with an efficient interface between the conscan computer and the receiver subsystem, it is conceivable that the mirror-based conscan system can actually achieve pointing correction update rates as ambitious as those shown in Fig. 2.

VII. Summary and Future Work

General pointing error and variance estimators for conscan have been derived in order to characterize the estimated performance in terms of the operator-selectable input to the algorithm and carrier-to-noise ratio. After assuming constant measurement noise on the carrier power inputs in the variance equation, it was shown that the magnitude of the beam-pointing error being estimated degrades performance in each axis of the estimator. The effect is especially dramatic when conscaning at the Ka-band frequency due to the the narrow antenna half-power beamwidths. For the closed-loop conscan tracking application, the zero pointing error assumption was applied in order to express the pointing error estimation accuracy as

a function of the selectable algorithm input (radius and number of input samples) for any given operating carrier power to the uncertainty ratio and antenna half-power beamwidth. The performance equation was then applied to a 34-m beam-waveguide, mirror-based conscan system interfaced with the Block V Receiver Subsystem tracking a Ka-band downlink. Simulation of conscan pointing error uncertainty against conscan estimation period showed that for a carrier-to-noise power $\geq$ 30 dB-Hz, the period for Ka-band operation may be chosen well below the current DSN minimum of 32 sec and still maintain estimation accuracy of less than 1.55 mdeg.

The analysis presented forms the basis of future conscan work in both research and development as well as for the upcoming DSN antenna controller upgrade for the new DSS-24, 34-m beam-waveguide antenna. The conscan model and performance equations derived will be used in designing advanced tracking algorithms and generating predictions for experimentation on the new beam-waveguide mirror conscan system currently being implemented at the DSS-13 antenna. These equations will also be utilized in the DSN conscan upgrade, which will use an automatic algorithm parameter selection as a function of signal-to-noise input ratio. Lastly, an augmented analysis that integrates the effects of spacecraft spin (as in [5]) and dynamic wind loading on the antenna structure needs to be pursued in order to more precisely simulate open- and closed-loop conscan performance at the Ka-band frequency.

Acknowledgment

The author wishes to thank Mike Thorburn for his helpful technical discussions and encouragement.

References

[1] G. Biernson, *Optimal Radar Tracking Systems*, New York, New York: John Wiley and Sons, Inc., 1990.

[2] A. E. Bryson and Y. Ho, *Applied Optimal Control*, Waltham, Massachusetts: Blaisdell Publishing Company, 1969.

[3] A. M. Monk, "Carrier-to-Noise Power Estimation for the Block V Receiver," *The Telecommunications and Data Acquisition Progress Report 42-106, vol. April-June 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 353–363, August 15, 1991.

[4] T. J. Brunzie, "The Parkes Front-End Controller and Noise-Adding Radiometer," *The Telecommunications and Data Acquisition Progress Report 42-102, vol. April-June 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 119–137, August 15, 1990.

[5] A. Mileant, and T. Peng, "Pointing a Ground Antenna at a Spinning Spacecraft Using Conscan-Simulation Results," *The Telecommunications and Data Acquisition Progress Report 42-95, vol. July-September 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 246–260, November 15, 1988.

Appendix A

Simplification of the Conscan Pointing Error Estimation Variance

Assuming that the uncertainties $\sigma_{P_c}(i)$ on the ith carrier power sample are random and independent, then the covariance matrix of the error in the weighted least squares estimator $\hat{\mathbf{C}}$ of Eq. (13) is given by

$$\mathbf{V} = (\mathbf{A}^t\mathbf{R}\mathbf{A})^{-1} \tag{A-1}$$

where the measurement distribution matrix $\mathbf{A}$ is

$$\mathbf{A} = \begin{bmatrix} 1 & \cos(\frac{2\pi}{n}) & \sin(\frac{2\pi}{n}) \\ 1 & \cos(2\frac{2\pi}{n}) & \sin(2\frac{2\pi}{n}) \\ & \vdots & \\ 1 & \cos(2\pi) & \sin(2\pi) \end{bmatrix} \tag{A-2}$$

and the weighting matrix $\mathbf{R}$ is

$$\mathbf{R} = \text{diag}\left(\frac{1}{\sigma^2_{P_c}(1)}, \frac{1}{\sigma^2_{P_c}(2)}, \cdots, \frac{1}{\sigma^2_{P_c}(n)}\right) \tag{A-3}$$

Given the above expressions, Eq. (A-1) can be expanded as follows

$$\mathbf{V} = \begin{bmatrix} \sum \frac{1}{\sigma^2_{P_c(i)}} & \sum \frac{1}{\sigma^2_{P_c(i)}} \cos(\frac{2\pi}{n}i) & \sum \frac{1}{\sigma^2_{P_c(i)}} \sin(\frac{2\pi}{n}i) \\ \sum \frac{1}{\sigma^2_{P_c(i)}} \cos(\frac{2\pi}{n}i) & \sum \frac{1}{\sigma^2_{P_c(i)}} \cos^2(\frac{2\pi}{n}i) & \sum \frac{1}{\sigma^2_{P_c(i)}} \cos(\frac{2\pi}{n}i)\sin(\frac{2\pi}{n}i) \\ \sum \frac{1}{\sigma^2_{P_c(i)}} \sin(\frac{2\pi}{n}i) & \sum \frac{1}{\sigma^2_{P_c(i)}} \cos(\frac{2\pi}{n}i)\sin(\frac{2\pi}{n}i) & \sum \frac{1}{\sigma^2_{P_c(i)}} \sin^2(\frac{2\pi}{n}i) \end{bmatrix}^{-1} \tag{A-4}$$

where the summations run over $i = 1, 2, \cdots, n$, and n is the total number of samples taken over the scan. If the measurement uncertainties are assumed to be constant over this period and equal to σ_{P_c}, then Eq. (A-4) simplifies to

$$\mathbf{V} = \sigma^2_{P_c} \begin{bmatrix} n & \sum \cos(\frac{2\pi}{n}i) & \sum \sin(\frac{2\pi}{n}i) \\ \sum \cos(\frac{2\pi}{n}i) & \sum \cos^2(\frac{2\pi}{n}i) & \sum \cos(\frac{2\pi}{n}i)\sin(\frac{2\pi}{n}i) \\ \sum \sin(\frac{2\pi}{n}i) & \sum \cos(\frac{2\pi}{n}i)\sin(\frac{2\pi}{n}i) & \sum \sin^2(\frac{2\pi}{n}i) \end{bmatrix}^{-1} \tag{A-5}$$

Now if the samples $P_c(i)$ are measured uniformly over the scan period as indicated above and $n \geq 3$, then Eq. (A-5) simplifies to

$$\mathbf{V} = \sigma^2_{P_c} \begin{bmatrix} \frac{1}{n} & 0 & 0 \\ 0 & \frac{2}{n} & 0 \\ 0 & 0 & \frac{2}{n} \end{bmatrix} \tag{A-6}$$

The conscan pointing-error estimation variances are derived from the following pointing-error estimator equations

$$\hat{\varepsilon}_{xel} = \frac{h\hat{\mathbf{C}}_2}{k_s\hat{\mathbf{C}}_1} \tag{A-7}$$

$$\hat{\varepsilon}_{el} = \frac{h\hat{\mathbf{C}}_3}{k_s\hat{\mathbf{C}}_1} \tag{A-8}$$

To find the variances of these estimates (each of which is a function of two random variables from $\hat{\mathbf{C}}$), the following formula is applied:

$$VAR(\hat{\varepsilon}) = \left(\frac{\partial\hat{\varepsilon}}{\partial\hat{\mathbf{C}}_1}\right)^2 VAR(\hat{\mathbf{C}}_1) + \left(\frac{\partial\hat{\varepsilon}}{\partial\hat{\mathbf{C}}_j}\right)^2 VAR(\hat{\mathbf{C}}_j) + 2\frac{\partial\hat{\varepsilon}}{\partial\hat{\mathbf{C}}_1}\frac{\partial\hat{\varepsilon}}{\partial\hat{\mathbf{C}}_j}VAR(\hat{\mathbf{C}}_1, \hat{\mathbf{C}}_j) \tag{A-9}$$

for $j = 2, 3$ and where the partial derivatives are evaluated at the statistical averages of the estimators $\hat{\mathbf{C}}_i, i = 1, 2, 3$. It can easily be proved that the statistical average of $\hat{\mathbf{C}}$ is just the vector $\mathbf{C}$, whose elements are defined by the conscan received signal model given in Eq. (8). Thus, the partial derivatives above are to be evaluated at $\mathbf{C}_1 = P_o$, $\mathbf{C}_2 = ((k_s/h)\varepsilon_{xel})\,P_o$, and $\mathbf{C}_3 = ((k_s/h)\varepsilon_{xel})\,P_o$. Applying the formula in Eq. (A-9) to Eqs. (A-7) and (A-8) yields

$$\sigma^2_{\hat{\varepsilon}_{xel}} = \left(\frac{h}{k_s}\right)^2\left(\frac{\mathbf{C}_2^2}{\mathbf{C}_1^4}\mathbf{V}(1,1) + \frac{1}{\mathbf{C}_1^2}\mathbf{V}(2,2) - 2\frac{\mathbf{C}_2}{\mathbf{C}_1^3}\mathbf{V}(1,2)\right) \tag{A-10}$$

and

$$\sigma^2_{\hat{\varepsilon}_{el}} = \left(\frac{h}{k_s}\right)^2\left(\frac{\mathbf{C}_3^2}{\mathbf{C}_1^4}\mathbf{V}(1,1) + \frac{1}{\mathbf{C}_1^2}\mathbf{V}(3,3) - 2\frac{\mathbf{C}_3}{\mathbf{C}_1^3}\mathbf{V}(1,3)\right) \tag{A-11}$$

Next, inserting the given expressions for the elements of $\mathbf{C}$ and using $\mathbf{V}$ defined by Eq. (A-6) for the constant measurement error case, the above variance equations can be rewritten as

$$\sigma^2_{\hat{\varepsilon}_{xel}} = \left(\frac{h}{k_s}\right)^2\left(\left(\frac{k_s}{hP_o}\varepsilon_{xel}\right)^2\frac{\sigma^2_{P_c}}{n} + \frac{2\sigma^2_{P_c}}{nP_o^2}\right) \tag{A-12}$$

and

$$\sigma^2_{\hat{\varepsilon}_{el}} = \left(\frac{h}{k_s}\right)^2\left(\left(\frac{k_s}{hP_o}\varepsilon_{el}\right)^2\frac{\sigma^2_{P_c}}{n} + \frac{2\sigma^2_{P_c}}{nP_o^2}\right) \tag{A-13}$$

Finally, after some rearranging, Eqs. (A-12) and (A-13) can be rewritten as

$$\sigma^2_{\hat{\varepsilon}} = \left(\frac{h}{k_s}\right)^2\left(\frac{1}{n}\right)\left(\frac{\sigma_{P_c}}{P_o}\right)^2\left(\left(\frac{k_s}{h}\varepsilon\right)^2 + 2\right) \tag{A-14}$$

where ε represents the actual pointing error in either the elevation or cross-elevation axis.

Appendix B

Received Carrier Power Uncertainty for the Block V Receiver

Inputs to the conical scan algorithm are estimates of received carrier power P_c and its uncertainty σ_{P_c}. A brief summary of the Block V receiver variance estimator is given below; however, a more detailed derivation may be obtained from [3] and footnotes 1 and 2. The receiver calculates σ_{P_c} from its estimates of signal-to-noise ratio (P_c/N_o) and variance $\sigma^2_{(P_c/N_o)}$, and system noise power N_o and variance $\sigma^2_{N_o}$. The system noise power over a 1-Hz bandwidth is $N_o = \kappa T_{sys}$, where T_{sys} is the system noise temperature with standard deviation $\sigma_{T_{sys}}$ (defined in [4]) and κ is Boltzmann Constant. The estimated noise power variance is then given by $\sigma^2_{N_o} = \kappa^2 \sigma^2_{T_{sys}}$. Multiplication negates the noise power from the received carrier signal power as follows:

$$P_c = \left(\frac{P_c}{N_o}\right) N_o \tag{B-1}$$

Now, assuming that the above equation is the product of two independent random variables, the carrier power variance can then be shown to be

$$\sigma^2_{P_c} = \sigma^2_{N_o} \left(\frac{\overline{P_c}}{N_o}\right)^2 + \sigma^2_{\left(\frac{P_c}{N_o}\right)} \overline{N}_o^{\,2} + \sigma^2_{N_o} \sigma^2_{\left(\frac{P_c}{N_o}\right)} \tag{B-2}$$

where the overbar denotes average values. Equation (B-2) gives the uncertainty of the received carrier power P_c in terms of the uncertainties of the estimates of (P_c/N_o) and (N_o). As described in [3] and footnote 1, the statistics of (P_c/N_o) are a function of receiver parameters (i.e., P_c/N_o, tracking loop error, tracking loop bandwidth, estimation interval, etc.) while the statistics of N_o are given as a function of T_{sys}, noise diode temperature, noise bandwidth, and estimation interval.

For this article, it is best to express the uncertainty on the carrier power in terms of the receiver carrier power P_c and the system noise power N_o. In [3], in-phase arm (I-arm) and in-phase/quadrature-phase arm (IQ-arm) P_c/N_o algorithms are presented. Of the two, a slightly modified version of the I-arm estimator will be implemented in the Block V receiver (see footnote 1). The equation for the variance of this estimator given in footnote 1 can be approximated by

$$\sigma^2_{\left(\frac{P_c}{N_o}\right)} \approx 2\left(\frac{\overline{P_c}}{N_o}\right) \tag{B-3}$$

assuming a 1-sec receiver estimation period when the loop tracking error ϕ is small enough so that $\cos(\phi) \approx 1$. Numerical simulation of Eq. (B-2) shows that for $P_c/N_o < 40$ db-Hz the contribution of the noise power variance σ_{N_o} in the calculation of σ_{P_c} is minimal and can be neglected. From Eq. (B-3), the carrier power uncertainty can then be reduced to

$$\sigma_{P_c} \approx \sigma_{\left(\frac{P_c}{N_o}\right)} \overline{N}_o \tag{B-4}$$

or inserting $\sigma_{(P_c/N_o)}$ from Eq. (B-3) and simplifying yields

$$\sigma_{P_c} \approx \left(2\overline{P}_c \overline{N}_o\right)^{\frac{1}{2}} \tag{B-5}$$

which is the desired approximation.

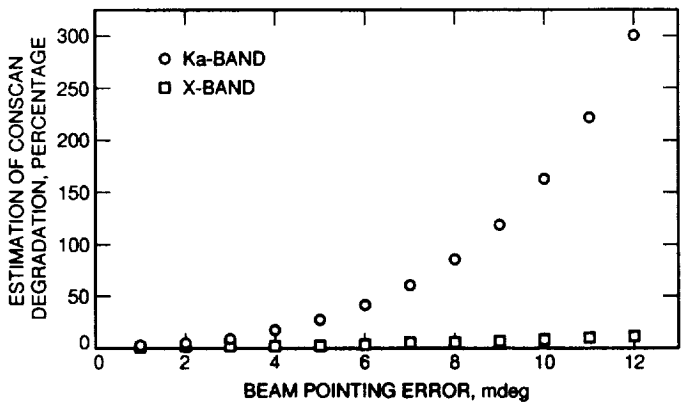


Fig. 1. Conscan estimation performance degradation due to increasing pointing error for a 34-m antenna.

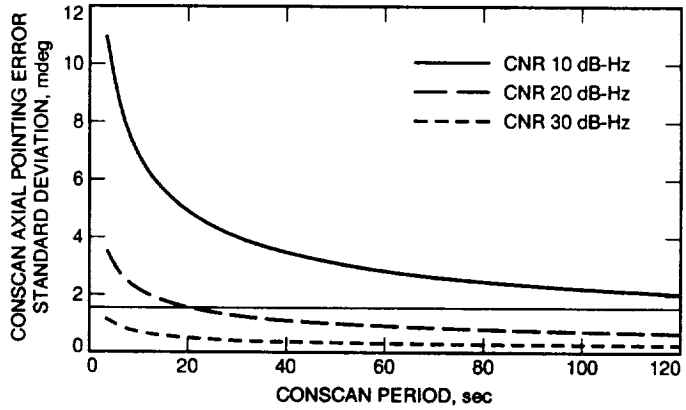


Fig. 2. Open-loop conscan axial pointing-error estimation accuracy for a 34-m antenna operating at Ka-band.

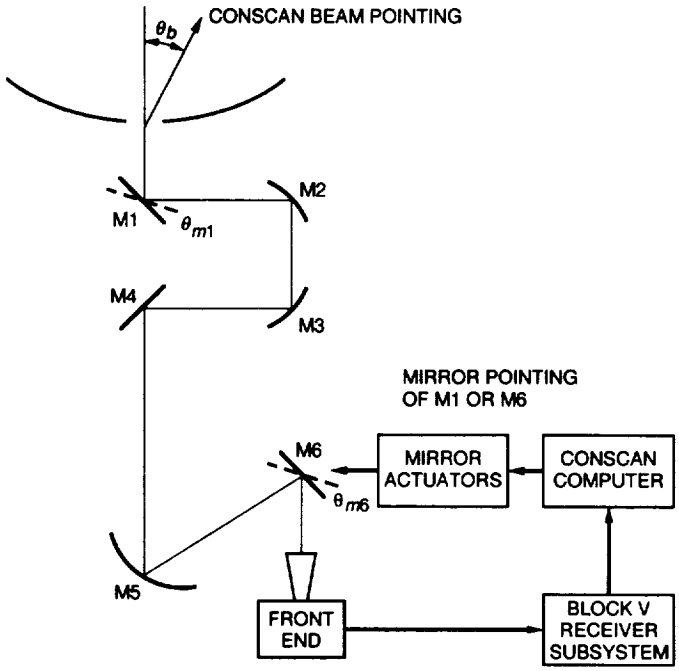


Fig. 3. Conceptual diagram of a 34-m beam-waveguide antenna mirror-based conscan system.

Spur-Reduced Digital Sinusoid Synthesis

M. J. Flanagan and G. A. Zimmerman
Communications Systems Research Section

This article presents and analyzes a technique for reducing the spurious signal content in digital sinusoid synthesis. Spurious-harmonic (spur) reduction is accomplished through dithering both amplitude and phase values prior to word-length reduction. The analytical approach developed for analog quantization is used to produce new bounds on spur performance in these dithered systems. Amplitude dithering allows output word-length reduction without introducing additional spurs. Effects of periodic dither similar to those produced by a pseudonoise (PN) generator are analyzed. This phase-dithering method provides a spur reduction of 6(M+1) dB per phase bit when the dither consists of M uniform variates. While the spur reduction is at the expense of an increase in system noise, the noise power can be made white, making the power spectral density small. This technique permits the use of a smaller number of phase bits addressing sinusoid lookup tables, resulting in an exponential decrease in system complexity. Amplitude dithering allows the use of less complicated multipliers and narrower data paths in purely digital applications, as well as the use of coarse-resolution, highly linear digital-to-analog converters (DACs) to obtain spur performance limited by the DAC linearity rather than its resolution.

I. Introduction

It is well known that adding a dither signal to a desired signal prior to quantization can render the quantizer error independent of the desired signal [1,2,3]. Classic examples of this deal with the quantization of analog signals. Advances in digital signal processing speed and large-scale integration have led to the development of all-digital receiver systems, direct digital frequency synthesizers, and direct digital arbitrary waveform synthesizers. Since finite-word-length effects are a major factor in system complexity, in all these applications, these effects may ultimately determine whether it is efficient to digitally implement a system with a particular set of specifications. Earlier work [4] has presented a technique for reducing the complexity of digital oscillators through phase dithering, with the claim of increased frequency resolution. Recent research [5] has suggested mitigation of finite-word-length effects in the synthesis of oversampled sinusoids through noise shaping. This article shows how the analysis techniques used for quantization of analog signals can be applied to overcome finite-word-length effects in digital systems. The analysis in this article shows how appropriate dither signals can be used to reduce word lengths in digital sinusoid synthesis without suffering the normal penalties in spurious signal performance. Furthermore, the dithering technique

presented in this article is not limited to the synthesis of oversampled signals.

Conventional methods of digital sinusoid generation [6], e.g., those in Fig. 1, result in spurious harmonics (spurs) that are caused by finite-word-length representations of both amplitude and phase samples [7]. Because both the phase and amplitude samples are periodic sequences, their finite-word-length representations contain periodic error sequences, which cause spurs. The spur signal levels are approximately 6 dB per bit of representation below the desired sinusoidal signal.

The technique presented in this article reduces the representation word length without increasing spur magnitudes, by first adding a low-level random noise, or dither, signal to the amplitude and/or the phase samples, which are originally expressed in a longer word length. The resulting sum, a dithered phase or amplitude value, is truncated or rounded to the smaller, desired word length. Of course, either the amplitude or the phase or both can be dithered. In phase dithering, the spurious response is determined by the type of dithering signal employed. In amplitude dithering, the spurious response is determined by the original, longer word length. While the amplitude-related spurious response is generally related to the phase-related spurious response, we will make the predither amplitude word length long enough to satisfy spur power specifications. Then the exact relationship is unimportant, and since the phase dither signal is independent of the amplitude dither signal, the amplitude and phase dithering processes can be treated independently.

The next section describes the quantizer model. Amplitude and phase quantization effects are reviewed in Sections III and IV, and simple new bounds on spurious performance are presented. In contrast to bounds in the existing literature, the new bounds are straightforward and require little information about the signal to be quantized. The derivations of the new bounds provide motivation for the new analysis of dithered quantizer performance that occurs later in this article. An analysis of dithering with a periodic noise source is presented in Section VI. The periodic noise source is considered because of its similarity to implementations involving linear feedback shift registers (LFSRs) or pseudonoise (PN) generators. A new analysis of phase dithering effects is presented in Sections VII and VIII, followed by simulation results and a design example.

II. Quantizer Model

When a discrete-time input signal, $x[n]$, is passed through a uniform midtread quantizer [8], the output signal, $y[n]$, can always be expressed as $y[n] = x[n] + e[n]$, where $e[n]$ is the quantization error, a deterministic function of $x[n]$. The input to the quantizer is mapped to 1 of 2^b levels, where b is the number of bits that digitally represent the input sample. Output levels are separated by one quantizer step size, $\Delta = 2^{-b}$. Throughout this article, Δ_A will be used as the step size for amplitude quantization results; Δ_P will be used for phase quantization results; and Δ will be used if the result applies to both amplitude and phase quantization. Similar subscripting will be used on the quantization error.

The input/output relation of a midtread quantizer appears in Fig. 2. If the input does not saturate the quantizer, then the quantizer error is [8]:

$$e[n] = \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} (-1)^k \frac{\Delta}{j2\pi k} \exp\left(\frac{j2\pi k x[n]}{\Delta}\right) \qquad (1)$$

If the input signal is bounded so that $|x[n]| \leq A_Q$ where $A_Q = 1/2 - \Delta$, then the quantizer does not saturate and $|e[n]| \leq \Delta/2$. Throughout this article, quantizers are always operating in nonsaturation mode.

III. Amplitude Quantization Effects

Let a discrete-time sinusoid with amplitude $A \leq A_Q$ and frequency ω_0 be the input to a midtread quantizer. If the sinusoid is generated in a synchronous discrete-time system, ω_0 can be expressed as 2π times the ratio of two integers. The input sequence is then periodic with a finite period. Since the error sequence, $e_A[n]$, is a deterministic function of the input sequence, it is periodic with a finite period as well. Therefore, the spectrum of the error sequence will consist of discrete frequency components (spurs) that contaminate the spectrum of $x[n]$.

The following argument leads to an upper bound on the size of the largest frequency component in the spectrum of $e_A[n]$. Assuming the quantizer is not saturated by the input signal $x[n]$, the maximum possible quantization error is $\Delta_A/2$, where Δ_A is the amplitude quantization step size. The total power in $e_A[n]$ is then bounded by $\Delta_A^2/4$. By Parseval's relation, the sum of the spur powers in the spectrum of $e_A[n]$ equals the power in $e_A[n]$. In order to maximize the power in a given spur, the total number of spurs must be minimized. Since $e_A[n]$ is real, the maximum power in a spur occurs when there are two frequency components, at $+\omega_{spur}$ and $-\omega_{spur}$, with equal

power.[1] With two frequency components, the power in a single spur is $\leq \Delta_A^2/8$.

Since $x[n]$ is real, its spectrum consists of a positive and a negative frequency component, each having power $A^2/4$. Using the above bound on spur power, the spurious-to-signal ratio ($SpSR$) is $\leq \Delta_A^2/(2A^2)$. If $A = A_Q \approx 1/2$ provided b is not small, then in decibels with respect to the carrier (dBc), $SpSR \leq 3 - 6b$ dBc, where $\Delta_A = 2^{-b}$, and b is the word length in bits. In summary, this upper bound on power in a spur caused by amplitude quantization exhibits -6 dBc per bit behavior.

IV. Phase Quantization Effects

Let a phase waveform, $\phi[n]$, be the input to the mid-tread quantizer. The phase waveform, $\phi[n] = \langle fn + \Phi/2\pi \rangle$, is a sampled sawtooth with amplitude ranging from 0 to 1. The fractional operator, $\langle x \rangle$, is defined so that $\langle x \rangle = x \bmod 1$, e.g., $\langle 1.3 \rangle = 0.3$. Since $\phi[n]$ is generated by a synchronous, finite-word-length, discrete-time system, it has a finite period. The signal output from the quantizer can be expressed as $\phi[n] + e_P[n]$, where $e_P[n]$ represents the error introduced by quantization. Since $\phi[n]$ is periodic, $e_P[n]$ is periodic with a period less than or equal to the period of $\phi[n]$. After multiplication by 2π and passage through the ideal function generator, the output signal is $y[n] = A\cos(2\pi\phi[n] + 2\pi e_P[n])$. If the quantizer has many levels, i.e., >16, $e_P[n] \ll 1$ and the small angle approximation $y[n] \approx A\cos(2\pi\phi[n]) - 2\pi A e_P[n]\sin(2\pi\phi[n])$ may be used.

Since $e_P[n]$ and $\phi[n]$ are periodic, the total error $2\pi A e_P[n]\sin(2\pi\phi[n])$ is periodic. The total error power is bounded by $\pi^2 A^2 \Delta_P^2$ because $e_P[n]$ is bounded by $\Delta_P/2$ and the magnitude of a sinusoid is bounded by unity. Recalling the arguments in the previous section on amplitude quantization effects, the maximum spur power of the real error signal is bounded by placing the total error power into two spectral components. Therefore, the maximum spur power is $\pi^2 A^2 \Delta_P^2/2$, where $\Delta_P = 2^{-b}$ and b bits are used to represent phase samples. By the above approximation for $y[n]$ and the bound on the spur power, the spurious-to-signal ratio bound is $SpSR \leq 2\pi^2\Delta_P^2 = 13 - 6b$ dBc, independent of the signal amplitude, A. This simple proof demonstrates the -6 dBc per phase bit behavior. More complicated arguments [7] improve the bound by about 9 dB.

[1] DC offsets and half sampling rate spurs are excluded because they can be corrected by appropriate calibration and filtering.

V. Amplitude Dithering

In this section, rounding the sum of an already quantized sinusoid and an appropriate dither signal is shown to cause spurious magnitudes that depend on the original (longer) word length, not on the output (shorter) word length. This phenomenon occurs at the expense of increased system noise from the addition of the dithering signal. An important finite-word-length dithering system is subsequently shown to be equivalent to the continuous-amplitude uniformly dithered system.

Consider the conceptual block diagram for a waveform generator shown in Fig. 3. The b-bit quantizer can be split into two parts, as in Fig. 4: a high-resolution B-bit quantizer ($B > b$) followed by truncation or rounding to b bits. Thus, the generation process consists of two separate steps: production of a high-resolution waveform and reduction of the word length. The number of bits used to represent the high-resolution samples should be sufficient to guarantee the desired spectral purity. Then the word length should be reduced without creating excess signal-dependent quantization error.

The input in Fig. 5 is a B-bit representation of a sinusoid, $x[n] = A\sin(2\pi\phi[n]) + e_{A0}[n]$, where $e_{A0}[n]$ is the quantization error. The dither signal, $z_u[n]$, is white noise uniformly distributed in $[-\Delta_A/2, \Delta_A/2)$, where $\Delta_A = 2^{-b}$. The sum $z_u[n] + x[n]$ is rounded to retain only the b most significant bits. The rounding can be modeled as a uniform quantizer with step size Δ_A. The amplitude A is chosen to avoid saturating this quantizer when the dither signal is added, i.e., $A + \Delta_A/2 \leq A_Q$.

The output from the quantizer can be expressed as $y[n] = x[n] + z_u[n] + e_A[n]$. The characteristic function of the dither signal, $z_u[n]$, is

$$F_z(\alpha) = E\{\exp(j\alpha z[n])\}$$

$$= \frac{2\sin(\alpha\Delta_A/2)}{\alpha\Delta_A} = \operatorname{sinc}\left(\frac{\alpha\Delta_A}{2\pi}\right) \qquad (2)$$

which has zeros at nonzero integer multiples of $2\pi/\Delta_A$. Thus, as shown in [1], $e_A[n]$ will be a white, wide-sense stationary process, uniformly distributed over $[-\Delta_A/2, \Delta_A/2)$, and it will not contribute spurious harmonics to the output spectrum of $y[n]$. Any spurious components in $y[n]$ are therefore due to $e_{A0}[n]$, which are present in the B-bit input, $x[n]$.

It remains to comment on the noise power not isolated in discrete spurious frequency components. If the

sequences $e_A[n]$ and $z_u[n]$ are uncorrelated, adding the variances of the quantization error, $\Delta_A^2/12$, and the dither process, $\Delta_A^2/12$, yields a white noise power of $\Delta_A^2/6$. This approximation is twice the variance of a quantization system with no dithering signal. Note that $e_{A0}[n]$ also contributes a white noise term, and that, in general, $e_A[n]$ and $z_u[n]$ are not uncorrelated. However, these two effects are dominated by the $\Delta_A^2/6$ behavior of the white noise power. In summary, $y[n]$, which is quantized to b bits, exhibits spurious performance as if it were quantized to B bits ($B > b$), at the expense of doubling the white noise power.

Because the input $x[n]$ is expressed as a B-bit value, an important system equivalent to continuous-amplitude, uniformly dithered word-length reduction can be constructed. Replace the uniformly distributed dither signal, $z_u[n]$, by a finite word-length representation of it, $z[n]$, which is said to be discreetly and evenly distributed over the $(B-b)$-bit quantized values in the region $[-\Delta_A/2, \Delta_A/2)$. Heuristically, $z[n]$ randomizes the portion of the finite word-length input, $x[n]$, that is about to be thrown away by the rounded truncation. This process is equivalent to continuous uniform dithering, since if $x[n]$ is padded out to an infinite number of bits by placing zeros beyond the least significant bit, then only the $B-b$ most significant bits of $z_u[n]$ will have an effect on the resulting sum, $x[n]+z_u[n]$. All of the bits below the most significant $B-b$ are added to zero and cannot beget a carry. The output, $y[n]$, is identical in both systems. Therefore $z_u[n]$, continuously, uniformly distributed over $[-\Delta_A/2, \Delta_A/2)$, can be replaced by the discretely valued $z[n]$ and yield the same spurious response for $y[n]$.

It appears that the finite word-length dither signal, $z[n]$, could be generated by an LFSR, or PN generator. This will be strictly true if and only if the PN generator has an infinite period, since, at this time, the dither signal is required to be white. However, it is not surprising that ideal behavior is approached as the period of the PN generator gets longer. With a sufficiently long period, the case where spur magnitudes are limited by the original word length can be achieved. The following section gives a simple model for a system implementation using a periodic random sequence that can be approximated by a PN generator.

VI. Effect of Periodic Dither

This section analyzes the use of a periodic dither signal with a long period, L, for both amplitude and phase dithering. Since the dither signal is periodic, the discrete frequency components in its spectrum will contaminate the desired signal. It is shown that the period can be chosen to satisfy worst-case spurious specifications. In this section, the case where the dither signal is generated using one uniform variate ($M = 1$) is given. When the dither signal is the sum of M independent uniform variates ($M > 1$), as in Section VIII, the analysis is the same because the resulting signal is an independent identically distributed (i.i.d.) sequence of random variables.

Instead of using the white dither process, $z_u[n]$, described in the previous section, consider a substitute, $z_L[n]$, which is periodic with period L. Any two samples, $z_L[n]$ and $z_L[n+m]$, where $m \neq 0 \bmod L$, are independent. Samples of $z_L[n]$ are uniformly distributed between $[-\Delta/2, \Delta/2)$, and the quantization step size is Δ.

When $z_L[n]$ is used as the dither signal, let the quantizer error be called $e_L[n]$. The autocorrelation of $z_L[n]$ when the lag, m, is an integer multiple of L is equal to $R_{z_L z_L}[0] = \Delta^2/12$. In the PN generator approximation to this noise source, $L = 2^l - 1$, where l is the length of the shift register in bits. At other lag values, the samples of $z_L[n]$ are independent, and since they have zero mean, the autocorrelation is zero. Therefore,

$$R_{z_L z_L}[m] = \frac{\Delta^2}{12}\delta[m \bmod L] = \sum_{l=0}^{L-1} \frac{\Delta^2}{12L} \exp\left(\frac{j2\pi ml}{L}\right)$$

and $z_L[n]$ contains L discrete frequency components, each with power $\Delta^2/(12L)$.

In the autocorrelation expression for $e_L[n]$, the expectation is taken over the random variables $z_L[n]$ and $z_L[n+m]$:

$$R_{e_L e_L}[n, n+m] = \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \sum_{\substack{l=-\infty \\ l\neq 0}}^{\infty} \alpha_k[n]\alpha_l^*[n+m] \times E\left\{\exp\left(\frac{j2\pi}{\Delta}(kz_L[n] - lz_L[n+m])\right)\right\} \quad (3)$$

where:

$$\alpha_k[n] = \frac{\Delta(-1)^k}{j2\pi k} \exp\left(\frac{j2\pi ks[n]}{\Delta}\right)$$

The desired signal to which the dither signal $z_L[n]$ is added is $s[n]$. Using the notation from earlier sections, in-phase quantization $s[n] = \phi[n]$, and in amplitude quantization $s[n] = x[n]$. When the lag is not a nonzero integer multiple of L,

$$E\left\{ \exp\left(\frac{j2\pi}{\Delta}\left(kz_L[n] - lz_L[n+m]\right)\right)\right\} =$$

$$E\left\{ \exp\left(\frac{j2\pi k z_L[n]}{\Delta}\right)\right\}$$

$$\times E\left\{\exp\left(\frac{-j2\pi l z_L[n+m]}{\Delta}\right)\right\}$$

$$= F_z\left(\frac{2\pi k}{\Delta}\right) F_z\left(\frac{-2\pi l}{\Delta}\right) = \delta[k]\delta[l]$$

This last fact is true because the characteristic function of $z_L[n]$ has zeros at all nonzero integer multiples of $2\pi/\Delta$, Eq. (2). But since the sums over k and l never assume the value 0, the autocorrelation function is 0 when the lag is not 0 mod L. When the lag is 0 mod L,

$$E\left\{ \exp\left(\frac{j2\pi}{\Delta}\left(kz_L[n] - lz_L[n+m]\right)\right)\right\} =$$

$$E\left\{ \exp\left(\frac{j2\pi(k-l) z_L[n]}{\Delta}\right)\right\} = \delta[k-l]$$

This results in

$$R_{e_L e_L}[n, n+m] =$$

$$\frac{\Delta^2}{2\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\cos\left(\frac{2\pi k}{\Delta}\left(s[n] - s[n+m]\right)\right) \qquad (4)$$

Setting $m = 0$ in Eq. (4) and evaluating the resulting summation [9, p. 7] yields the power in $e_L[n]$: $R_{e_L e_L}[n, n] = \Delta^2/12$. From Eq. (4), $e_L[n]$ is a cyclostationary process because $s[n]$ has a finite period, N. Using the results of Ljung [10], spectral information is obtained when Eq. (4) is averaged over time. Note that when the lag, m, is not only an integer multiple of L, the period of the dither, but also an integer multiple of N, the autocorrelation function equals $\Delta^2/12$, independent of n. The smallest nonzero lag that satisfies these two conditions is the least common multiple of L and N, denoted by qL where q is an integer. Therefore, the period of the time-averaged autocorrelation function, $\bar{R}_{e_L}[m] = \mathrm{Avg}_n(R_{e_L e_L}[n, n+m])$, is at least L and at most qL. Let the period equal cL, where c is an integer, $1 \le c \le q$. The function $\bar{R}_{e_L}[m]$ can be expressed as a sum of cL weighted complex exponentials:

$$\bar{R}_{e_L}[m] = \sum_{l=0}^{cL-1} p_l \exp\left(\frac{j2\pi ml}{cL}\right),\ m = \cdots, -1, 0, 1, 2, \cdots$$

where

$$p_l = \frac{1}{cL}\sum_{m=0}^{cL-1}\bar{R}_{e_L}[m]\exp\left(\frac{j2\pi ml}{cL}\right)$$

$$= \frac{1}{cL}\sum_{n=0}^{c}\bar{R}_{e_L}[nL]\exp\left(\frac{j2\pi nl}{c}\right)$$

The last equality is true since the autocorrelation function in Eq. (3) and its time-average, $\bar{R}_{e_L}[m]$, are zero for lags not equal to integer multiples of L. The weights, p_l, are the power magnitudes of the spurs. Since $\bar{R}_{e_L}[m] \le \Delta^2/12$, the spur power can be bounded: $p_l \le \Delta^2/(12cL) \le \Delta^2/(12L)$. Equality is achieved when the period of the time-averaged autocorrelation function is exactly L, the period of the dither.

As $L \to \infty$, the spacing between spurs goes to zero in the spectra of both $e_L[n]$ and $z_L[n]$. The power in an individual spur goes to zero, but the density (power per unit of frequency) tends to a constant. Thus, ideal white noise behavior is approached. While $z_L[n]$ and $e_L[n]$ are correlated in general, the worst-case spur power scenario coherently adds the power spectra from both processes. For this reason, L should be chosen to satisfy $\Delta^2/(6L) < P_{\max}$, where $P_{\max}$ is the maximum acceptable spur power. When constructing a dither signal as the sum of $M \ge 1$ independent, uniform variates, the noise autocorrelation becomes $R_{z_L z_L}[m] = (M\Delta^2/12)\delta[m \bmod L]$. The analysis follows closely that for $M = 1$, and L should be chosen to satisfy $(M+1)\Delta^2/(12L) < P_{\max}$.

As in the previous section, since the desired signal has finite word length, it is equivalent to rounding or truncating the dither signal to an appropriate word length. Such a truncated periodic noise source is an approximation to

an implementation using a PN generator that produces a periodic sequence of discreetly and evenly distributed random numbers.

VII. Phase Dithering

In this section, phase dithering is analyzed using a continuous, zero-mean, wide-sense stationary sequence. As described in Section V on amplitude dithering, an evenly distributed discrete random sequence is equivalent to continuous uniform dithering when the initial phase word is quantized to a finite number of bits.

Let the digital sinusoid to be generated be

$$x[n] = \cos\ (2\pi(\phi[n] + \epsilon[n])) \tag{5}$$

so that the desired phase is $\phi[n]$, as defined in Section IV. The total quantization noise is $\epsilon[n] = e_P[n] + z[n]$, the sum of the dither signal and the quantizer error. Using small angle approximations,

$$x[n] = \cos\ (2\pi fn + \Phi) - 2\pi\epsilon[n]$$

$$\times\ \sin\ (2\pi fn + \Phi) + O\left([\max\ (\epsilon[n])]^2\right)$$

The total quantization noise will be examined by considering the first two terms above, and then the second-order, $O\left([\max\ (\epsilon[n])]^2\right)$, effect.

A. First Order Analysis

Since the quantization error after dithering is independent of the input signal [2], $\epsilon[n]$ is uncorrelated with the desired sinusoids. Without loss of generality, and for ease of notation, let us shift the uniformly distributed dither random variate range to $[0,\Delta_P)$. The total phase quantization noise $\epsilon[n]$ will be $\epsilon[n] = -p[n]\Delta_P$ with probability $(1 - p[n])$, and $\epsilon[n] = (1 - p[n])\Delta_P$ with probability $p[n]$. The value $p[n]$ is the distance from the initial high-precision phase value, $\phi[n]$, to the nearest greater quantized value normalized by the phase quantization step size Δ_P. The value of the probability sequence $p[n]$ varies periodically, since $p[n] = \phi[n] \bmod \Delta_P$, and $\phi[n]$ is periodic; however, at all sample times n, the first moment of the total phase quantization noise, $E\{\epsilon[n]\}$, is zero.

Information about the spurs and noise in the power spectrum of $x[n]$ is obtained from the autocorrelation function. The autocorrelation of $x[n]$ is

$$\begin{aligned} E\{x[n]x[n+m]\} &= \cos\ (2\pi fn + \Phi) \\ &\quad \times \cos\ (2\pi f(n+m) + \Phi) \\ &\quad +\ 4\pi^2 \sin\ (2\pi fn + \Phi) \\ &\quad \times \sin\ (2\pi f(n+m) + \Phi) \\ &\quad \times E\{\epsilon[n]\epsilon[n+m]\} + O(\Delta_P^4) \end{aligned}$$

Spectral information is obtained by averaging over time [10], resulting in

$$\bar{R}_{xx}[m] \approx \frac{1}{2}\left[1 + 4\pi^2 \bar{R}_{\epsilon\epsilon}[m]\right] \cos\ (2\pi fm)$$

where $\bar{R}_{\epsilon\epsilon}[m] = \mathrm{Avg}_n(E\{\epsilon[n]\epsilon[n+m]\})$, the time-averaged autocorrelation of the total quantization noise.

The power spectrum of $x[n]$, the Fourier transform of the autocorrelation, is the power spectrum of the desired sinusoid of frequency f plus the total quantization noise amplitude modulated on the desired sinusoid. Note that since $\bar{R}_{\epsilon\epsilon}[m] = O(\Delta_P^2)$, and $\Delta_P << 1$, the modulation index is small.

The amplitude modulation (AM) signal produced by phase dithering is clear of spurious harmonics down to the level due to periodicities in the dither sequence. The next section will examine spur performance in more detail, but first it is important to consider the noise power spectral density resulting from the phase dithering process.

Recall that for any fixed time n, the probability distribution of $\epsilon[n]$, a function of $p[n]$, is determined by the input signal, but the outcome of $\epsilon[n]$ is determined entirely by the outcome of the dither signal $z[n]$. When $z[n]$ and $z[n+m]$ are independent random variables for nonzero lag m, $\epsilon[n]$ and $\epsilon[n+m]$ are also independent for $m \neq 0$, and hence $\epsilon[n]$ is spectrally white. In this case, the autocorrelation becomes

$$\bar{R}_{xx}[m] \approx \frac{1}{2}\ \cos\ (2\pi fm) + 2\pi^2\delta[m]\ \mathrm{Var}\ (\epsilon)$$

where Var (ϵ) is the time-averaged variance of the total quantization noise, and $\delta[m]$ is the Kronecker delta function ($\delta[0] = 1$, $\delta[m] = 0$, $m \neq 0$). The resulting signal-to-noise ratio (SNR) is approximately $1/(4\pi^2\ \mathrm{Var}\ (\epsilon))$.

When the dither signal is constructed from one uniform $[-\Delta_P/2, \Delta_P/2)$ random variate, the error $\epsilon[n]$ is bounded between $-\Delta_P$ to Δ_P with $\Delta_P = 2^{-b}$, and b is the number of bits in the phase representation after the word length is reduced. The number of bits, b, must be large enough to satisfy the earlier small angle assumption, typically, $b \geq 4$. The time-averaged variance of $\epsilon[n]$ is less than or equal to $2^{-2b}/4$, and the SNR is $2^{2b}/\pi^2 = 6.02b - 9.94$ dB.

Since the sinusoid generated is a real signal, the signal power in the SNR will be divided between positive and negative frequency components. If the sinusoid is the result of a discrete-time random process with sampling frequency f_s, then the resulting noise power spectral density (NPSD) will be given by:

$$\text{NPSD} \approx -\left[\frac{\text{SNR}}{2} + 10 \log_{10}\left(\frac{f_s}{2}\right)\right] \text{ dBc/Hz}$$

$$\leq 6.93 - 10 \log_{10}\left(\frac{f_s}{2}\right) - 6.02b \text{ dBc/Hz} \quad (6)$$

Table 1 gives noise power spectral densities as a function of the number of bits per cycle, b, at a 160-MHz sampling rate, calculated according to the above formula.

B. Second Order Analysis: Residual Spurs

For a worst-case analysis of second order effects, expand the initial cosine from Eq. (5) by the sum of angles formula:

$$x[n] = \cos(2\pi\epsilon[n]) \cos(2\pi f n + \Phi)$$

$$- \sin(2\pi\epsilon[n]) \sin(2\pi f n + \Phi)$$

Information about the spurs in the power spectrum of $x[n]$ is obtained from the autocorrelation function at nonzero lags. When the dither sequence, $z[n]$, is a sequence of i.i.d. variates, the autocorrelation function for $x[n]$, with lag m not equal to zero, is

$$R_{xx}[n, n+m] = E\{x[n]x[n+m]\}$$

$$= E\{x[n]\}E\{x[n+m]\}$$

The expected value of $x[n]$ is a deterministic function of time. From the above expression, it follows that spectral information about the random process $x[n]$, with the exception of noise floor information, is contained in $E\{x[n]\}$, which we call the "expected waveform." Since $\epsilon[n]$ is zero mean at all sample times, the expected waveform reduces to

$$E\{x[n]\} = (1 - 2\pi^2 E\{\epsilon^2[n]\}) \cos(2\pi f n + \Phi) + O(\Delta_P^3)$$

It remains to consider the second moment of the total phase quantization noise, $E\{\epsilon^2[n]\}$, which we evaluate by using the probability sequence $p[n]$ from the previous section as $E\{\epsilon^2[n]\} = \Delta_P^2(p[n] - p^2[n])$. Since $p[n]$ is bounded between 0 and 1, the function $u[n] = p[n] - p^2[n]$ is bounded between 0 and 1/4, with its maximum value of 1/4 at $p[n] = 1/2$.

Since $u[n]$ is bounded between 0 and 1/4, it must have some nonzero dc (average) component. Any remaining components can be periodic in the worst case. Since all nonlinear operations have been performed, conservation of power (energy) arguments can be used to determine the total non-dc error power. The total power in the dc component of $u[n]$ is equal to the square of the average value of $u[n]$. Similarly, the total power in $u[n]$ is equal to the average value of $u^2[n]$. Thus, the power remaining for time-varying components of $u[n]$ is

$$\text{Avg}(u^2[n]) - (\text{Avg}(u[n]))^2 = \text{Avg}\left[(u[n] - \text{Avg}(u[n]))^2\right]$$

This value is maximized by maximizing the dispersion of the samples about the mean. When the sample values are bounded, this maximization is achieved by placing half of the samples at each bound, so that the mean is equidistant from each bound. Since $0 \leq u[n] \leq 1/4$, the maximum power present in harmonic components is 1/64.

Recall that at this worst case, half of the values of $E\{\epsilon^2[n]\}$ are zero and half are $\Delta_P^2/4$. Since $\epsilon^2[n]$ is nonnegative, $E\{\epsilon^2[n]\} = 0$ implies that $\epsilon[n] = 0$. Note that the difference between $\epsilon[n]$ and $\epsilon[n+1]$ is the phase increment modulo the quantization step size. If, for any n and $n+1$, $\epsilon[n] = \epsilon[n+1] = 0$, the phase increment can be exactly expressed in the new quantization step. By induction, $\epsilon[n]$ will be zero for all n if any two adjacent values $E\{\epsilon^2[n]\}$ and $E\{\epsilon^2[n+1]\}$ are both zero. The only possible sequence $E\{\epsilon^2[n]\}$ achieving the worst case is, therefore, 0, 1/4, 0, 1/4, 0, 1/4.... This sequence has a single sinusoidal component at the Nyquist frequency, which is half the sampling rate.

In the worst case, the model to consider is $u[n] = 1/8 - (1/8)\cos(\pi n)$ since $\cos((2\pi f + \pi)n + \Phi) = \cos((2\pi f - \pi)n + \Phi)$. The expected waveform is

$$E\{x[n]\} = \left(\left(1 - \pi^2\frac{\Delta_P^2}{4}\right) + \left(\pi^2\frac{\Delta_P^2}{4}\right)\cos(\pi n)\right)$$

$$\times \cos(2\pi f n + \Phi) + O(\Delta_P^3)$$

$$= \left(1 - \pi^2\frac{\Delta_P^2}{4}\right)\cos(2\pi f n + \Phi) + \left(\pi^2\frac{\Delta_P^2}{4}\right)$$

$$\times \cos((2\pi f + \pi)n + \Phi) + O(\Delta_P^3)$$

clearly showing the desired signal and spur components. Thus, dropping the $O(\Delta_P^3)$ term, a -18 dB per bit power behavior, the worst-case spur level relative to the desired signal after truncating to b bits is $SpSR$

$$SpSR \approx \frac{\pi^4\Delta_P^4/16}{(1 - \pi^2\Delta_P^2/4)^2} \approx \frac{\pi^4\Delta_P^4}{16} = 7.84 - 12.04b \text{ dBc}$$

This worst case is achieved for a large class of frequencies. Let the phase of the desired signal $\Phi = 0$, and let the desired frequency be R cycles in 2^{b+1} samples, where R is an odd integer. The sequence $\epsilon^2[n]$ will be deterministic: 0, $\Delta_P^2/4$, 0, $\Delta_P^2/4$..., exactly the worst case analyzed above. The frequency of the spur is the reflection of the desired signal across 1/4 the sampling rate, and, as a result, it can be as close as $1/2^b$ of the sampling rate from the desired sinusoid.

In summary, if b bits of phase are output to a lookup table, and B bits of phase ($B > b$) are used prior to truncation, then the addition of an appropriate dithering signal using $(B - b)$ bits will allow the word length reduction without introducing spurs governed by the usual $-6b$ dBc behavior. If a single random variate is added as a dither signal (first-order dithering), the spur suppression is accelerated to 12 dB per bit of phase representation. Since the table size is affected only linearly by the number of bits in a table entry, rather than exponentially, as it is by the number of phase bits, the amplitude word length is of secondary importance to the phase word length, especially in all-digital systems. For example, -90 dBc spur performance would nominally require $b = 16$ bits of phase and a 65,536-entry table. With first-order dithering, this level of performance requires only $b \geq 8.1$ bits of phase per cycle in the lookup table addressing. Worst-case spur performance of -100.5 dBc is achieved with 9 bits, a 512-entry table at most, and, at a 160-MHz sampling rate, Table 1 shows that with these realistic system parameters, the noise power spectral density is at a low -126 dBc/Hz.

VIII. Accelerated Spur Suppression

Further analysis [11] based on an extension of results by Gray [12] indicates that the phase spur suppression rate can be increased in steps of 6 dBc per bit by adding multiple uniform random deviates to the phase value prior to truncation. The addition of M uniform random deviates produces a dither signal with Mth-order zeros in its characteristic function, thus making the Mth moment of the quantization error independent of the input sequence [12].

An example of this technique providing 18 dBc per phase bit spur performance is shown in Fig. 6. This technique involves adding two $(B - b)$-bit uniform deviates to produce a $(B - b + 1)$-bit dither signal, which achieves the accelerated spur reduction due to second-order zeros in the dither characteristic function. Simulation results for when two uniform variates are added to the phase are presented in the next section. A straightforward extension of this technique to a polynomial series allows spur-reduced synthesis of periodic digital signals with arbitrary waveforms.

IX. Simulation Results

Simulations were performed to validate the results of this analysis. These results were obtained using 8192-point unwindowed fast Fourier transforms (FFT's), and the synthesized frequencies were chosen to represent worst-case amplitude and phase spur performance. Figure 7 shows the power spectrum of a sine wave of one-eighth the sampling frequency truncated to 8 bits of amplitude without dithering. Figure 8 shows the same spectrum with a 16-bit sinusoid amplitude dithered with 1 uniform variate prior to truncation to 8 bits. Note that the spurs have been eliminated to the levels consistent with those imposed by the initial 16-bit quantization.

Figure 9 shows the spectrum of a 5-bit phase-truncated sinusoid with high-precision amplitude values. A worst-case example of first-order phase dithering is shown in Fig. 10. The measured noise power spectral density in Fig. 10 is -62.3 dBc per FFT bin, giving a noise density of $-23.2 - 10\log(f_s/2)$ dBc, in agreement with the upper bound derived in Eq. (6). The spur level is -52.3 dBc in the first-order dithering shown in Fig. 10.

Figure 11 shows the same example using second-order ($M = 2$) dithering using the sum of two uniform deviates. While the spectrum in Fig. 10 shows the residual spurs at -12 dBc per bit due to second-order effects, Fig. 11 shows no visible spurs, indicating better than -63 dBc spurious performance. Additional simulations involving megapoint FFTs and not represented by figures confirm the

−18 dBc per bit performance of the second-order phase-dithered system.

Finally, Fig. 12 shows a worst-case result for first-order phase dithering together with first-order amplitude dithering. The amplitude samples are truncated to 8 bits, as are the phase samples. Note that the spurs are not visible in the spectrum; however, close analysis has demonstrated that they are present at the −88 dBc level expected due to second-order effects.

X. A System Design Example

The block diagram of a direct digital frequency synthesizer based on the techniques presented here is shown in Fig. 13. The following system would perform at a sampling rate of 160 MHz, producing 8-bit digital sinusoids spur-free to −90 dBc with better than −120 dBc/Hz noise power spectral density. The system parameters are

(1) Phase bits are in unsigned fractional cycle representation with phase accumulator word length determined by frequency resolution and ≥ 16 bits prior to the addition of one uniform phase dither variate, with ≥ 9 bits after dither addition and truncation.

(2) Amplitude lookup table with

   (a) $\geq 2^7 = 128$ entries (using quadrant symmetries) of ≥ 16 bits each, normalized so that the sinusoid amplitude equals 512 16-bit quantization steps less than the full-scale value.

   (b) Linear feedback shift register PN generator with ≥ 16 lags producing one 8-bit amplitude dither variate.

   (c) One LFSR PN generator with ≥ 18 lags for generation of the 7-bit phase dither variate.

XI. Conclusion

A digital dithering approach to spur reduction in the generation of digital sinusoids has been presented. A class of periodic dithering signals has been analyzed because of its similarity to LFSR PN generators.

The advantage gained in amplitude dithering provides for spur performance at the original longer word length in an ideal system when the digital dithering signal is white noise distributed evenly, not uniformly, over one quantization interval. The reduced word length allows the use of less complicated multipliers and narrower data paths in purely digital applications. If the waveform is ultimately converted to an analog value, the reduced word length allows the use of fast, coarse-resolution, highly linear digital-to-analog converters (DAC's) to obtain sinusoids or other periodic waveforms whose spectral purity is limited by the DAC linearity, not its resolution. These results suggest that coarsely quantized, highly linear techniques for digital-to-analog conversion, such as delta-sigma modulation, would be useful in direct digital frequency synthesis of analog waveforms.

The advantage gained in the proposed method of phase dithering provides for an acceleration beyond the normal 6 dB per bit spur reduction to a $6(M+1)$ dB per bit spur reduction when the dithering signal consists of M uniform variates. Often the most convenient way to generate a periodic waveform is by table lookup with a phase index. Since the size of a lookup table is exponentially related to the number of phase bits, this can provide a dramatic reduction in the complexity of numerically controlled oscillators, frequency synthesizers, and other periodic waveform generators.

The advantages of dithering come at the expense of an increased noise content in the resulting waveform. However, the noise energy is spread throughout the sampling bandwidth. In high bandwidth applications, dithering imposes modest system degradation. It has been shown that high performance synthesizers with dramatically reduced complexity can be designed using the dithering method, without resulting in high noise power spectral density levels.

Acknowledgment

Michael J. Flanagan was supported in part by a National Science Foundation Fellowship.

References

[1] A. B. Sripad and D. L. Snyder, "A Necessary and Sufficient Condition for Quantization Errors to be Uniform and White," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-25, pp. 442–448, October 1977.

[2] L. Schuchman, "Dither Signals and Their Effects on Quantization Noise," *IEEE Transactions on Communication Technology*, vol. COM-12, pp. 162–165, December 1964.

[3] N. S. Jayant and L. R. Rabiner, "The Application of Dither to the Quantization of Speech Signals," *Bell System Technical Journal*, vol. 51, pp. 1293–1304, 1972.

[4] S. C. Jasper, "Frequency Resolution in a Digital Oscillator," U.S. Patent Number 4,652,832, Washington, DC, March 24, 1987.

[5] P. O'Leary and F. Maloberti, "A Direct-Digital Synthesizer with Improved Spectral Performance," *IEEE Transactions on Communications*, vol. COM-39, pp. 1046–1048, July 1991.

[6] J. Tierney, C. Rader, and B. Gold, "A Digital Frequency Synthesizer," *IEEE Transactions on Audio and Electroacoustics*, vol. AU-19, no. 1, pp. 48–57, March 1971.

[7] T. Nicholas and H. Samueli, "An Analysis of the Output Spectrum of Direct Digital Frequency Synthesizers in the Presence of Phase Accumulator Truncation," *Proceedings of the 41st Annual Frequency Control Symposium*, pp. 495–502, 1987.

[8] L. E. Brennan and I. S. Reed, "Quantization Noise in Digital Moving Target Indication Systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-2, pp. 655–658, November 1966.

[9] I. S. Gradsteyn and I. M. Rizhk, *Table of Integrals, Series, and Products*, corrected and enlarged edition, New York: Academic Press, 1980.

[10] L. Ljung, *System Identification: Theory for the User*, Englewood Cliffs, New Jersey: Prentice-Hall, 1987.

[11] M. J. Flanagan and G. A. Zimmerman, "Spur-Reduced Digital Sinusoid Generation Using Higher-Order Phase Dithering," *Conference Record: Papers Presented at the 27th Annual Asilomar Conference on Signals, Systems and Computers* (in press), Pacific Grove, California: IEEE Computer Society Press, November 1993.

[12] R. M. Gray and T. G. Stockham, Jr., "Dithered Quantizers," presented at *1991 IEEE International Symposium on Information Theory*, Budapest, Hungary, June 1991.

Table 1. Noise power spectral densities for 160-MHz sampling rate.

b, bits/cycle	Noise power spectral density, dBc/Hz
5	−102.20
6	−108.22
7	−114.24
8	−120.26
9	−126.28
10	−132.30
11	−138.32
12	−144.35

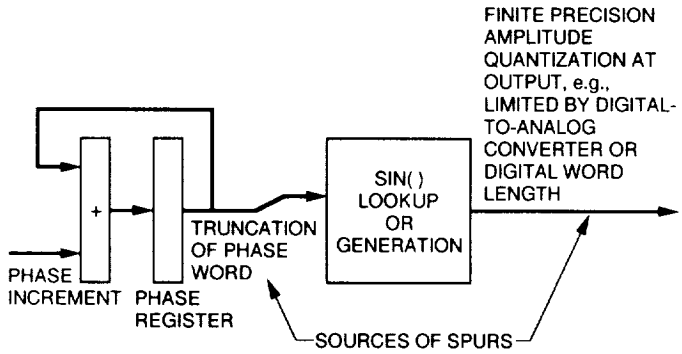


Fig. 1. Spur generation in conventional digital sinusoid generation.

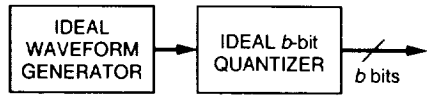


Fig. 3. Conceptual waveform generator model.

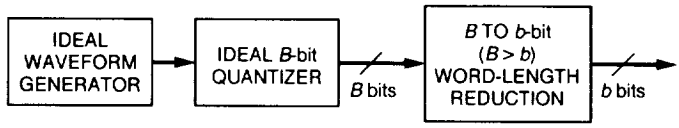


Fig. 4. Two-step waveform generator model.

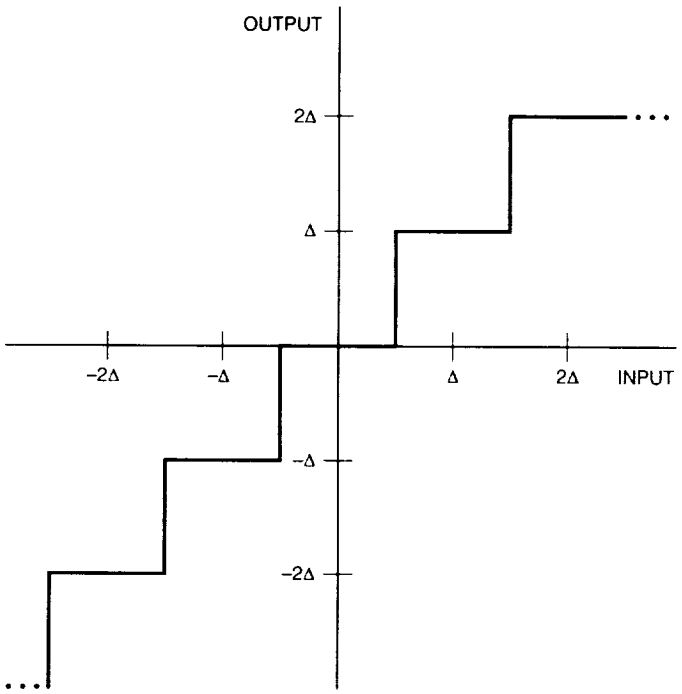


Fig. 2. Input/output relation of a midtread quantizer.

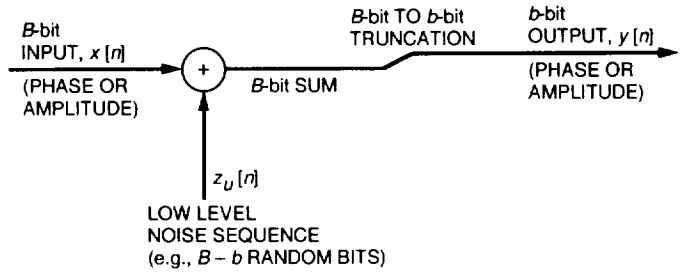


Fig. 5. Uniform dithered quantizer.

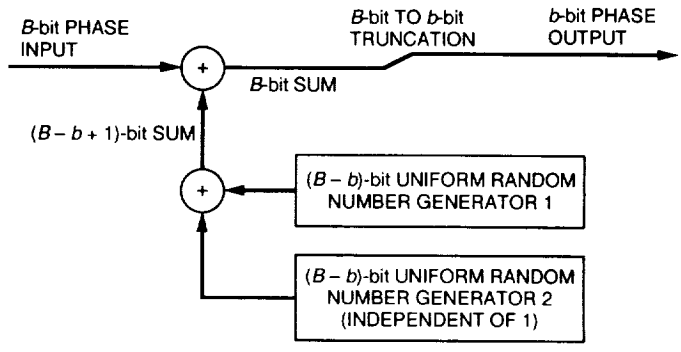


Fig. 6. System for 18-dBc-per-phase-bit spur reduction.

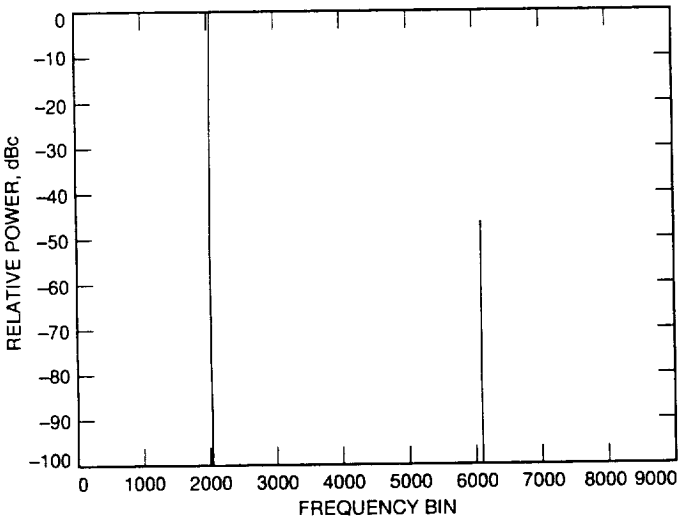


Fig. 7. Power spectrum of 8-sample/cycle sine wave without dithering (8-bit amplitude quantization).

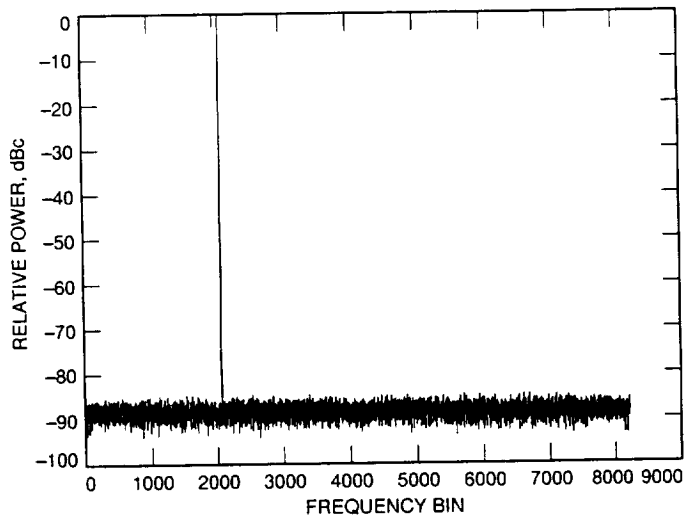


Fig. 8. Power spectrum of 8-sample/cycle sine wave with amplitude dithering (8-bit amplitude quantization).

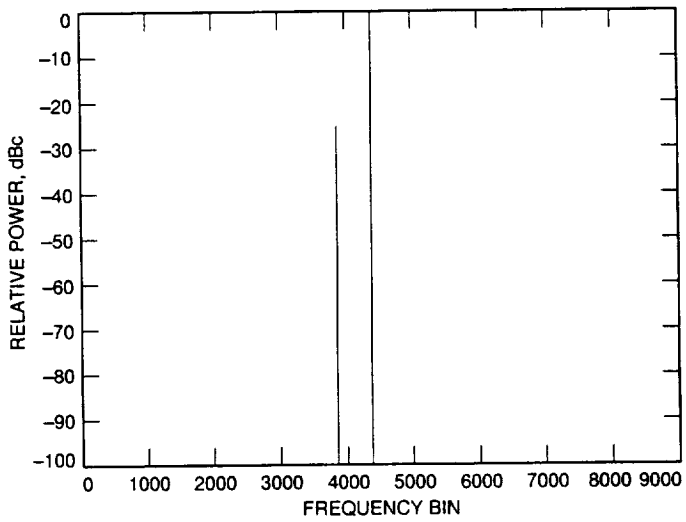


Fig. 9. Power spectrum of 5-bit phase-truncated sine wave without phase dithering (high-precision amplitude).

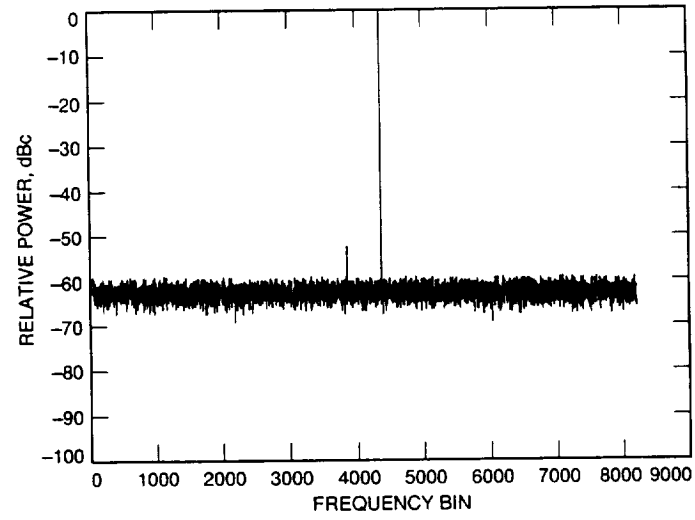


Fig. 10. Power spectrum of 5-bit phase-truncated sine wave with first-order phase dithering (high-precision amplitude).

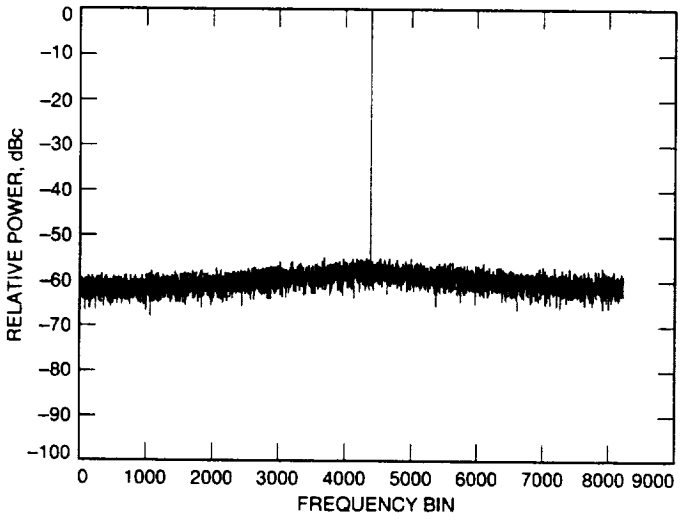


Fig. 11. Power spectrum of 5-bit phase-truncated sine wave with second-order phase dithering (high-precision amplitude).

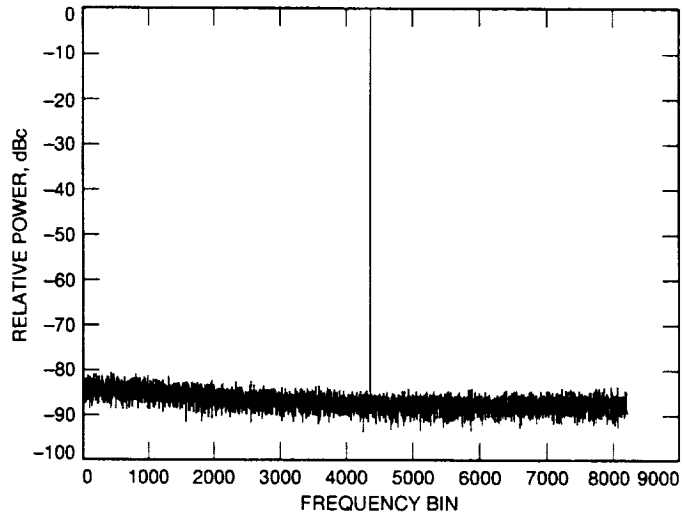


Fig. 12. Worst-case power spectrum of sinusoid with first-order phase dithering and amplitude dithering (8 bits each).

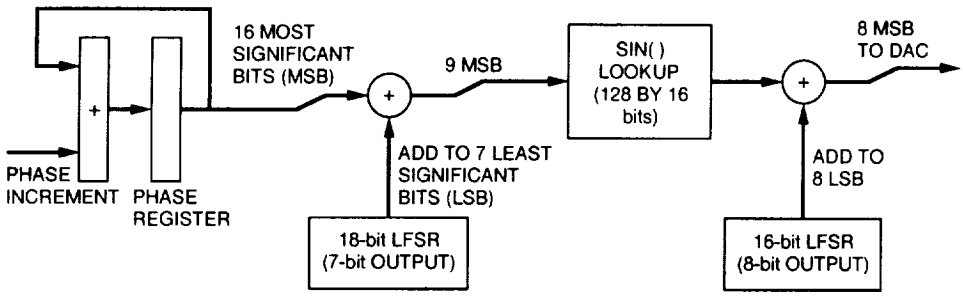


Fig. 13. Spur-reduced direct digital frequency synthesizer.

Box Codes of Lengths 48 and 72

G. Solomon[1] and Y. Jin[2]

A self-dual code of length 48, dimension 24, with Hamming distance essentially equal to 12 is constructed here. There are only six codewords of weight 8. All the other codewords have weights that are multiples of 4 and have a minimum weight equal to 12. This code may be encoded systematically and arises from a strict binary representation of the (8,4;5) Reed–Solomon (RS) Code over GF(64). The code may be considered as six interrelated (8,7;2) codes. The Mattson–Solomon representation of the cyclic decomposition of these codes and their parity sums are used to detect an odd number of errors in any of the six codes. These may then be used in a correction algorithm for hard or soft decision decoding.

A (72,36;15) box code was constructed from a (63,35;8) cyclic code. The theoretical justification is presented herein.

A second (72,36;15) code is constructed from an inner (63,27;16) Bose–Chaudhuri–Hocquenghem (BCH) code and expanded to length 72 using box code algorithms for extension. This code was simulated and verified to have a minimum distance of 15 with even weight words congruent to 0 modulo 4. The decoding for hard and soft decision is still more complex than the first code constructed above.

Finally, an (8,4;5) RS Code over GF(512) in the binary representation of the (72,36;15) box code gives rise to a (72,36;16) code with nine words of weight 8, and all the rest have weights $\geq$ 16.*

I. Codes of Length 48

The self-dual (48,24;12) Quadratic Residue Code had a history of difficulty and complexity in decoding for five errors algebraically as well as decoding for soft decision. This led us to apply the techniques of box codes as successfully developed for Golay Codes to rate 1/2 codes of length 48. See [1,2]. Subcodes of dimension 23 and Hamming distance 12 were easily found. In addition, the box structure gave parity information to detect odd errors in rows that simplify decoding procedures.

The attempt to avoid the six codewords of weight 8 in the natural box code construction yielded two self-dual

[1] Independent consultant to the Communications Systems Research Section.

[2] Student at the California Institute of Technology, Pasadena, California.

(48,24;12) codes [1]. Upon closer examination of computer simulation, these codes were found to contain 42 words of weight 8.

In [1], the two codes constructed were designed to be self-dual. The (48,23;12) systematic subcodes of each were easily found. The 24th dimension in each was more elaborately constructed with the proviso that odd parities of the rows were induced to be used as tools in an erasure-error correcting decoding procedure. A search of the codeword weights' structure indicated the presence of 42 words of weights 8 and 40 in both these codes. The remaining nonzero words were of minimum weight 12. There exists a straight systematic construction of the Reed–Solomon (RS) (8,4;5) Code over $GF(64)$ for the 24th dimension given below, still using the particular binary representation in [1], which yields only six codewords of weights 8 and 40. This gives a box code with even parity on the rows. So for a low signal-to-noise ratio, this code and the previously constructed codes of dimension 48, rate 1/2, are effectively of minimum distance 12. The decoding procedure for soft decision mentioned in [1] is still applicable and preferred over any current soft decoding of the (48,24;12) Quadratic Residue Code.

In [2], a code of length 72 and distance 15 was constructed specifically to have simplified soft decoding. The (72,35;16) subcode was constructed with even parity on the nine rows in a nonsystematic manner as a subcode of the Reed–Solomon (8,4;5) Code over $GF(512)$. The 36th dimension was constructed to give odd parity on the rows and yield a code of minimum distance 15. The full code was designed to have a systematic encoding. This code, however, upon investigation, was found to have a very small number of words of length 11.

To meet this emergency, a new (72,36;15) box code is constructed here with rows of even or odd parity, and so it possesses, perhaps, a simple hard decision 7–8 error correcting procedure. This code has been simulated and verified to have a minimum distance of 15 and even weight words congruent to 0 modulo 4.

II. (8,4;5) RS Code Over *GF* (64)

Represent the Reed–Solomon (8,4;5) Code over $GF(64)$ in binary using the particular normal basis in [1]. One can generate a rate 1/2 self-dual code of length 48 and dimension 24 with weights that are multiples of 4.

This binary representation of the RS (8,4;5) Code over $GF(64)$ yields six (8,7;2) codewords whose decomposition via Mattson–Solomon into two cyclic code components and a constant component looks like (6,4;3) and (6,2;5) RS Codes over $GF(8)$ and a (6,6;1) binary code, respectively.

In particular, let γ be a root of the polynomial $f(x) = x^6 + x^5 + x^4 + x + 1$, where γ is a primitive generator of the 63 roots of unity. Represent the elements of $GF(64)$ in the normal representation using the roots of $f(x)$. The roots are γ^j; $j \in J$; $J = \{1, 2, 4, 8, 16, 32\}$.

NOTE: For this particular choice of $f(x)$, we have

$$\mathrm{Tr}\gamma^j = 1; \quad j \in J; \quad J = \{1, 2, 4, 8, 16, 32\}$$

$$\mathrm{Tr}(\gamma^i \gamma^k) = 0; \quad i \neq k; \quad i, k \in J$$

Let β be a root of the polynomial $g(x) = x^3 + x^2 + 1$. β is an element of $GF(8)$, a subfield of $GF(64)$, and $\beta = \gamma^9$.

A. Encoding

Now use the recursion or check polynomial $h(x) = \prod_{i=0}^{3}(x + \beta^i)$ to generate an extended (8,4;5) RS Code over $GF(64)$. This means that the initial shift register contains four elements in $GF(64)$ expressed as coefficients in the normal representation above. The cyclic portion of the code is of length 7, and the overall parity symbol is the eighth dimension. Represent the binary code as components $\mathrm{Tr}(P(x)\gamma^j)$; $j = 1, 2, 4, 8, 16, 32$.

The general Mattson–Solomon (M–S) polynomial of a codeword $\mathbf{a}$, similar to the Golay codeword over $GF(8)$, is $P_{\mathbf{a}}(x) = C_0 + C_1 x + C_2 x^2 + C_3 x^3$ where $C_i \in GF(64)$ for $0 \leq i \leq 3$ and $x \in GF(8)$.

Encode the codeword in its cyclic portion. The extended codeword $\mathbf{a}$ expressed in terms of the M–S polynomial is

$$\mathbf{a} = (P_{\mathbf{a}}(\beta^i); 0 \leq i \leq 6, P_{\mathbf{a}}(0))$$

Writing the codewords in binary using the normal basis γ^j, $j \in J$ above, there are six binary codewords of length 8:

$$\mathrm{Tr}(P(x)\gamma^j); \quad j = 1, 2, 4, 8, 16, 32$$

where $\mathrm{Tr}a$ denotes the value in $GF(2)$ given by the Trace of an element $a \in GF(64)$:

$$\mathrm{Tr}a = a + a^2 + a^4 + a^8 + a^{16} + a^{32}$$

Consider one of the six binary words in its Mattson–Solomon setting,

$$\begin{aligned}\mathrm{Tr}(P_{\mathbf{a}}(x)\gamma^j) &= \mathrm{Tr}((C_0 + C_1x + C_2x^2 + C_3x^3)\gamma^j) \\ &= \mathrm{Tr}(C_0\gamma^j) + \mathrm{Tr}'[(C_1x + C_2x^2 + C_3x^3)\gamma^j \\ &\quad + ((C_1x + C_2x^2 + C_3x^3)\gamma^j)^8]\end{aligned}$$

$$\mathrm{Tr}'a = a + a^2 + a^4; \quad a \in GF(8)$$

Set $C_0 = 0$ temporarily, as this does not affect the arguments to follow.

$$\begin{aligned}\mathrm{Tr}(P(x)\gamma^j) = \mathrm{Tr}'[&(C_1\gamma^j + (C_1\gamma^j)^8 + (C_2\gamma^j)^4 \\ &+ (C_2\gamma^j)^{32})x + ((C_3\gamma^j)^2 + (C_3\gamma^j)^{16})x^6]\end{aligned}$$

Lemma: The coefficient of x is a (6,4:3) code over $GF(8)$. The coefficient of x^3 is a (6,2;5) code over $GF(8)$. The code is indexed by the values of $\gamma^j; j \in J = \{1,2,4,8,16,32\}$.

Proof: The set $\gamma^j; j \in J = \{1,2,4,8,16,32\}$ is a linear independent set and thus can take zero values only one less than the number of terms in the coefficient of x, x^6. The term $\mathrm{Tr}(C_0\gamma^j)$ in the code's expression when $\mathrm{Tr}C_0 = 0$, the constant terms, forms a (6,5;2) binary code.

Theorem: The RS Code determined by codewords with M–S polynomials $P_{\mathbf{a}}(x); \mathrm{Tr}C_0 = 0$ forms a (48,23;12) binary code with weight multiples of 4.

Proof: The multiple of 4 property of the weights follows, using the Solomon–McEliece Γ_2 Formula.

$$\begin{aligned}\mathrm{Tr}(P(x)\gamma^j) = \mathrm{Tr}'[&(C_1\gamma^j + (C_1\gamma^j)^8 + (C_2\gamma^j)^4 \\ &+ (C_2\gamma^j)^{32})x + ((C_3\gamma^j)^2 + (C_3\gamma^j)^{16})x^6]\end{aligned}$$

where Tr is defined in $GF(64)$ and Tr′ is defined in $GF(8)$.

Now

$$\begin{aligned}\Gamma_2(\mathrm{Tr}P(x)\gamma^j) &= \mathrm{Tr}'(C_1C_3^2\gamma^{3j} + C_1^8C_3^2\gamma^{10j} + C_1C_3^{16}\gamma^{17j} \\ &\quad + C_1^8C_3^{16}\gamma^{24j} + C_2^{32}C_3^2\gamma^{34j} + C_2^4C_3^2\gamma^{6j} \\ &\quad + C_2^{32}C_3^{16}\gamma^{48j} + C_2^4C_3^{16}\gamma^{20j}) \\ &= \mathrm{Tr}(C_1C_3^2\gamma^{3j} + C_1^8C_3^2\gamma^{10j} + C_2^4C_3^2\gamma^{6j} \\ &\quad + C_2^4C_3^{16}\gamma^{20j})\end{aligned}$$

and therefore $\sum_{j\in J}\Gamma_2(\mathrm{Tr}P(x)\gamma^j) = 0$.

Recall that the normal basis was chosen so that $\mathrm{Tr}\gamma^j = 1; j \in J\mathrm{Tr}(\gamma^i\gamma^k) = 0;\ i \neq k;$ and $i, k \in J$.

It has been demonstrated that the binary weight of any codeword in the RS Code above is a multiple of 4. Since the symbol distance of the code is greater than 5, we have narrowed the weights down to $8, 12, 16, 20, \cdots, 40$.

III. Structure of the Code

Using the same arguments given in [1], the minimum weight of the code for $\mathrm{Tr}C_0 = 0$ is equal to 12. However, note that for each $C_0 = \gamma^i;\ i = 1,2,4,8,16,32$ and $C_i = 0$; and $i = 1,2,3$, one obtains a codeword of weight 8. We proved that these six are the only words of weight 8. A counting argument on the weights would do the same. Since all words have weight multiples of 4, the code is self-dual.

IV. (72,36;15) Code

In [2], an alternate (72,36;15) box code was constructed from the (63,35;8) cyclic code, generated by the check polynomial $f(x) = \prod f_i(x); i = 1,3,5,7,9,13,21$ where $f_i(x)$ is a polynomial irreducible over $GF(2)$ with β^i as a root where β is a primitive 63rd root of unity. We now present the theoretical justification.

Place the codewords in the usual 9×7 box code matrices corresponding to their values $7i + 9j \pmod{63}$ for $0 \leq i \leq 8$, $0 \leq j \leq 6$. Let $z = xy$ where $x^7 = 1, y^9 = 1, x = \beta^{9j}$, and $y = \beta^{7i}$. Indexing the rows by y, the M–S polynomial for each row y is

$$P_y(x) = \mathrm{Tr}(C_1 z + C_3 z^3 + C_5 z^5 + C_7 z^7 + C_{13} z^{13})$$

$$+ C_9 z^9 + C_{18} z^{18} + C_{36} z^{36} + C_{21} z^{21} + C_{42} z^{42}$$

$$= C_{21} y^3 + C_{21}^2 y^6 + \mathrm{Tr}(C_7 y^7)$$

$$+ \mathrm{Tr}'[(C_9^4 + C_1 y + C_1^8 y^8)x + (C_5^4 y^2 + C_5^{32} y^7$$

$$+ C_3^{16} y^3 + C_3^2 y^6 + C_{13} y^4 + C_{13}^8 y^5) x^6]$$

where Tr is defined in $GF(64)$ and Tr′ is defined in $GF(8)$.

Thus the coefficient of x is a (9,3;7) code over $GF(8)$, and the coefficient of x^6 is a (9,6;4) code over $GF(8)$.

Construct an eighth column on the nine rows by the usual parity rule. The eighth column will have the same Γ_2 value as the original (63,35;8) code. Then we immediately have the following lemma:

Lemma 1: The extended box code is a (72,35;16) code.

Proof: Consider the Solomon–McEliece Formula. The sum Γ_2 over the nine rows and eight columns gives 0, showing that the weight of every codeword is a multiple of 4. The properties of the coefficients of x and x^6 for the subcode or dimension 27 imply the minimum weight of the entire code to be 16.

Now adjoin a vector of all ones to the original 9×7 matrix setting. This will make the rows have odd parity and will complement the column sums. It is easy to show that all odd-weight codewords have weights of the form $4m - 1$. We will prove that the minimum code distance is 15.

The degree of the Mattson–Solomon polynomial for the entire 63 length code is 56; the next highest degree is 52. From this and the properties of the coefficients of x and x^6, one can easily see that the weight of the inner cyclic codeword is less than or equal to 54. If the inner weight is 54, the nine-row weight patterns 6 6 6 6 6 6 6 6 6, 7 6 6 6 6 6 6 6 5, and 7 7 6 6 6 6 6 6 4 could generate codewords of weights less than 15. If the inner weight is 52, the weight pattern 6 6 6 6 6 6 6 6 4 could generate codewords of weights less than 15.

Lemma 2: For the original cyclic (63,35;8) code, none of the weight patterns above are possible.

Proof: Weight pattern 6 6 6 6 6 6 6 6 6 gives the sum Γ_2 to be 1. This is impossible.

Let

$$P(y) = C_9^4 + C_1 y + C_1^8 y^8$$

$$Q(y) = C_5^4 y^2 + C_5^{32} y^7 + C_3^{16} y^3 + C_3^2 y^6 + C_{13} y^4 + C_{13}^8 y^5$$

then

$$\Sigma_y P(y) Q(y) = 0$$

$$\deg(P^6(y) + Q(y)) = 7$$

If the weight pattern is 7 6 6 6 6 6 6 6 5, then $\Sigma_y P(y)Q(y) = 1 + \alpha \neq 0$ for some $\alpha \neq 1$; $\alpha \in GF(8)$.

If the weight patterns are 6 6 6 6 6 6 6 6 4 or 7 7 6 6 6 6 6 6 4, the polynomial $P^6(y) + Q(y)$ has eight zeros. Then $P^6(y) = Q(y)$. But $P(y)Q(y) = 0$ and $P^6(y) = Q(y)$ cannot give weight 4 for any row indexed by y.

Theorem: The box code is a (72,36;15) code, where the even-weight subcode is a (72,35;16) code with all codewords having weights of the form $4m$, and the odd-weight subcode is a (72,35;15) code with all codewords having weights of the form $4m - 1$. **QED**

V. (72,36;15) Alternate Code

One can construct the Bose–Chaudhari–Hocquenghem (BCH) (63,27;16) code generated by the check polynomial $f(x) = \prod f_i(x), i = 1, 3, 5, 9, 11$. The cyclic decomposition in the box code setting yields a (9,5;5) code over $GF(8)$ for the coefficient of x and a (9,4;6) code over $GF(8)$ for the coefficient of x^6. This does extend to a (72,36;15) code, too. In fact, this code has been simulated and verified to have a minimum distance of 15 with even weight words congruent to 0 modulo 4. If we try all possibilites for the check polynomial $g(x) = \prod f_i(x)$; $i = 7, 21$, which totals to 256 codewords, we are left with an inner BCH code that can algebraically correct 7 errors. This leaves soft decoding still very complex and unworkable.

Note that because of the Mattson–Solomon decomposition here, one may correct five errors easily by hard decision, but six or more take more trials. Similarly, a soft

decision would require $8^4 = 2^{12}$ trials. This requires more trials for both hard and soft decisions than the alternate code mentioned above in [2]. Note the advantage that the (9,3;7) code over $GF(8)$, the coefficient of x^6, has over the (9,4;6) code over $GF(8)$, the coefficient of x.

VI. (8,4;5) RS Code Over $GF(512)$

The (8,4;5) RS Code over $GF(512)$ in the binary representation of [2] gives rise to a systematic (72,36;16*) code with nine words of weight 8, and all the rest have weights ≥ 16. The normal basis consists of $\gamma^i; i = 2^j; 0 \leq j \leq 8$ with γ a root of $f(x) = x^9 + x^8 + x^6 + x^5 + x^4 + x + 1$. The proof that there are no words of weight 12 is a simple counting argument. We prove there are no words of weight 60 in the code of dimension 35 given by $C_0 = 0$.

Represent the elements of $GF(512)$ in the normal representation using the roots of $f(x)$. The roots are $\gamma^j; j \in J$; and $J = \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$.

For this particular choice of $f(x)$, we have

$$\mathrm{Tr}\gamma^j = 1; \quad j \in J; \quad J = \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$$

$$\mathrm{Tr}(\gamma^i \gamma^k) = 0; \quad i \neq k; \quad i, k \in J$$

Represent the binary code as components $\mathrm{Tr}(P(x)\gamma^i)$; $i = 1, 2, 4, 8, 16, 32, 64, 128, 256$, giving nine words of length 8.

Let β be a root of the polynomial $g(x) = x^3 + x^2 + 1$. β is an element of $GF(8)$, a subfield of $GF(512)$ and $\beta = \gamma^{73}$.

A. Encoding

Now use the recursion or check polynomial $h(x) = \prod_{i=0}^{3}(x + \beta^i)$ to generate an extended (8,4;5) RS code over $GF(512)$. This means that the initial shift register contains four elements in $GF(512)$ expressed as coefficients in the normal representation above. The cyclic portion of the code is of length 7, the overall parity symbol; the eighth dimension is the usual sum over the seven symbols.

In the (72.36;16*) binary representation of the RS (8,4;5) Code over $GF(512)$, any codeword with the coeffiecients of x and x^6 nonzero has a minimum weight of 16. When these are zero, then clearly there are only nine words of weight 8, which come from the encoding of the symbol γ^j; $j \in J$; $J = \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$. A similar proof would argue that there are only six words of weight 8 in the binary representation of the (48,24;12*) RS Code over $GF(64)$.

To prove there are no words of weight 12, a counting argument notes that there are no words of weight 60 in the even-weight codes, where $\mathrm{Tr}C_0 = 0$. Words of weight 60 must possess a weight distribution over the nine words in any permutation of 8 8 8 6 6 6 6 6 6. This implies that three rows are zero and six rows are nonzero with weights 6 or $\Gamma_2 = 1$, to say the least. However, in [2], we note that this cannot be. The cyclic coefficients are (9,3;7) and (9,6;4) codes over $GF(8)$, so there are at least seven rows with $\Gamma_2 = 1$. **QED**

References

[1] G. Solomon, "Self-Dual (48,24;12) Codes," *The Telecommunications and Data Acquisition Progress Report 42-111, vol. July-September 1992*, Jet Propulsion Laboratory, Pasadena, California, pp. 75–79, November 15, 1992.

[2] G. Solomon, "A (72,36;15) Box Code," *The Telecommunications and Data Acquisition Progress Report 42-112, vol. October-December 1992*, Jet Propulsion Laboratory, Pasadena, California, pp. 19–21, February 15, 1993.

Integer Cosine Transform Compression for Galileo at Jupiter: A Preliminary Look

L. Ekroot, S. Dolinar, and K.-M. Cheung
Communications Systems Research Section

The Galileo low-gain antenna mission has a severely rate-constrained channel over which we wish to send large amounts of information. Because of this link pressure, compression techniques for image and other data are being selected. The compression technique that will be used for images is the integer cosine transform (ICT). This article investigates the compression performance of Galileo's ICT algorithm as applied to Galileo images taken during the early portion of the mission and to images that simulate those expected from the encounter at Jupiter.

I. Introduction

The Galileo low-gain antenna (LGA) mission will rely heavily on source coding, i.e., data compression, to maximize the return over the severely rate-constrained channel. A major portion of Galileo's information return from Jupiter will be images from the solid-state imaging (SSI) camera. Image frames contain either 800 × 800 or 400 × 400 8-bit pixels, and would require 5.12 Mbits or 1.28 Mbits, respectively, to transmit uncompressed. Image compression gains of 10 dB and higher are obtainable through known coding techniques such as the discrete cosine transform (DCT), which introduces hardly noticeable losses in data fidelity. However, Galileo cannot directly utilize off-the-shelf compression algorithms due to its limited onboard computing capability. Instead, Galileo will use a computationally tractable algorithm called the integer cosine transform (ICT). The ICT operates and performs comparably to the DCT. This article investigates the compression performance of Galileo's ICT algorithm, as applied to actual Galileo images and to simulated images constructed to mimic the conditions that the Galileo camera is expected to encounter at Jupiter.

The amount of compression achieved by the ICT is adjustable by selecting the degree to which the transform coefficients are quantized. Coarser quantization generally produces higher compression and larger errors in the reconstructed image. Thus, the study of ICT compression performance is in fact an analysis of the possible tradeoffs between compression of a source image and the resulting distortions in the reconstructed image. It is important to understand the extent to which the compression is predictable by using anticipated image characteristics, and how compression versus distortion varies with the selectable quantization step size. The overall statistical variability of the compression performance is an important

factor that influences link usage planning. This article is a preliminary attempt to examine these issues.

The ICT is described in [1] and [2], and the specific application of the ICT to the Galileo LGA mission is discussed in [3]. For a given quantization step size, quantization matrix or template, Huffman table, and, of course, image, the ICT produces a compressed image. The compression ratio indicates the ratio of the compressed image size to the original image size. Distortion indicates the difference between the reconstructed image and the original one. Neither the compression ratio nor the distortion is input for the ICT. This article describes the behavior of the compression ratio and distortion as a function of the quantization step size, q, and any prior information about the image.

An additional level of control may be achievable by selecting one of a few specially constructed quantization templates and using a corresponding Huffman table. For this article, the algorithm uses only a template for uniform quantization and a Huffman table that has been trained for the template on space images [3]. The tailoring and use of different quantization matrices is the subject of ongoing work, and will not be discussed here.

II. The Images in the Data Set

The ICT has been tested on various images. Ideally, the test images would be ones of Jupiter taken by Galileo's SSI camera. Instead, the available images are ones taken by Galileo of Venus, the Earth, the Moon, and black sky, and ones taken by Voyager of Jupiter and a few of Jupiter's satellites. There are four image sets provided for this study by JPL's Multimission Image Processing Laboratory (MIPL).

(1) The Earth 1 images are Galileo images from volumes 2 through 6 of the Earth 1 CD ROMs. They consist of 970 raw_cal, 1798 Earth, 81 Venus, and 773 Moon images for a total of 3622 images. A randomly selected one-third of the Earth 1 images has been used for this article to identify trends and performance predictors. The other two-thirds of the data has been reserved for future determination of the effectiveness of the predictors and the consistency of the trends.

(2) The 13 selected Galileo SSI images consist of 12 recent Galileo images and 1 of the Earth 1 images selected by MIPL to illustrate a typical range of image types. These images are designated by names beginning with rq534.

(3) The 19 simulated Galileo images were obtained from Voyager images of the Jovian system by removing Voyager camera characteristics and introducing the Galileo SSI camera signature. These images are designated by names r.1, r.2, ..., r.19.

(4) The 16 radiation-noise-added simulated Galileo images are 7 of the simulated images and 9 of the selected images with 1 of 4 available noise frames (A, B, C, and D) added by MIPL to simulate the radiation effects at Jupiter. These images are designated by names beginning with rq538.

Some examples of the selected Galileo images are shown in Fig. 1. The smaller data sets are useful for illustrating different behaviors but do not provide sufficient data for analysis. The Earth 1 images provide a larger number of images for drawing statistical inferences.

Images were compressed using different values of q, and reconstructed. Quantization step sizes 2 through 18, 20, and 24 were tested on all images. The compression ratio, root mean squared error (rmse), error variance, and error mean were noted along with image statistics such as the mean, minimum and maximum pixel values, the sample zeroth-order entropy, and the sample difference entropy. From this information and some a priori information about the image, patterns in the compression and distortion performance have been sought.

III. Performance as a Function of Quantization

This section shows the relationships between compression ratio and image distortion for the ICT as the quantization step size is varied.

A. Compression Behavior

The compression ratio achieved by the ICT depends on the image. This section shows the gross behavior with only the small representative image sets. Scatter plots illustrate the compression behavior as a function of the quantization step size, the spread of compression ratios for different images, the limiting compression for very large q, and the effect of the simulated radiation noise on the compression.

For the Huffman codes used in this work, there is a limit of 102.4 on the compression ratio achievable by the 8×8 ICT. This is because an all-zero quantized block in the transform domain can be coded in 5 bits, and an

uncompressed block is 8×64 bits. In order for a quantized block to be all zero, either the original block had to be nearly black or the quantization step size had to be very large. In either case, the reconstructed block would be entirely black. Among the images tested, the nearly black ones approached this limit with the quantization step sizes tested.

Figure 2 has three scatter plots showing the compression ratio versus quantization step size for the images in each of the three small data sets. The plots in Fig. 2 show that the compression trend is typically monotonic, increasing with q, but that it starts to level off for higher q. In most instances, as q increases, the spread on the log scale of the compression ratio decreases; this is consistent with the upper limit on compression ratio. The unlabeled scatter plots also show that without taking into account information about the specific image, it is difficult to predict the compression.

To illustrate the effects on compression of adding the simulated radiation noise to images, Fig. 3 has one plot for each of the four noise frames used. The plots show how the noise changed the compression ratio for the individual images. Solid symbols are used for images in the sets of selected Galileo SSI images or simulated Galileo images, and hollow symbols with the same shape are used for the corresponding images with radiation noise added. These plots seem to indicate that the noise frames dominate the compression performance; this is more apparent for frames A and B and less so for D.

B. Distortion

The rmse is a measure of the pixel-by-pixel difference between the original image and the reconstructed image. It is not the best measure of the distortion of an image, but it is easily quantifiable and commonly used as a guideline. A better measure of the distortion is how it affects the conclusions and analysis of the images sent back from Jupiter, but this is hard to quantify. The SSI team, consisting of scientists who will be using Galileo's images, has been meeting with both JPL and the Ames Research Center to try to determine what types of distortion are of more concern than others. The appropriate distortion measure for this analysis is not the subject of this article, and for now the rmse will serve as a rough guideline.

For a given image, the rmse typically increases as the quantization step size increases. The three scatter plots in Fig. 4 show rmse versus quantization step size for the three smaller image sets. Although there is a monotonic trend, there are examples of images for which the rmse is unexpectedly high for a particular q or q's. These high rmse situations may be avoided most of the time by careful selection of the quantization step size.

C. Selection of Quantization Step Size

For a particular image, there may be a quantization step size that is inferior, meaning that for that image there is another quantization step size that gives both better compression and better rmse. An inferior q causes a non-monotonic spike on the compression versus distortion curve. The plot in Fig. 5 shows the distortion versus compression ratio for the selected Galileo SSI images in Fig. 1, namely rq534.litn, rq534.ausvn, rq534.lun, rq534.pct, rq534.ear9n, and rq534.gas. The solid lines show the curves traced out by the rmse and compression ratios at all of the tested q's; the dashed lines show how the curves would be smoother and more predictable if the inferior q's could be predicted and avoided. Images that are predominately black sky, such as the image of Gaspra, rq534.gas, exhibit the most dramatic rmse fluctuations with q.

In order to get a reasonable profile for quantization step sizes that are likely to be inferior, the larger image set is necessary. The number of images for which a particular q was inferior was tallied over the randomly selected third of the Earth 1 images. Figure 6 is a histogram showing the frequency that different q's were inferior for the raw_cal or black sky images, and for the non-raw_cal images that are of the Earth, Moon, and Venus. The histogram shows that some q's are inferior more often than others, but sometimes only for certain subsets of the images. For instance, $q = 6$ is inferior for a quarter of the raw_cal images, but not often for the interesting ones. For $q = 13$, we see the opposite trend. Further study is important and should include a measure of how bad a q is for a particular image (as opposed to simply counting how often q occurs). Also, if the images will be classified into types that give improved prediction of the compression, then the study of which q's are inferior should be redone for each subset.

IV. Performance as a Function of Image Characteristics

Using the randomly selected one-third of the Earth 1 images, correlations were done between the camera settings and the compression ratio, the compressed image size, and the rmse in order to direct the study towards likely parameters for predictions, or at least away from

parameters that appear to be uncorrelated. The camera settings were not continuously valued, and so were not ideal for a correlation analysis. A portion of the correlation matrix of interest is shown in Fig. 7.

The filter width has a small amount of correlation with the performance. The middle wavelength of the camera filters seemed independent of the compression and the rmse. By dividing the selected data set into subsets defined by the different camera settings, statistics for each subset could be found. Scatter plots of compression versus rmse show the trends for different subsets. However, an analysis of variance test with a level of significance of $\alpha = 0.05$ indicated that the subsets probably come from distributions with the same mean.

The black sky images may need to be treated separately because compression performance effects that are second order for more interesting images are apparently first order for these images. For instance, the raw_cal images that are mostly black sky tend to cluster according to the gain state on the camera (not shown in this article). The effect is less noticeable for images with less black sky.

V. Overall Statistical Variability

With so many images in the Earth 1 image set, the compression and rmse distortion behavior can be displayed in histograms where the range of values is divided into bins and in cumulative distribution type plots. Figures 8 and 9 show the histograms and cumulative distributions of both compression ratio and rmse for the selected third of the images using $q = 8$. On the cumulative plots there are three curves: one is for the Earth, Moon, and Venus images; one is for the raw_cal images; and one is for all of the images. There are separate histograms corresponding to each curve on the cumulative plots. Histograms of this sort for different quantization step sizes can assist Galileo mission planners in deciding how many pictures to attempt to send back, and at what risk this can be done.

VI. Conclusions

This preliminary study has generally validated the assumptions regarding high image compression gains that the Galileo LGA mission is counting on. But it has also flagged several areas of concern that require further study.

The scatter plots of Figs. 2–4 and the histograms of Fig. 8 show that compression ratios of 10:1 to 40:1 are not unreasonable to expect at moderately small values of rmse. On the other hand, there is a statistical variability of the achievable compression ratio on the order of 5 to 10 dB from image to image, and slightly less variability in the rmse. The variability of compressibility can greatly complicate the job of planning how many images can be sent back and how large buffers have to be in the onboard computer.

The study of correlations between the known camera settings and the resulting compression performance suggested no significant correlation that could be used to appreciably improve the predictability of performance.

There needs to be more analysis of what can be calculated or predicted about each image to aid downlink planning and, perhaps, compression control. For instance, better predictors might be obtained from anticipatable or measurable image characteristics, such as the percentage of black sky expected to fill the image frame. Alternatively, an onboard calculation could be used to adjust the quantization step size over small portions of the image in order to target a particular compression ratio. An example of such a calculation of local image statistics is an estimate of the entropy using the buffered portion of the image awaiting compression. Such ideas are under examination.

Further improvement in the reconstructed images can be achieved by developing more sophisticated reconstruction techniques to exploit continuities across block boundaries and other spatial correlations in the image. There is ongoing work to develop special purpose quantization matrices for improved image quality at good compression ratios.

References

[1] W. Cham, "Development of Integer Cosine Transform by the Principle of Dyadic Symmetry," *IEE Proceedings*, vol. 136, no. 4, pp. 276–282, August 1989.

[2] K.-M. Cheung, F. Pollara, and M. Shahshahani, "Integer Cosine Transform for Image Compression," *The Telecommunications and Data Acquisition Progress Report 42-105, vol. January–March 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 45–53, May 15, 1991.

[3] K.-M. Cheung and K. Tong, "Proposed Data Compression Schemes for the Galileo S-Band Contingency Mission," *1993 Space and Earth Science Data Compression Workshop Proceedings*, Snowbird, Utah, pp. 99–109, April 2, 1993.

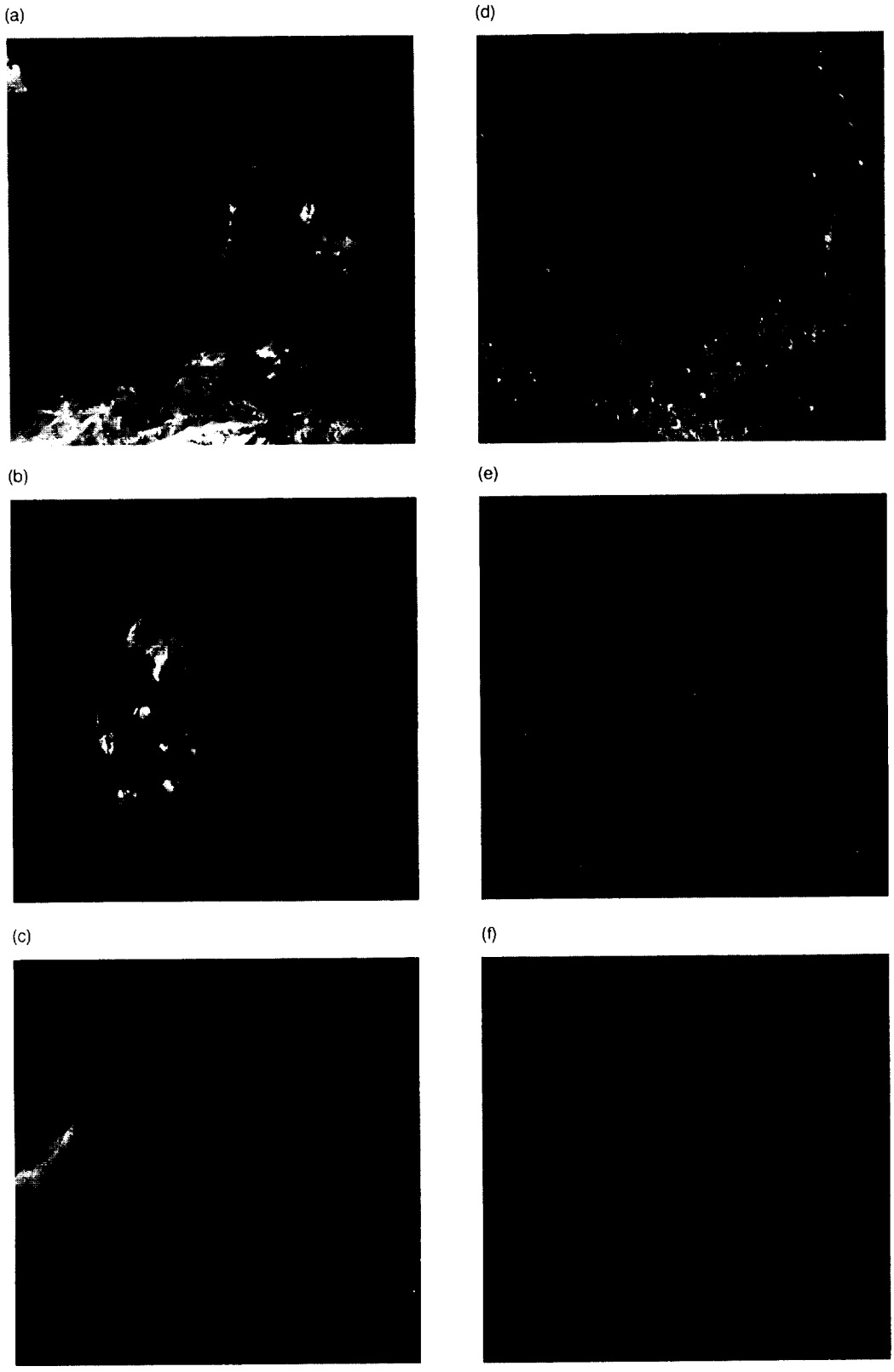


Fig. 1. Selected Galileo Images: (a) rq534.ausvn, (b) rq534.ear9n, (c) rq534.gasn, (d) rq534.lunvn, (e) rq534.litn, and (f) rq534.pctn.

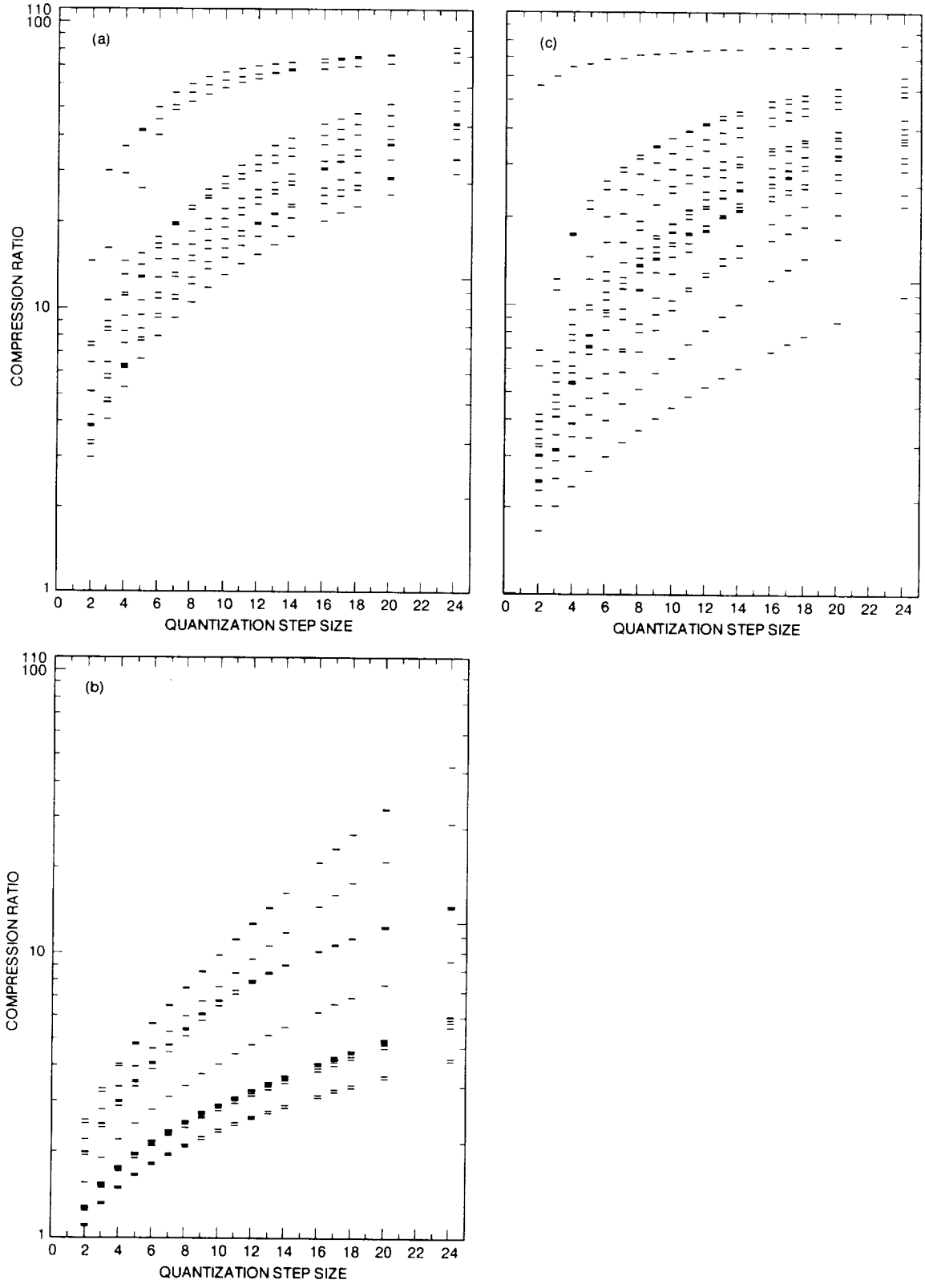


Fig. 2. Compression behavior for the small image sets: (a) compression ratio for selected Galileo SSI images, (b) compression ratio for images with radiation noise added, and (c) compression of simulated Galileo images from Voyager images.

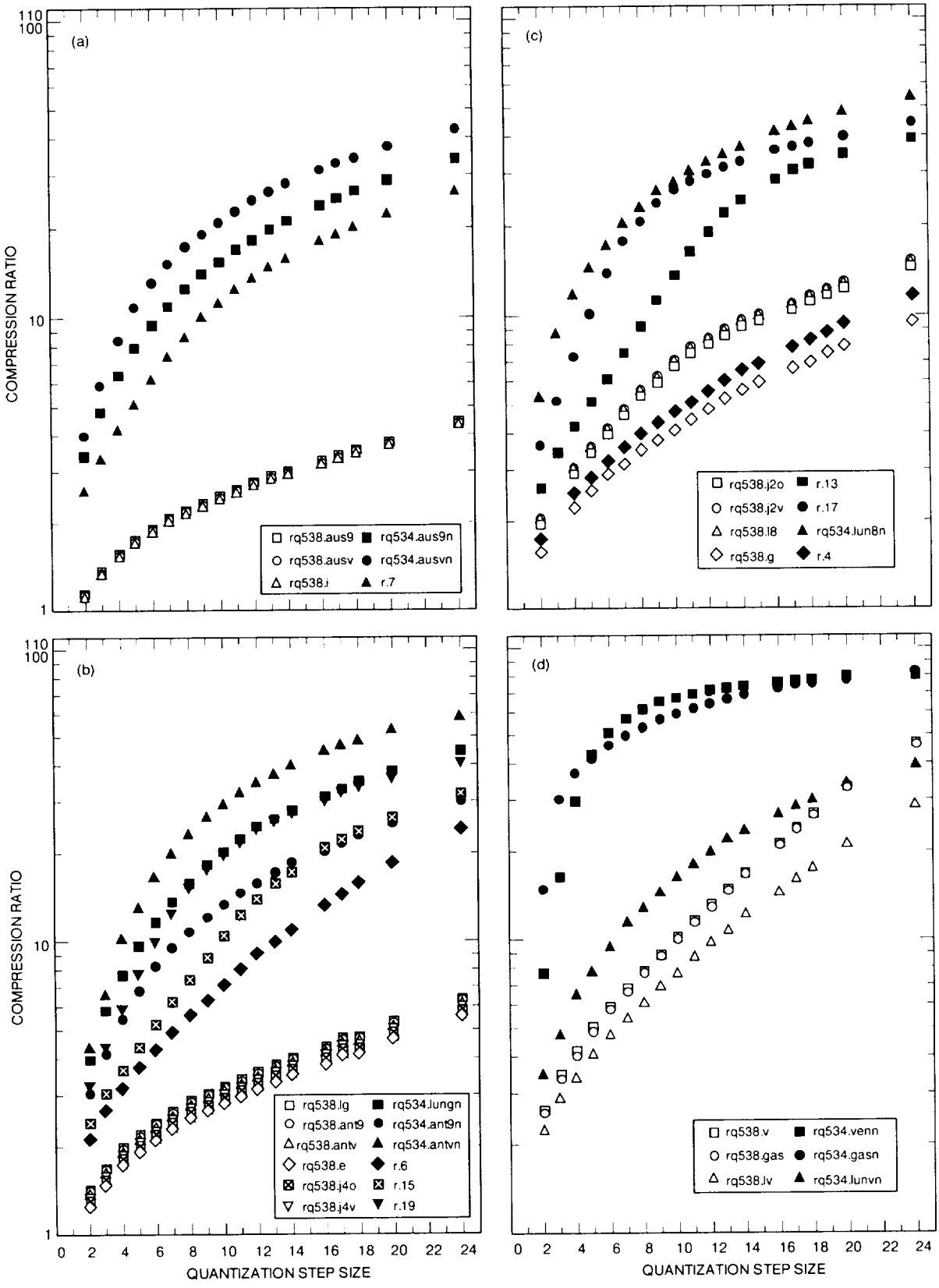


Fig. 3. Compression behavior for images with radiation noise added, showing the compression ratios for images with and without noise for (a) frame A, (b) frame B, (c) frame C, and (d) frame D.

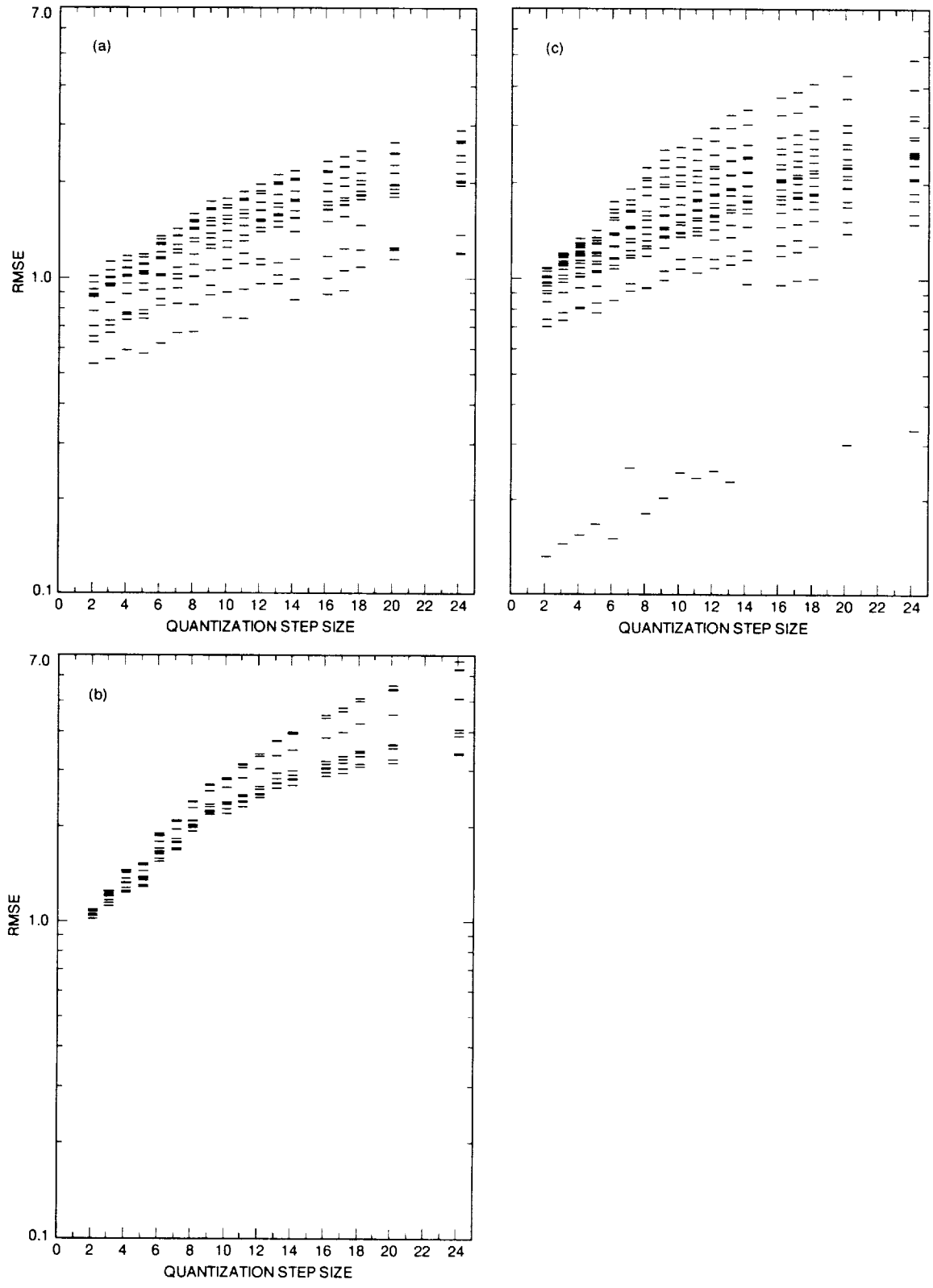


Fig. 4. RMSE distortion for the small image sets: (a) selected Galileo SSI images, (b) images with radiation noise added, and (c) simulated Galileo images from Voyager images.

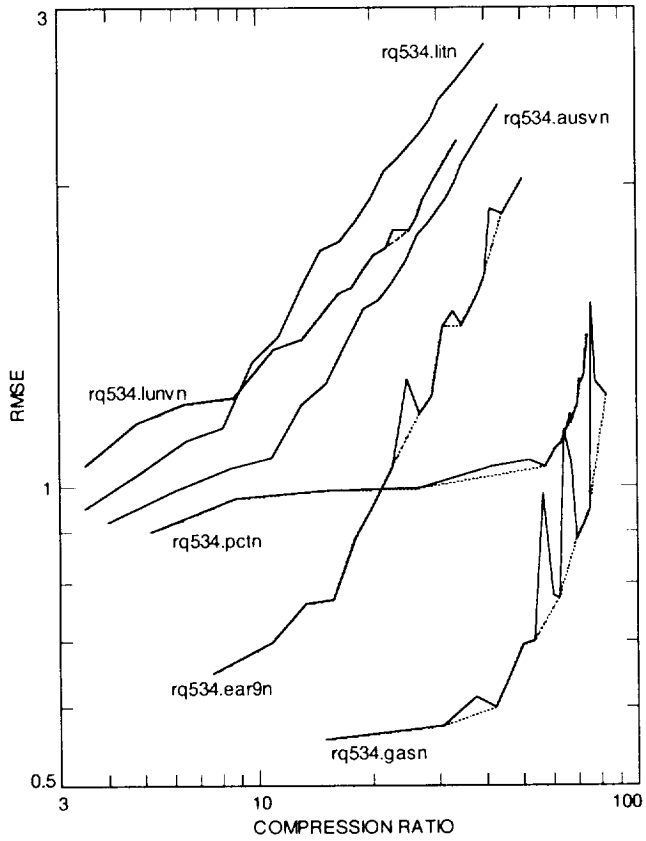


Fig. 5. Effects of inferior quantization step sizes on distortion versus compression curves.

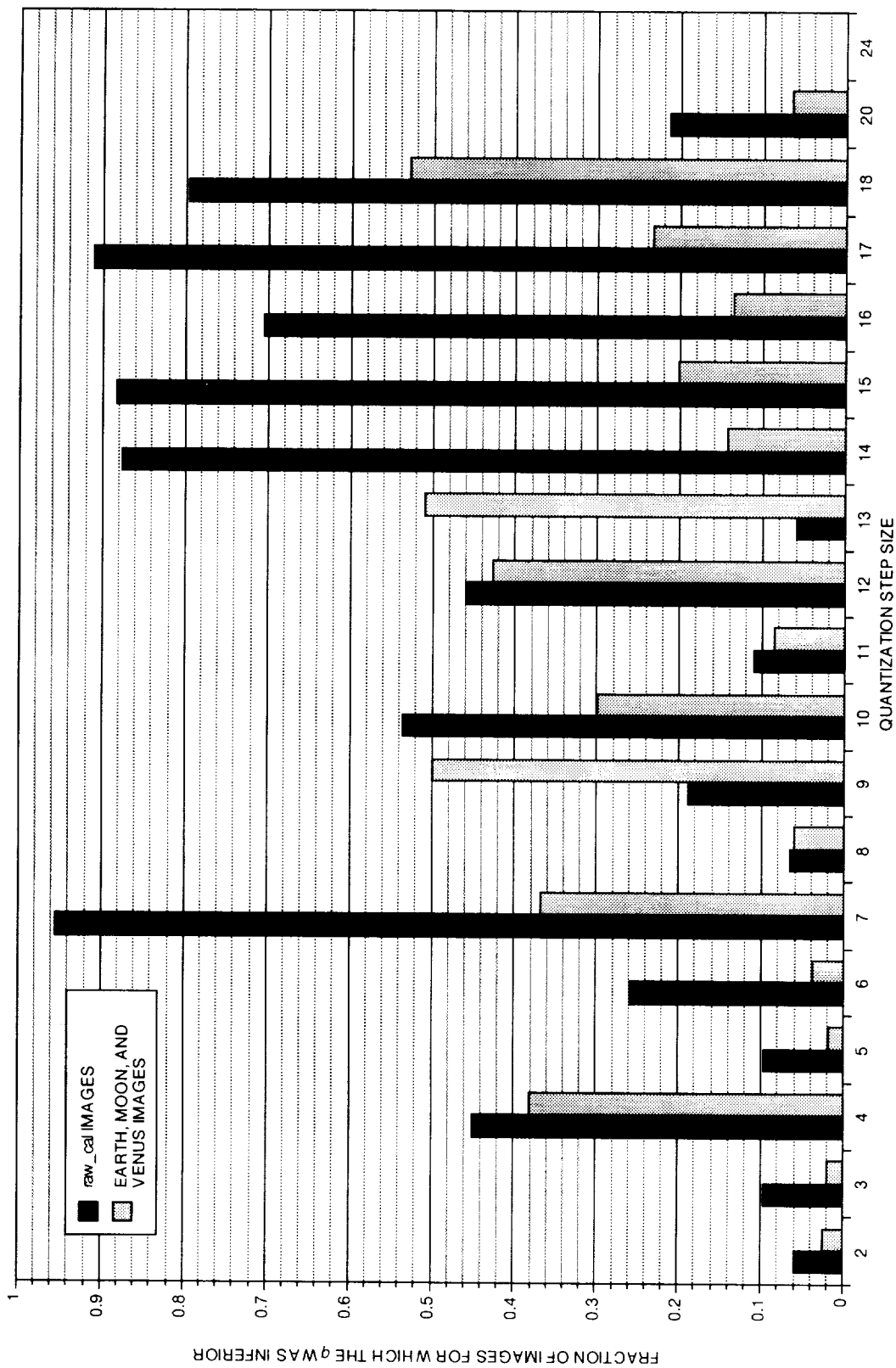


Fig. 6. Frequency of occurrence of inferior quantization step sizes over randomly selected images from the Earth 1 image set. 318 raw_cal images and 931 Earth, Moon, and Venus images were used to generate this histogram.

	filter mid-wave-length	filter width	exposure time	gain	frame duration
filter middle wavelength	1				
filter width	0.166	1			
exposure time	0.064	0.239	1		
gain	-0.32	-0.327	-0.1943	1	
frame duration	0.04	-0.228	-0.2214	0.025	1
mean	-0.04	0.196	0.62413	-0.08	-0.194
variance	0.02	0.032	-0.0763	0.1	-0.045
minimum	0.022	0.213	0.34753	-0.04	-0.595
maximum	0.071	0.214	0.09676	-0.22	0.0974
sample entropy	0	-0.266	-0.2223	0.023	0.2884
variance of differences	0.005	0.075	0.0184	-0.05	0.012
difference entropy	0.059	-0.045	-0.223	-0.23	0.2878
compression q=2	-0.04	0.151	0.25208	0.195	-0.359
compression q=3	-0.03	0.093	0.20209	0.145	-0.421
compression q=4	-0.03	0.089	0.18404	0.109	-0.433
compression q=5	-0.02	0.13	0.17664	0.062	-0.447
compression q=6	-0.01	0.201	0.17326	-0.01	-0.46
compression q=7	-0.07	0.223	0.14175	-0.06	-0.625
compression q=8	2E-04	0.282	0.16688	-0.1	-0.466
compression q=9	0.003	0.293	0.16243	-0.11	-0.466
compression q=10	0.003	0.301	0.15819	-0.13	-0.464
compressed filesize q=2	0.022	0.12	-0.1111	-0.24	0.1075
compressed filesize q=3	0.011	0.092	-0.0934	-0.17	0.1115
compressed filesize q=4	6E-04	0.016	-0.0938	-0.09	0.1278
compressed filesize q=5	-0.01	-0.058	-0.0951	-0.01	0.1396
compressed filesize q=6	-0.02	-0.097	-0.0956	0.035	0.1444
compressed filesize q=7	0.047	-0.026	-0.0809	0.025	0.2337
compressed filesize q=8	-0.02	-0.107	-0.0937	0.057	0.1425
compressed filesize q=9	-0.02	-0.101	-0.0914	0.057	0.1395
compressed filesize q=10	-0.01	-0.094	-0.0885	0.057	0.1345
rmse q=2	0.06	0.116	-0.2145	-0.36	0.1536
rmse q=3	0.063	0.125	-0.2009	-0.37	0.1604
rmse q=4	0.053	0.07	-0.2189	-0.33	0.1862
rmse q=5	0.047	0.073	-0.2123	-0.31	0.1888
rmse q=6	0.051	0.013	-0.2225	-0.27	0.2277
rmse q=7	0.085	0.06	-0.0063	-0.27	0.384
rmse q=8	0.014	-0.063	-0.2213	-0.17	0.2604
rmse q=9	0.024	-0.166	-0.2414	-0.13	0.3272
rmse q=10	0	-0.116	-0.2267	-0.11	0.2737

Fig. 7. A portion of the correlation matrix.

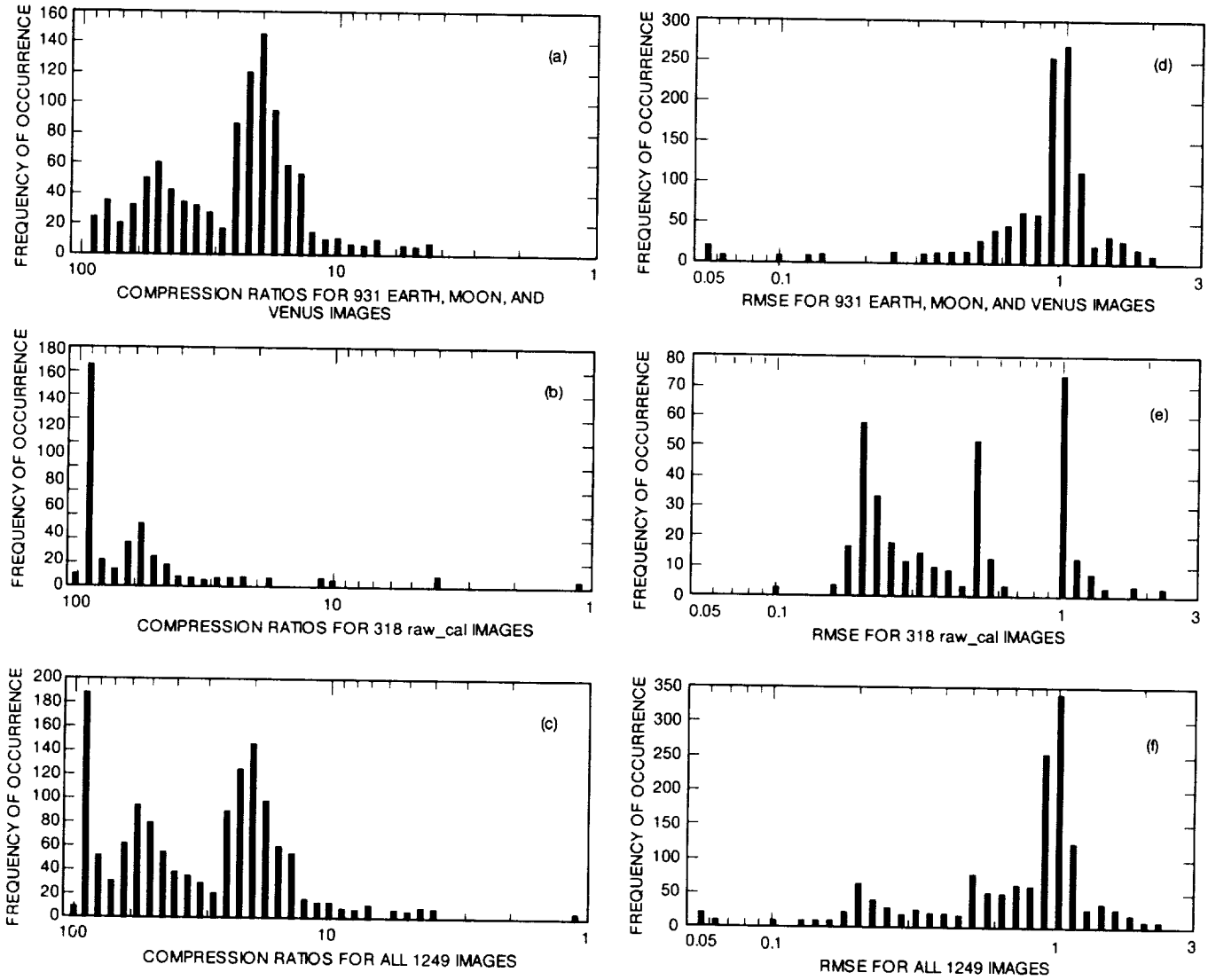


Fig. 8. Histograms of the compression ratios with quantization step size 8 for (a) 931 Earth, Moon, and Venus images, (b) 318 raw_cal images, and (c) all 1249 images, and of the rmse for (d) 931 Earth, Moon, and Venus images, (e) 318 raw_cal images, and (f) all 1249 images.

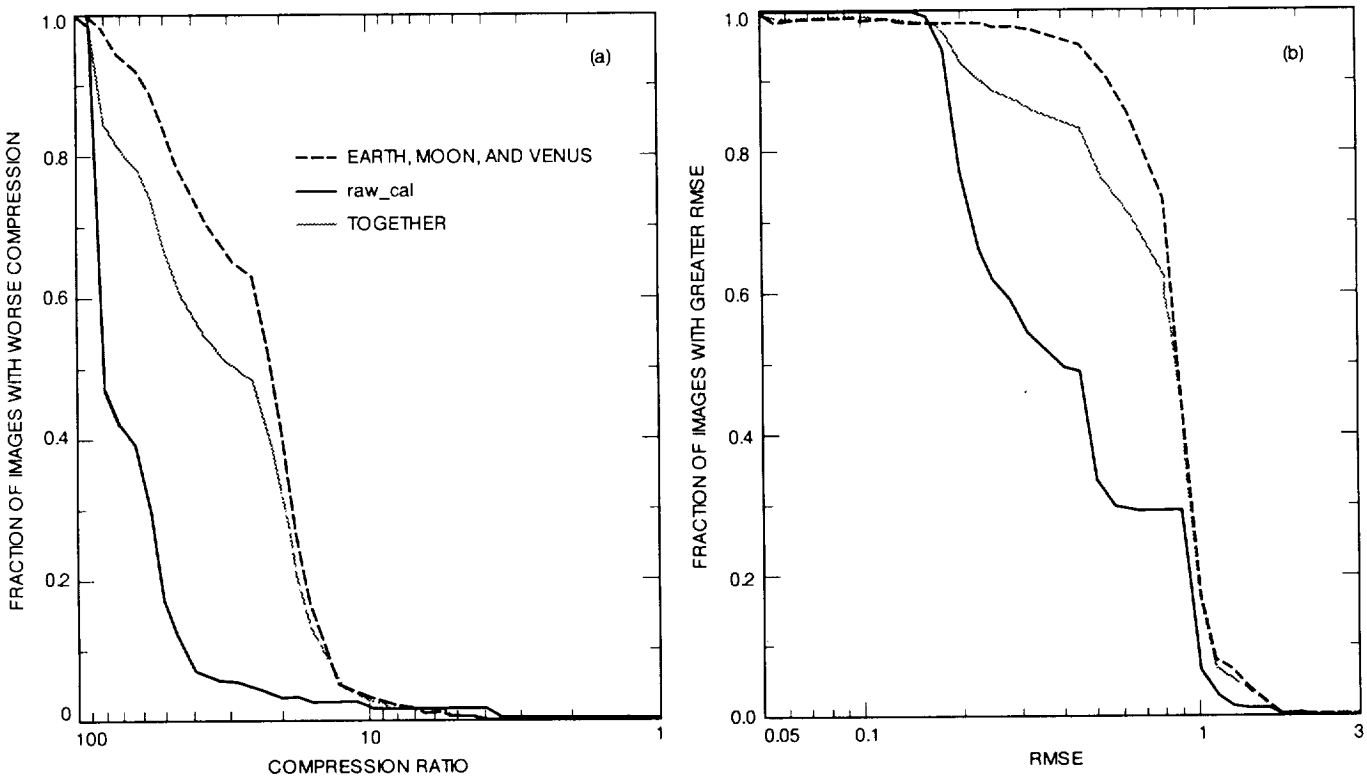


Fig. 9. Cumulative distributions for (a) compression ratios and (b) rmse for the randomly selected third of the Earth 1 images for quantization step size 8.

Link Monitor and Control Operator Assistant: A Prototype Demonstrating Semiautomated Monitor and Control

L. F. Lee and L. P. Cooper
Advanced Information Systems Section

This article describes the approach, results, and lessons learned from an applied research project demonstrating how artificial intelligence (AI) technology can be used to improve Deep Space Network operations. Configuring antenna and associated equipment necessary to support a communications link is a time-consuming process. The time spent configuring the equipment is essentially overhead and results in reduced time for actual mission support operations. The NASA Office of Space Communications (Code O) and the NASA Office of Advanced Concepts and Technology (Code C) jointly funded an applied research project to investigate technologies which can be used to reduce configuration time. This resulted in the development and application of AI-based automated operations technology in a prototype system, the Link Monitor and Control Operator Assistant (LMC OA). The LMC OA was tested over the course of 3 months in a parallel experimental mode on very long baseline interferometry (VLBI) operations at the Goldstone Deep Space Communications Center. The tests demonstrated a 44 percent reduction in precalibration time for a VLBI pass on the 70-m antenna. Currently, this technology is being developed further under Research and Technology Operating Plan (RTOP)-72 to demonstrate the applicability of the technology to operations in the entire Deep Space Network.

I. Introduction

The Jet Propulsion Laboratory (JPL) manages a worldwide network of antennas, the Deep Space Network (DSN), that provides a communications link with spacecraft. DSN operations personnel are responsible for creating and maintaining this communications link. Their tasks involve configuring the required subsystems and performing test and calibration procedures. The task of creating a communications link is known as precalibration and is a manual, time-consuming process that requires both operator input of more than a hundred control directives and monitoring of more than a thousand event messages and several dozen displays to determine the execution status of the system. The existing Link Monitor and Control (LMC) system requires the operator to perform a large number of textual keyboard entries and to monitor and interpret a large number of messages in order to determine the state of the system and to selectively identify relevant information from dozens of predefined, data-intensive displays. The tasks required by the LMC create an environment in which it is difficult to operate efficiently.

The goal of the Link Monitor and Control Operator Assistant (LMC OA) task is to demonstrate automated operations techniques that will improve operations efficiency and reduce precalibration time. The LMC OA is a knowledge-based prototype system that incorporates artificial intelligence (AI) technology to provide semiautomated[1] monitor-and-control functions to support operations of the DSN 70-m antenna at the Goldstone Deep Space Communications Complex (DSCC). The AI technology improves operations by using a flexible and powerful procedural representation, by reducing the amount of operator keyboard entries, and by providing explicit closed-loop communications and control through an expert-system module.

The precalibration process used for VLBI on the 70-m antenna was selected as the test domain for the prototype. The LMC OA was field tested at the Goldstone DSCC and performed a semiautomated precalibration for VLBI using actual operational equipment. The test demonstrated that precalibration time can be reduced by 40 percent with the LMC OA prototype.

The LMC OA has three major components: the Temporal Dependency Network (TDN), the Execution Manager (EM), and the Situation Manager (SM). These three components work together to provide a closed-loop, system-level control system for precalibration. The TDN is a directed network that represents parallel procedural paths, precedence relations, preconditions, and postconditions. The TDN is the primary knowledge base for the system. The EM is responsible for traversing the TDN and sending control directives to the subsystems while maintaining the precedence, parallel, and sequential constraints specified in the TDN. The SM works in step with the EM and provides the situational awareness necessary to close the control loop, to detect anomalies, and to support recovery from anomalies. The SM maintains an internal model of the expected and actual states of the subsystems in order to determine if each control directive is executed successfully and to provide feedback to the user.

This article describes the LMC OA prototype and test results. Section II describes the problems with the existing LMC system. The following sections will explain the LMC OA approach used to address the identified problems. The two major concepts that drive the LMC OA design will be presented, followed by a description of the TDN and the primary knowledge representation in the LMC OA. A detailed discussion of the three major modules and an operational scenario will be presented. In conclusion, the results of operational field testing and the lessons learned from this applied research project will be discussed.

II. Problem Overview

Currently, for standard operations, an operator is allocated 45 min[2] to perform a precalibration. In the case of more complex operations, like VLBI, an operator may be allocated much more time. Precalibration is a time-consuming process because of limitations in the existing operational monitor-and-control system. Precalibration is a command-line, keyboard-entry system that requires operators to manually send hundreds of directives to subsystems and monitor more than a thousand incoming messages on a text-based scrolling log. The system lacks explicit, informative responses about the state of a directive and does not provide guaranteed communications between the monitor-and-control system and subsystems being controlled. For each directive sent by the operator, the subsystem usually returns a *directive response*; this is simply an acknowledgment from the subsystem informing the operator whether the directive was received or rejected. A directive response does not indicate the success or failure of the directive's execution. The subsystem may also send out *event notice messages*, which relay information about the state of some device in a subsystem. These messages, however, are not explicitly tied to any directive sent. Operators, therefore, must rely on their experience to determine which directive was most likely to have caused the subsystem to send the event notice message. *Monitor data*, which are sent periodically by the subsystems, also provide information about device states. However, monitor data are never displayed automatically or tied to any directive. Instead, a subset of monitor data is formatted into predefined displays that the operators can call up. The operators then must decide which piece of the data they need and which display contains that piece of information. Oftentimes, a display contains many pieces of information of which operators only need one or two.

The inability of the monitor-and-control system to keep up with input from the subsystems causes messages to be dropped at monitor and control. To compound the problem, the subsystems cannot detect when a message has

[1] The LMC OA provides *semiautomated* precalibration because operator interaction is required. Precalibration currently requires several manual operations which could not be done through existing DSCC monitor and control interfaces. Furthermore, certain support and subsystem data were inaccessible to the LMC OA, thus requiring input from the operator.

[2] The standard time allocated for precalibration and postcalibration for each user or project activity is listed in Appendix A of the *DSN Scheduling Code Dictionary*, JPL document 842-204: 10-009, Rev. B (internal document), Jet Propulsion Laboratory, Pasadena, California, June 20, 1989.

been dropped and, thus, cannot resend information. This situation causes false alarms that can inundate the user with messages and often hide real alarm situations. Finally, the system is prone to input errors. A simple precalibration pass requires more than a hundred directives. It also requires the operator to manually identify and type each directive and its parameters. A subsystem, therefore, can take several minutes to recover from a simple typographic error.

Operators use a variety of support data—schedules, predict files, sequence of events, and pass briefings—to determine the type of pass, the spacecraft being tracked, and the method used to configure the communications and processing equipment. Information contained in the support data files is also used to determine the correct parameters for the control directives. Because these files are not available electronically for easy viewing and usage, the operator must refer to the hard copy version of these files and manually enter numerical parameters for control directives, when the numbers oftentimes are accurate to 10 decimal places. An entry error in any one of the digits could cause a major problem in the system.

The most difficult part of precalibration is the determination by the operator of what directives need to be sent and how the directives should be ordered. Currently, end-to-end representation of operations procedures does not exist. The documentation that is available addresses a specific subsystem or spacecraft or provides a general overview of an activity. As a result, operators must rely on their own experience to assemble an end-to-end operational sequence. Thus, the operational sequences vary from operator to operator, leading to inconsistencies in operations and making recovery from anomalies difficult.

The following are the specific operability problems identified with the existing LMC system.

(1) Extensive manual entry is required of the operator.

(2) The lack of integrated monitoring tools for the operator makes it difficult or nearly impossible to perform parallel operations. The operator must mentally interpret displays and text messages to determine correct execution of a directive.

(3) False alarms due to dropped messages occur frequently, and because dropped messages are not detected, they are retransmitted by the subsystem, giving the operator an incomplete picture of the system state.

(4) The lack of on-line access to usable support data increases the need to integrate information from multiple sources. Entry of complex numerical parameters increases the chances of typographical errors.

(5) There is no end-to-end representation of the operations procedures.

III. Closed-Loop Control and Situational Awareness

Two major design concepts found in the LMC OA system are closed-loop control and situational awareness. In the LMC OA context, closed-loop control means that all control actions (i.e., directives) have explicit feedback regarding the success or failure of the requested action. Under the existing monitor-and-control system, no single message can report the status of a directive. Rather, the operator must sift through many different data messages returned by the subsystems and many different displays to determine the status of the directive. Moreover, this present process of filtering and identifying pertinent data is time consuming. The LMC OA, however, integrates all available information sources and provides the operator with clear, consistent, explicit feedback for every control action.

Situational awareness, another feature of the LMC OA, allows the operator visibility into the state of the system and the state of procedure execution. In the current LMC, a large set of displays provides the operator with visibility into the state of the system. However, the information is difficult for the operator to interpret. Information important to the operator is not easily accessible because there are too many displays and none of them are user-definable. The LMC OA team did not redesign the displays because the resources to tackle such a significantly large problem were not available. Rather, the LMC OA prompts the user with the name of the display and the value to look for. In this manner, the LMC OA makes it slightly easier for the operator to determine the state of the system by explicitly providing the display name and monitor item to look at.

The second criterion for situational awareness is visibility into the state of procedure execution, which means that the operator knows the progress and status of procedure execution. Currently, since there does not exist end-to-end procedural documentation, the operator depends on experience to determine the procedure. To determine the state of procedure execution, the operator must interpret a large number of messages from the subsystem. However, through an extensive knowledge engineering effort, an end-to-end integrated procedure for VLBI was created and represented in a TDN. The TDN is a clear and intuitive way of representing the procedure to the user. Furthermore,

through the color-coded, graphical display of the TDN, the operator can immediately determine the execution status of the procedure.

IV. Temporal Dependency Network

One of the problems with the existing LMC is the lack of end-to-end procedural documentation. To perform a VLBI precalibration, the operator must refer to several operation manuals which describe individual subsystems or portions of the procedure. The operator must then manually create an integrated procedure. In some cases, operators create and use, as a reference, personal "cheat sheets" that describe what needs to be done. The lack of a single source of documentation that describes the VLBI precalibration procedure results in inconsistent operations. Actual operations rely heavily on an individual operator's experience and expertise. To automate operations, an integrated procedure for VLBI precalibration was created.

The approach to knowledge engineering involved first learning about the system through existing documentation and noting inconsistencies and missing information. The next step involved discussing the procedure with operators, engineers, technicians, and scientists to get their viewpoints and to clear up inconsistencies as much as possible. This led to the development of an initial TDN. The TDN became the much-needed common language between the knowledge engineers and the knowledge sources. This LMC OA knowledge engineering effort is the only known attempt, within the DSN, to produce a single, coherent, and consistent baseline operational sequence for precalibration that merges the viewpoints of all users.

The TDN, shown in Fig. 1, illustrates an end-to-end operational sequence for VLBI. Sequential, parallel, and optional operation sequences are identified in the TDN. Each block in the TDN contains directives that are sent to the subsystems sequentially. Blocks have precedence constraints where the directives cannot be sent until all of its predecessor blocks' directives have successfully completed execution. Each block has associated precondition and postcondition constraints. These constraints define the state the system must be in *before* starting each block of directives and *after* successful execution of those directives, respectively. Each block may also have temporal constraints that limit the start and completion of the directives to a specific time or time interval.

V. LMC OA Design

The goal of the LMC OA is to provide both closed-loop control and closed-loop communications for the operator. There are two major modules in the LMC OA: the TDN Execution Manager (EM) and the Situation Manager (SM). Other modules that will be discussed include the Block Execution module, Router, Monitor Data Handler, and DSN Data Simulator. An overview of the design is presented in Fig. 2.

A. TDN Execution Manager and Block Execution Modules

The TDN EM traverses the TDN identifying blocks that are ready to execute. Blocks whose precedence constraints are satisfied are started. When a block is started, the user is asked to parameterize any unparameterized directives. The preconditions are then evaluated by the SM. A block's directives are sent only after the SM verifies that the preconditions are satisfied. Once a directive is sent, a directive response must be received before the next directive in the block can be sent to a subsystem. After the last directive is sent and its corresponding response is received, the block's postconditions are checked by the SM. If the postconditions are satisfied, the block of directives is considered completed.

B. Situation Manager

The SM provides situational awareness within the LMC OA. It is also an AI-based module that verifies correct execution of blocks of directives by checking postcondition constraints. Problems can be detected and simple recovery assistance provided. To keep track of the state of the system, the SM keeps an internal model of all hardware and software devices that can be monitored in the system. Each device represented in the model has attributes that reflect the state of the device. Each attribute has a pair of values: an *expected value* and an *actual value*. The expected value of an attribute, in the form of a postcondition, is set when a directive is sent to the subsystem. The actual value of an attribute is set when the subsystem sends messages noting state changes in the subsystem. Every directive sent to a subsystem is expected to cause certain known changes on the states of the devices in the subsystem. Each time a directive is sent, the expected values of the attributes in the device model are updated.

In addition, several data types are used to set the actual values of the device attributes: event notice messages, directive responses, monitor data, and operator input. Event notice messages describe explicitly the actual states of devices. Directive responses provide information on whether the directive has been received by the subsystem. In some cases, these responses also provide progress and completion data. Monitor data are blocks of status information that are sent periodically by the subsystems. Monitor data

usually provide more information than event notice messages. In certain situations, operator input is requested. Although the operator is provided a set of predefined monitor displays, the information in these displays is not always available from the monitor data blocks. These displays are generated as bit-map displays at the subsystem level and are unavailable to the LMC OA because of format and DSN operational restrictions. Therefore, for certain directives, the operator must obtain information from the displays and enter it into the LMC OA. This information is used to set the actual value of an attribute in the SM internal device model. All four electronic data types provide information about the actual state of a device, but they do not give explicit information about whether a directive was executed successfully. However, by using information about the expected and actual states of devices, the success of a directive can be inferred. With the SM maintaining its device models, information about the state of the system and the state of the procedure is always available to the operator.

C. Router, Monitor Data Handler, and DSN Subsystem Simulator

In addition to the TDN EM, block execution, and SM modules, there are several other supporting modules. The Router handles all communications between the LMC OA and the DSN subsystems and serves as a translator between the DSN 890-132 protocol and the LMC OA internal data representations. It receives and decodes input from the DSN subsystems and directs the input to either the TDN EM, the SM, or both. It also formats the directives into communication packages that are sent to the subsystems. The Monitor Data Handler receives Monitor Data blocks from the subsystems and stores them in the Monitor Data database. Since access to the operational environment is limited, a DSN Subsystem Simulator was implemented to simulate the directive responses and event notice messages from the subsystems for testing.

D. User Interaction and Status Displays

One of the LMC OA goals is to provide consistent interaction and meaningful displays to keep the user aware of what is transpiring in the system. The primary method of interaction is through menu or button selections using a mouse. The operator may be asked occasionally to enter a value or response. The primary interaction window is a block-level display of the TDN and provides a high-level, end-to-end sequence of operations. A color bar in each block shows the status and progress of each block: a gray bar means the block is inactive, a green bar means the directives are executing, a red bar means an anomaly has occurred, and a blue bar means the directives have been completed successfully. The portion of the color bar that is green is proportional to the number of executed directives in the block.

The operator can call up a lower-level display for each block that lists the preconditions and postconditions for each block and shows the state of the block and the state of each directive in the block (inactive, executing, paused, anomaly, etc.). At the TDN level, there are controls to pause, resume, and stop execution. Block-level and directive-level controls allow the user to pause, resume, and skip execution. Icons are used to show the user whether a block is paused or skipped.

The SM anomaly messages that require a user response are displayed in a separate window. A synopsis is displayed in a scrolling portion at the top of the window. By selecting a synopsis, the operator can display a description of the anomaly in the bottom portion of the window. The operator can then enter the requested input or select a default option.

The scrolling event log lists all the input to and output from the LMC OA system. A command line window allows operator control outside of the TDN. Another display shows the end-of-pass report as it is being filled in by the LMC OA. With the existing LMC system, the operator must write down the time at which certain directives were executed and their results. At the end of the pass, the operator must also write a set of paper reports. The LMC OA system, however, internally logs the time, parameters, and responses for each directive and automatically generates reports.

VI. Operational Scenario

A typical operations scenario using the LMC OA follows: The operator starts the LMC OA system and selects a specific precalibration task, like VLBI. The corresponding TDN and knowledge bases are loaded, and the TDN is graphically displayed. The operator then enters the specific parameters for the next pass based on the support data. Directives that require real-time data input, like weather information, contain place holders for parameters. The operator can also tailor the TDN, skipping any unnecessary blocks, entering special directives, or establishing break points, as needed. The process of preparing the TDN for a specific pass can be done at any time preceding the designated pass start time. At the start of the pass, the operator selects the start option by a single click of the mouse to start execution of the TDN. The TDN can be paused or halted at any time during the process. The

operator watches the execution of the TDN by following the color coding in the graphical user interface.

At any time, the operator may bring up low-level displays to see the execution state of the individual control directives. The low-level display is updated automatically when each directive is sent and when completion is verified. The SM and TDN EM work in tandem to ensure that the control directives are correlated with the monitor data and event messages. This correlated information is then summarized and presented to the operator. If the SM detects a problem, it reports the problem along with recovery suggestions to the user. The user selects a recovery option which will cause the TDN execution to continue or halt the execution of the TDN. A command window is provided so the operator can enter any control directives into the system. Another window displays a scrolling log of all incoming directives, directive responses, and event notice messages. Additional windows provide a pass summary report and link status. During execution of the TDN, the operators are able to view the detailed subsystem displays on the LMC console. (These displays were not reimplemented in the LMC OA due to resource constraints.)

VII. Results

The LMC OA prototype was tested at the Goldstone DSCC's 70-m antenna while performing a VLBI precalibration procedure. The LMC OA was successfully tested over a 3-month period at the Goldstone DSCC. The tests were made in conjunction with maintenance and, despite interruptions, the LMC OA performed a VLBI precalibration in 27 min, compared with the standard time of 45 min. This is a 44 percent reduction in precalibration time. In addition, the LMC OA reduced the number of operator-entered directives from more than a 100 to 0 directives and 14 parameters.

The LMC OA prototype was implemented using Objective-C, Interface Builder, and the C Language Integrated Production System on a NeXT workstation running the MACH operating system. In addition, a 386 personal computer (PC) running translation software served as the gateway between the DSN network running proprietary protocols and the NeXT workstation running TCP/IP. The PC was equipped with an IEEE 988 card which could communicate with an Ungermann Bass Interface (UBI) Network Interface Unit (NIU) running DSN proprietary protocols. The PC was also equipped with an Ethernet card running TCP/IP and the PC-NFS package, which provided socket communications to the NeXT workstation. The translation software running on the PC gateway was developed by a previous project and modified by the LMC OA team as a gateway between the NeXT workstation and the DSN network.

VIII. Lessons Learned

Many elements contributed to the success of the LMC OA system, yet there were difficulties to overcome. This section examines both successes and difficulties.

(1) *The LMC OA prototype successfully applied AI technology to provide semiautomated precalibration.* The LMC OA prototype focused on two major concepts: closed-loop control and situational awareness. The LMC OA prototype demonstrates that with the right technology precalibration can be performed in significantly less time.

(2) *The TDN is a powerful and flexible procedural representation.* The TDN is powerful enough to represent end-to-end operations, constraints, and parallel paths. The TDN can also be used to handle contingencies and anomalies. It is flexible enough to be easily changed and can be adapted for other domains. The representation is simple enough so that it is easy to encode internally and easy to explain to the users.

(3) *The TDN, in addition to being a procedural representation, is a valuable knowledge engineering tool.* The TDN provides knowledge engineers and operators with a common focal point to work from. Its intuitive nature makes the format easy to understand and easy to use by both the operators and the knowledge engineers.

(4) *Station operators and JPL training and engineering personnel have valuable knowledge and experience that should be used.* The successes described are due to the excellent support provided by the personnel at the Goldstone DSCC, CTA-21, and other JPL personnel in training, operations, and engineering. Because of their experience, they have a wealth of information, but it is not yet documented. This information is critical to developing the TDN.

(5) *Knowledge engineering must be performed at the very beginning of the project.* The knowledge engineering process is an important part of building a knowledge-based system. The bulk of the LMC OA development effort was spent in knowledge engineering.

(6) *Documentation must be kept current, operability standards must be enforced, and documents must be*

integrated. A directive dictionary and device models should be provided by the subsystems. In the knowledge engineering process, knowledge engineers often found out-of-date documentation and operational modifications documented in operators' personal notes. This leads to error-prone and inconsistent operations. Furthermore, standards specifying format and content must be followed in order to ensure that each document contains information as expected and at the same level of usefulness and quality. Documents must be integrated and cross-referenced. There are hundreds of documents, and not having some form of cross-reference to easily identify the documents' contents makes operations and knowledge engineering difficult. Providing a directive dictionary and device models for each subsystem can identify many of the side effects that are currently being ignored.

(7) *The TDN should be used as end-to-end procedural documentation.* In the process of creating the TDN, the LMC OA knowledge engineers found many documentation manuals but none that contained descriptions of integrated, end-to-end procedures. End-to-end procedural information is necessary for efficient operations. The TDN is an effective tool for representing this type of information.

(8) *On-line documentation and capabilities to search the documentation should be provided.* This will assist with fast and efficient searches.

(9) *Complete access to monitor-and-control data must be provided.* This is essential for implementing closed-loop control, which in turn enables automation. The current environment does not provide electronic access to all data. Naming and usage of data items are inconsistent. There is also inconsistency in the meaning of directive responses. A centralized, automated monitor-and-control system must have electronic access and the ability to manipulate monitor-and-support data, reliable and guaranteed communications, and remote monitor and control of subsystems.

(10) *A data management strategy to store and easily retrieve data should be developed.* Information that needs to be stored includes monitor-and-event data, knowledge bases, support data, directive libraries, and TDN libraries.

(11) *An environment where both systems and knowledge verification testing can be performed needs to be built.* There are three types of testing for a knowledge-based system: compatibility, system, and knowledge verification. CTA-21 was used for network compatibility testing. However, CTA-21 no longer has a full suite of test equipment, and this makes it impossible to perform system or knowledge verification tests. Systems testing and testing of the LMC OA functions were conducted at JPL, using a subsystem simulator, and at DSS 14. Knowledge verification could only be done at DSS 14 because equipment was unavailable at CTA-21.

(12) *Visibility into the system and training tools that allows the operator to maintain operational skills must be provided.* There is no such thing as a fully automated system. When unforeseen events happen, operators must know what the system is doing so they can take over operations when the automated system is out of its league. Operators must have flexible control and the ability to completely override the system. Embedded training will provide operators with the mechanism for maintaining and further developing their analysis and problem-solving skills in an operational environment.

(13) *Smarter subsystems should be built.* Subsystems should be able to perform their own self-test and anomaly detection, isolation, and recovery whenever possible.

IX. Related Efforts

The technology demonstrated in the LMC OA can be extended to other operations in the DSN. Current efforts include extending the LMC OA to control multiple activities as well as using the system as the cornerstone of an operations automation thrust at DSS 13, the 34-m experimental antenna.

An architectural study of the DSN specifies that in the future one operator must be able to monitor and control multiple activities. At issue is the question of how much information an operator can process and how to organize the enormous amount of data so that the operator can at all times manage multiple tasks. The prototyping effort will involve the development of intelligent user interfaces and advanced data management systems. The current effort, sponsored by NASA Code O under Research and Technology Operating Plan (RTOP)-73 and NASA Code C under the AI-RTOP, is researching and identifying what technology is required to provide a multilink monitor-and-control capability for the operators. In 1994, the identified technology will be incorporated into the LMC OA to provide a semiautomated, multilink monitor-and-control capability.

The LMC OA is the starting point for an effort to demonstrate a systems approach to automation in the

DSN. The focus of the newly created RTOP-72, DSN Station Operability, is to research and identify technologies that will support automated and remote operations for the DSN. The technology will be demonstrated in the development of an automated monitor-and-control system with remote capabilities at DSS 13. The DSS-13 baseline monitor-and-control system, delivered for operational use in 1993, uses standard protocols such as Open Systems Interconnection (OSI), Manufacturing Messaging Specification (MMS), and commercially available packages such as RTWorks from the Talarian Corporation. The data communications and access infrastructure provided by the DSS-13 Monitor and Control system provides a strong baseline for testing automation concepts. In 1994, RTOP-72 will develop a systems approach to incorporate link- and subsystem-level automation. In addition, RTOP-72 will: (1) implement and deliver the LMC OA for DSS-13 operations; (2) develop an automated station monitoring prototype based on the Multimission Automation for Real-time Verification of Spacecraft Engineering Link (MARVEL) system; (3) develop a plan for integrated data management services, including the identification of data required for automation and improved operability; and (4) develop a prototype of a link health and performance monitoring system.

X. Conclusion

Knowledge-based systems will play a major and enabling role in improving operability and capabilities of future ground systems at the DSN. The LMC OA prototype demonstrates the feasibility and benefits of AI-based automation in DSN operations. The benefits of an operational, semiautomated monitor-and-control system are (1) reduction in precalibration time; (2) reduction in keyboard entry, which reduces occurrences of typographic errors; (3) capability of parallel operations; and (4) increased operator efficiency via closed-loop control. The LMC OA system demonstrates several operational improvements. It provides the operator with mechanisms for closed-loop control and situational awareness. It provides an end-to-end procedural representation for precalibration using a TDN. And it reduces the number of keyboard entries required by the operator. Furthermore, current efforts are showing that this technology is applicable to the DSN as a whole.

Acknowledgments

Past and present team members who have contributed to this version of the LMC OA are Kristina Fayyad, Randall Hill, Juan Urista, Kathryn Sturdevant, Jay Wyatt, Crista Smyth, and Patricia Fernandez. The contributions of Pam Wolken, Terry Dow, Dave Girdner, Dan Kiewicz, and the many operations personnel at the Goldstone Deep Space Communications Complex are gratefully acknowledged.

Appendix
Bibliography

L. P. Cooper, R. Desai, and E. Martinez, "Operator Assistant to Support Deep Space Network Link Monitor and Control," SOAR Symposium, Houston, Texas, August 1991.

L. P. Cooper, "Operations Automation Using Temporal Dependency Networks," Technology 2001, San Jose, California, December 3–5, 1991.

K. E. Fayyad and L. P. Cooper, "Representing Operations Procedures Using Temporal Dependency Networks," Space Ops '92, Pasadena, California, November 1992.

R. W. Hill, and L. F. Lee, "Situation Management in the Link Monitor and Control System," Space Ops '92, Pasadena, California, November 1992.

L. F. Lee and R. W. Hill, Jr. "Process Control and Recovery in the Link Monitor and Control Operator Assistant," Space Operations, Applications, and Research Symposium, Johnson Space Center, Houston, Texas, August 1992.

NeXT Operating System Software, Redwood City, California: NeXT Computer, Inc., 1990.

J. Urista, "Using MACH Threads to Control DSN Operational Sequences," Space Ops '92, Pasadena, California, November 1992.

B. Wilkinson, "Operability Engineering in the Deep Space Network," Space Ops '92, Pasadena, California, November 1992.

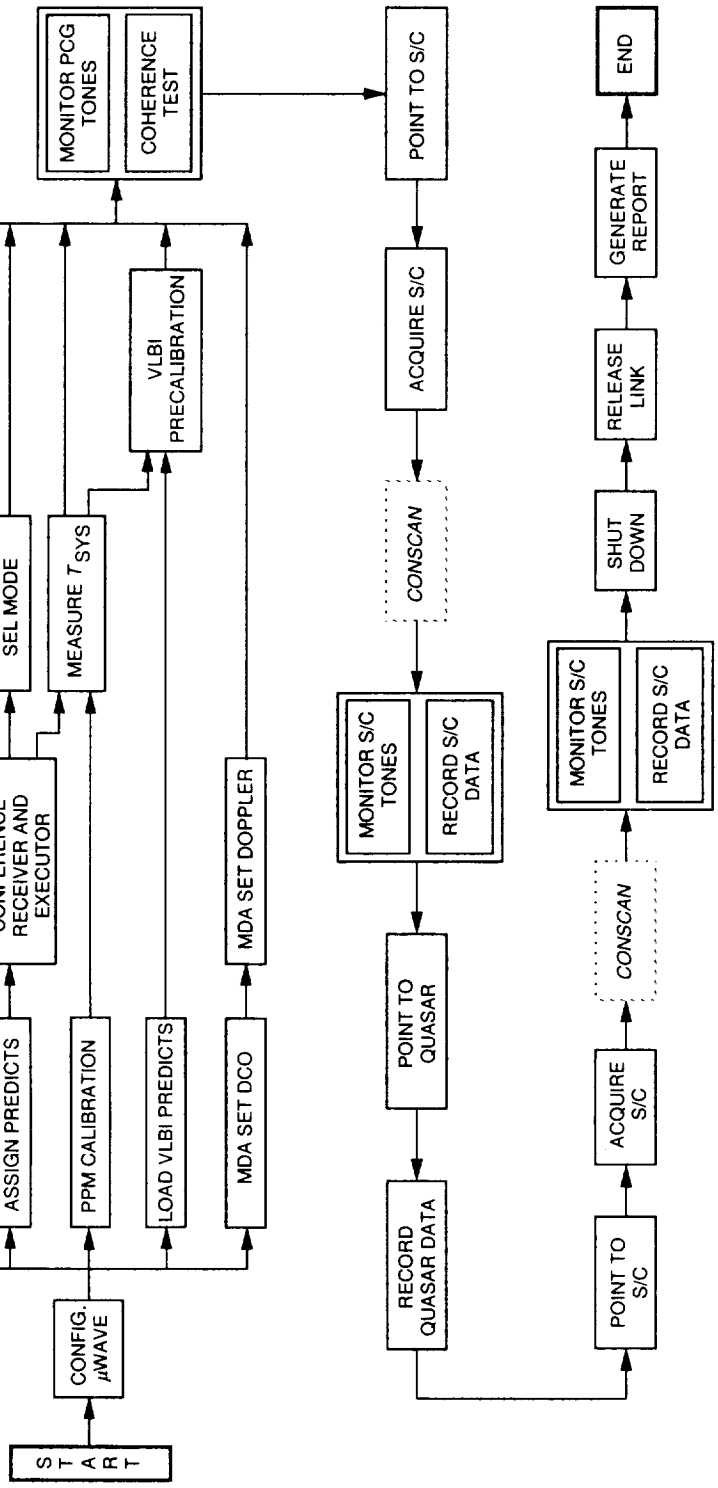


Fig. 1. A high-level VLBI temporal dependency network.

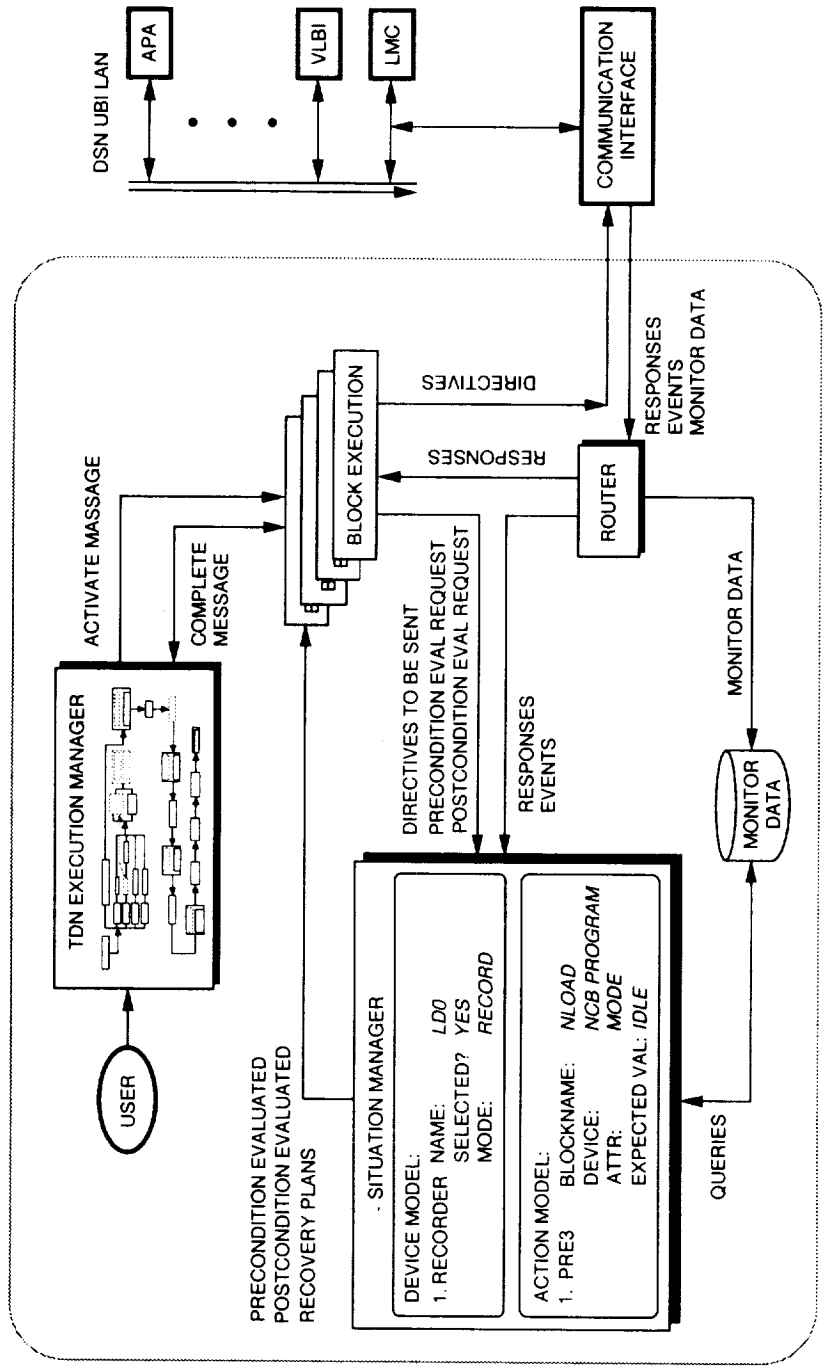


Fig. 2. The LMC OA design.

Beam-Waveguide Antenna Servo Design Issues for Tracking Low Earth-Orbiting Satellites

W. K. Gawronski and J. A. Mellstrom
Ground Antennas and Facilities Engineering Section

Upcoming NASA missions will require tracking of low-orbit satellites. As a consequence, NASA antennas will be required to track satellites at higher rates than for the current deep space missions. This article investigates servo design issues for the 34-m beam-waveguide antennas that track low-orbit satellites. This includes upgrading the servo with a feedforward loop, using a monopulse controller design, and reducing tracking errors through either proper choice of elevation pinion location, application of a notch filter, or adjustment of the elevation drive amplifier gain. Finally, improvement of the signal-to-noise ratio through averaging of the oversampled monopulse signal is described.

I. Introduction

Future NASA missions will include low-orbiting satellites with significantly higher antenna tracking rates, as compared with the deep space missions. Thus, the JPL/NASA antenna servos should be upgraded to be able to follow commands at higher rates. A feedforward upgrade, discussed in [1], is the simple and reliable choice. For tracking, a monopulse controller is an alternative to the existing conscan tracking, since the former is much faster than the latter. The design and performance of a monopulse controller is discussed. It is shown that its performance can be improved through proper choice of the location of the elevation pinion, the implementation of a notch filter, or the adjustment of the amplifier gain. Finally, the improvement of the signal-to-noise ratio (SNR) of the monopulse signal is presented. By averaging the redundant monopulse samples, the SNR improvement is from 7 to 17 dB.

II. Feedforward Controller Design

Tracking accuracy of fast moving objects can be improved if a proportional-and-integral (PI) control system is augmented with a feedforward loop [1], shown in Fig. 1. In this diagram, G_p, G_c, G_f, and G_w denote transfer functions of the antenna's rate loop, PI controller, feedforward gain, and wind disturbance, respectively; r is a command; y is output (elevation and azimuth angles); e is tracking error in azimuth and elevation; u is plant input; and w is wind disturbance. Almost perfect tracking ($e \cong 0$) in the absence of disturbances is obtained for the feedforward gain G_f such that $G_f = j\omega I_2$, c.f., [1].

The closed-loop transfer function (elevation encoder to elevation command) for a system with and without the feedforward gain is compared in Fig. 2. The figure shows that for frequencies up to 1 Hz, the system with the feedforward gain has superior tracking properties as

compared with the system without feedforward gain. This is confirmed by tracking simulations with a trajectory, as in Fig. 3. The DSS-13 antenna, with proportional gain $k_p = 0.5$ and integral gain $k_i = 1.8$ in azimuth and elevation, was investigated. For the servo with the feedforward loop, the error of 1.4 mdeg in elevation and 0.2 mdeg in cross-elevation was observed, which exceeds the requirements. However, for this controller, the high-frequency components of the command are strongly amplified, as can be observed from the transfer function plots in Fig. 2, where the resonance peaks of the system with feedforward gain are much higher than the ones of the system without feedforward gain. As a result, any sharp change in the command may cause excessive vibrations of the antenna.

Despite the increased sensitivity to the command inputs, the disturbance rejection of the antenna with a feedforward loop remains the same as that for the antenna without a feedforward loop. Thus, the pointing errors due to wind gust disturbances are comparable to the results obtained for the DSS-13 antenna with the PI servo.

III. Monopulse Controller Design

In monopulse tracking, a deviation of an antenna from a target is detected by four slightly displaced feedhorns, each receiving the signal from a slightly different angle. The received beams are added and subtracted to form a sum and a difference beam. The difference beam is zero when the target is on the antenna boresight, and the nonzero difference beam produces an error signal, which is used by the monopulse control system. A detailed description of the monopulse technique is given in [2–5].

The monopulse tracking control system is shown in Fig. 4. It consists of the plant, the monopulse feedhorns, and the monopulse controllers in azimuth (H_a) and elevation (H_e). The plant in this case is the antenna with the closed encoder position loop. The monopulse feedhorns detect the tracking errors in azimuth (e_a) and elevation (e_e). The encoder command is denoted r_a in azimuth and r_e in elevation. The feedhorns detect the tracking errors e_a and e_e directly, and the output signals y_a and y_e, as well as the commands c_a and c_e, are not available. Note that y_a and y_e signals are not the encoder output, but the antenna positions related to the focal location of the RF beam.

Denote the two-input, two-output plant transfer function G

$$G = \begin{bmatrix} G_{aa} & G_{ae} \\ G_{ea} & G_{ee} \end{bmatrix} \tag{1}$$

and introduce the following notations:

$$c = \begin{bmatrix} c_a \\ c_e \end{bmatrix}, \; r = \begin{bmatrix} r_a \\ r_e \end{bmatrix}, \; e = \begin{bmatrix} e_a \\ e_e \end{bmatrix},$$

$$y = \begin{bmatrix} y_a \\ y_e \end{bmatrix}, \; H = \mathrm{diag}(H_a, H_e) \tag{2}$$

It follows from Fig. 4 that

$$y = G_c c + G_r r \tag{3}$$

where $G_c = (I + GH)^{-1}GH$ and $G_r = (I + GH)^{-1}G$ are of dimension 2×2

$$G_c = \begin{bmatrix} G_{caa} & G_{cae} \\ G_{cea} & G_{cee} \end{bmatrix}, \; G_r = \begin{bmatrix} G_{raa} & G_{rae} \\ G_{rea} & G_{ree} \end{bmatrix} \tag{4}$$

The components of G have the following properties

$$| G_{aa} | \cong | G_{ee} | \cong 1 \text{ for } f < f_o \tag{5a}$$

$$| G_{aa} | \ll 1, \; | G_{ee} | \ll 1 \text{ for } f \gg f_o \tag{5b}$$

$$| G_{ae} | \ll 1, \; | G_{ea} | \ll 1 \text{ for all } f \tag{5c}$$

as illustrated in Fig. 5. The above properties yield the following monopulse loop properties

$$y \cong c \text{ for } \| GH \| \gg 1 \tag{6a}$$

$$y \cong r \text{ for } \| GH \| \ll 1 \tag{6b}$$

$$y \cong 0.5(c + r) \text{ for } GH \cong I \tag{6c}$$

In the first case of large open-loop gain, the closed-loop monopulse system follows the monopulse command. In the second case of small open-loop gain, the closed-loop system follows the encoder command. In the last case of unit monopulse gain, the system follows the average of the monopulse and the encoder commands. In order to prove this, note that, for $\| GH \| \gg 1$, one obtains $\| G_r \| \ll 1$ and $\| G_c \| \cong I$, hence $y \cong c$; for $\| GH \| \ll 1$, one obtains $\| G_c \| \ll 1$ and $G_r \cong G$, thus $y \cong r$; for $GH \cong I$, one obtains $G_r \cong 0.5G$, thus, from Eq. (3),

$y = 0.5GHc + 0.5Gr$; and, since $G \cong I$ (see Eq. (5)), one obtains $y = 0.5(c + r)$.

The transfer function H of the monopulse controller is determined as follows. The monopulse bandwidth f_m is smaller than the encoder bandwidth f_o; therefore, the monopulse tracker will compensate for a slowly varying error signal e. If the condition presented in Eq. (6a) is satisfied for $f < f_m$, the monopulse tracking system will follow the command c. And since $G \cong I$ for $f < f_m$, $\| H \| \gg 1$ is required to satisfy the condition presented in Eq. (6a). In addition, a rapid roll-off rate for $f > f_m$ would be an advantage. However, the roll-off rate is limited through the Bode conditions, as specified in [6, p. 25]. Namely, the roll-off rate in the region of the gain crossover frequency must not exceed 40 dB/decade, and for a reasonable stability margin it must actually be smaller than this. Due to this restriction, the following transfer function of the monopulse tracker is chosen

$$H = \frac{2\pi f_m}{s} I_2 \tag{7}$$

This transfer function satisfies Eq. (6a) for $f < f_m$ and has a roll-off rate of 20 dB/decade for $f > f_m$ (see Fig. 5). The parameter f_m of H is determined by analyzing the root locus of the monopulse closed-loop system with respect to f_m. The plot of real parts of closed-loop poles is shown in Fig. 6. It shows that for $f_m \geq 0.067$ Hz, the monopulse system is unstable. In order to maintain a reasonable stability margin, $f_m = 0.04$ Hz is chosen.

The plant transfer function G is obtained for the DSS-13 antenna with the encoder loop closed and the feedforward loop implemented. The magnitudes of the plant transfer function are shown in Fig. 7. From the figure, one can see that the condition of Eqs. (5a) and (5b) are satisfied, but the condition of Eq. (5c) is violated for some frequencies from the interval $f = [2, 10]$ Hz. This violation will cause some performance deterioration.

The azimuth and elevation components of the command signal r are shown in Fig. 3. The command c is slightly deviated from r by δ, i.e., $c = r + \delta$, where $\| \delta \| \ll \| r \|$. The plot of δ is shown in Fig. 8. Magnitudes of the transfer functions are shown in Fig. 9 from input r to output y, and in Fig. 10 from input c to output y. For the case of $\delta = 0$, the same input c and r are obtained and denoted u, i.e., $c = r = u$. In this case, one obtains from Eq. (3) $y = G_o u$, where $G_o = G_c + G_r$. The plots of the magnitudes of G_o are shown in Fig. 11. They indicate that the system follows low-frequency command c, high-frequency command r, and low- and high-frequency command u.

Implementation of the monopulse controller requires its discretization in time. The monopulse signal is supplied with the rate f_d Hz or with the sampling time $T = 1/f_d$ sec. In the case of the DSS-13 antenna, the sampling rate is 10 Hz. A block diagram of the discrete-time monopulse tracker is shown in Fig. 12. The main difference between the continuous-time and the discrete-time trackers lies in a delay of the tracking error.

The monopulse closed-loop systems with sampling rates of 10 and 50 Hz have been simulated. The 50-Hz sampled system has been simulated for evaluation of accuracy of the slower, sampled 10-Hz system. The simulations show similar results for 10- and 50-Hz sampling and are shown in Fig. 8 for the 10-Hz sampled system, where the solid line denotes the tracking error, e, and the dashed line the deviation, δ. The plots show that the pointing accuracy increased more than twofold in both cases. A sampling rate of 10 Hz is satisfactory to maintain the accuracy of the control system, and the 0.1-sec delay does not deteriorate the system performance.

IV. Improving Tracking Performance

As mentioned before, the implementation of the feedforward loop causes a significant excitation of flexible motion of the antenna, specifically in the elevation loop. The mode of deformation for the highest peak in the elevation-to-elevation transfer function is shown in Fig. 13. It is a bending mode of the antenna structure, strongly excited not only by the elevation command but also by the azimuth command. It impacts the stability and performance of an antenna. This mode is extremely difficult to control with elevation and/or azimuth torques, but any one of the following measures can be taken to reduce the impact of this mode on tracking performance: proper location of the elevation pinion, application of a notch filter, or adjustment of the amplifier gain in elevation drive. These measures are described below.

A. Choosing the Elevation Pinion Location

The antenna dynamics for the three positions of the elevation pinion, $\alpha = 0$, 60, and 90 deg, as shown in Fig. 14, have been simulated. The step responses are presented in Fig. 15, showing increased damping of transient motion for the higher location of the pinion. In consequence, the monopulse gains can be increased for the higher pinion, causing smaller tracking errors, as shown in Table 1. The decrease is almost proportional to $\cos\alpha$, which can be explained by the fact that the bending mode is excited mainly by the horizontal component F_h of the elevation pinion force F_t, proportional to the $\cos\alpha$, c.f., Fig. 14.

B. Implementing a Notch Filter

The critical elevation-to-elevation peak can be decreased by applying a notch filter. This narrowband filter removes a component of the specified frequency from the signal. The notch filter transfer function is as follows:

$$G_f(s) = \frac{s^2 + \omega_f^2}{s^2 + 2\delta\omega_f s + \omega_f^2} \tag{8}$$

where ω_f rad/sec is the filter frequency and δ is a damping coefficient which defines the bandwidth of the filter. The elevation input signal to the antenna excites the antenna vibration mode of frequency $f_f = 2.12$ Hz; thus, $\omega_f = 2\pi f_f = 13.32$ rad/sec and $\delta = 0.2$ have been chosen. In implementation, the matching of the filter frequency and the antenna resonance frequency is not a difficult task, since this particular resonance peak is strong and dominant, thus easily detected.

The notch filter is implemented as in Fig. 16. Let (A_f, B_f, C_f, D_f) and (A_r, B_r, C_r, D_r) be the filter and the antenna rate loop state-space representations, respectively. Their transfer functions are shown in Fig. 17, where the 2.12-Hz peak of the rate loop and 2.12-Hz dip of the notch are visible. The state-space representation of the connection is (A_o, B_o, C_o, D_o), where

$$A_o = \begin{bmatrix} A_f & 0 \\ B_r C_f & A_r \end{bmatrix}, \; B_o = \begin{bmatrix} B_f \\ B_r D_f \end{bmatrix},$$

$$C_o = [D_r C_f \; C_r], \; D_o = D_r D_f \tag{9}$$

and its transfer function is given by the dashed-line plot of Fig. 17. Note that the peak at 2.12 Hz has disappeared. The closed-loop properties of the antenna with and without a notch filter are compared in Fig. 18, where, again, the peak at 2.12 Hz has disappeared. The reduction of the peak allows one to increase significantly the gain of the monopulse loop without losing stability. Figure 19 shows the root locus of the antenna with and without the notch filter, showing that the antenna without the notch filter is stable for the gains $f_m < 0.06$, while the antenna with the notch filter is stable for $f_m < 0.5$. This gain yields about a tenfold reduction in the elevation pointing error, as follows from a comparison of Figs. 20 and 8(a).

Additional simulations have been performed to test the robustness of the system to filter frequency variations up to 10 percent of the nominal frequency, i.e., for filter frequency $f_f = f_{fn} \pm 0.1 f_{fn}$ (and f_{fn} is a nominal frequency). They show negligible deterioration of performance.

C. Adjusting Amplifier Gain

The impact of the critical elevation-to-elevation peak on monopulse controller stability and performance can be reduced by adjusting the open-loop gain. For example, the gain can be adjusted by varying the elevation drive amplifier gain. The elevation drive structure, presented in detail in [7], is shown in Fig. 21. It consists of amplifiers, motors, and gear boxes. The peak level of the elevation-to-elevation mode can be lowered by lowering the gain k_r of the elevation drive to $k_{ro} = \beta k_r$; $\beta = 0.33$ has been chosen through simulated tests. In doing so, the rate-loop transfer function has been lowered for higher frequencies, as in Fig. 22. The feedback gain, which contains an integrator, retains the tracking properties for low frequencies of the closed-loop transfer function, while the higher frequency part of the transfer function is not compensated, lowering the critical peak (Fig. 23). This simple approach allows one to increase gain of the monopulse loop, producing an improvement in tracking performance similar to that with the notch filter in Fig. 20. The explanation is as follows: Let G be the transfer function of the rate loop model from elevation to elevation, and K the transfer function of the PI controller. The closed-loop transfer function G_o is

$$G_o = \frac{KG}{1 + KG} \tag{10}$$

Since K consists of an integrator, therefore $|K| \to \infty$ for $\omega \to 0$, and $|K| \to 0$ for $\omega \to \infty$. It yields $|G_o| \to 1$ for $\omega \to 0$, and $|G_o| \to 0$ for $\omega \to \infty$; thus tracking for low frequencies is preserved, and the peaks in higher frequencies are suppressed. This can be seen in Fig. 23, where the low-frequency part of the closed-loop transfer function is the same, equal to 1 for $\beta = 1$ and $\beta = 0.33$, but for higher frequencies, the transfer function for $\beta = 0.33$ is lower than for $\beta = 1$.

V. Improving SNR

Typically, a monopulse signal is contaminated with measurement noise of significant intensity. Noise intensity is measured with the SNR

$$SNR = 10 \log_{10} \frac{P_s}{P_n} \text{ dB} \tag{11}$$

where P_s and P_n are signal and noise powers, respectively. The noise impacts the pointing accuracy of the control system. Here a simple method that improves SNR is discussed.

The monopulse signal $u(i\Delta t)$ (see Fig. 24) consists of a true measurement $u_o(i\Delta t)$ and a noise $n(i\Delta t)$

$$u(i\Delta t) = u_o(i\Delta t) + n(i\Delta t) \tag{12}$$

where $u_o(i\Delta t) = y_o(i\Delta t) - y(i\Delta t)$. It is assumed initially that the noise $n(i\Delta t)$ is a white noise with zero mean, $E(n(i\Delta t)) = 0$, where $E(.)$ is the expectation operator. The assumption is the worst-case scenario. White noise consists of components of all frequencies of equal intensity s_o, up to the Nyquist frequency f_c, as in Fig. 25(a). Typically, the measurement noise is rather a high-frequency noise; thus its impact on system performance is less severe than that of the white noise.

The monopulse signal $u(i\Delta t)$, shown in Fig. 26 for sampling time $\Delta t = 0.02$ sec, is transmitted to the antenna controller in clusters every N samples (typically $N = 5$); thus the new sampling period is

$$\Delta T = N\Delta t \tag{13}$$

and a cluster $U(i\Delta T) = \{u_1(i\Delta T), u_2(i\Delta T), \cdots, u_N(i\Delta T)\}$, consists of N measurements $u_k(i\Delta T)$

$$u_k(i\Delta T) = u(i\Delta T + k\Delta t),\ k = 1, \cdots, N \tag{14}$$

The mean value, $m_k = E(u_k(i\Delta t))$, and the variance, $\sigma_k = E(\Delta u_k(i\Delta T))^2$, of each component are the same in the cluster

$$m_k = m_N,\ \sigma_k^2 = \sigma_N^2,\ k = 1, \cdots, N \tag{15}$$

This assumption has the following meaning: the value of $u_o(i\Delta t)$ is considered constant within the period ΔT if the reaction of the antenna to $u_o(t_o + i\Delta t)$ is the same as to $u_o(t_o + N\Delta t)$ for $i = 1, \cdots, N$. This property has been confirmed by the earlier simulations reported in [1].

Although the monopulse signal is sent to the controller in clusters, only the last component, $u_N(i\Delta T) = u(i\Delta T + N\Delta t)$, is used to drive the monopulse controller. This excess information is used to reduce the signal-to-noise ratio by averaging the signal within a cluster. The average value, $u_{av}(i\Delta T)$, of the monopulse signal within the cluster of N samples is obtained

$$u_{av}(i\Delta T) = \frac{1}{N}\sum_{k=1}^{N} u_k(i\Delta T) \tag{16}$$

It is shown in the Appendix that in the case of white noise the mean value (m_{av}) of the averaged process, $u_{av}(i\Delta T)$, and the mean value (m_N) of the nonaveraged process, $u_N(i\Delta T)$, are the same, while the variance of the averaged process (σ_{av}^2) is smaller than the variance of the nonaveraged process (σ_N^2) by the factor N

$$m_{av} = m_N,\ \sigma_{av}^2 = \frac{\sigma_N^2}{N} \tag{17}$$

Define r_s, the ratio of variances of nonaveraged and averaged signals, as

$$r_s = \frac{\sigma_N^2}{\sigma_{av}^2} \tag{18a}$$

and its logarithmic counterpart, an SNR increase, ΔSNR, as

$$\Delta SNR = SNR_{av} - SNR = 10\log_{10} r_s \text{ dB} \tag{18b}$$

Then, for white noise, from the definition in Eq. (11), one obtains $r_s = N$ and $\Delta SNR = 10\log_{10} N$ dB.

Consider high-frequency noise with a constant spectrum within the interval $[f_o,\ f_c]$ (see Fig. 25(b)) such that $0 < f_o < f_c$, f_o is a cutoff frequency (the lowest frequency component of the noise), and f_c is the Nyquist frequency, $f_c = 0.5/\Delta t$. Results of noise reduction for the high-frequency noise, obtained through simulations, are shown in Fig. 27. From this plot of the ratio, r_s, versus cutoff frequency, f_o, it is evident that the high-frequency noise is more suppressed through averaging than is the white noise (r_s increases from 5 in the case of white noise to 50 in the case of high-frequency noise for cutoff frequencies of 8 Hz and higher, and SNR increase, ΔSNR, is from 7 to 17 dB, respectively). These results have also been confirmed by simulations of monopulse tracking with $SNR = 20$ dB, where the elevation pointing error for the

case of the nonaveraged signal is shown in Fig. 28(a), and the same error for the averaged signal is shown in Fig. 28(b), with the noise power ratio $r_s = 4.7$ (the SNR increase is $\Delta SNR = 6.7$ dB), which is close to the predicted $r_s = 5$ ($\Delta SNR = 7$ dB).

VI. Conclusions

It has been shown that the feedforward upgrade of the existing DSN antenna servos improves tracking at higher rates, and that monopulse tracking is an appropriate replacement of the conscan technique for the considered rates. A sampling rate of 10 Hz is satisfactory to maintain the accuracy of the monopulse control system, and the 0.1-sec delay does not deteriorate the system performance. Either repositioning of the elevation pinion, implementation of a notch filter, or adjustment of amplifier gain serves as a tool for improving tracking accuracy. The monopulse SNR is improved by averaging the high-frequency sampled signal.

Acknowledgment

The authors would like to thank Ben Parvin for his support and encouragement.

References

[1] W. Gawronski, "Feedforward Control Upgrade of the Deep Space Network Antennas," *The Telecommunications and Data Acquisition Progress Report 42-110, vol. April-June 1992*, Jet Propulsion Laboratory, Pasadena, California, pp. 253–262, August 15, 1992.

[2] D. K. Barton, *Modern Radar System Analysis*, Norwood, Massachusetts: Artech House, 1988.

[3] D. R. Rhodes, *Introduction to Monopulse*, New York: McGraw–Hill, 1959.

[4] J. C. Toomay, *Radar Principles for the Non-Specialist*, New York: Van Nostrand Reinhold, 1989.

[5] G. Biernson, *Optimal Radar Tracking Systems*, New York: Wiley–Interscience, 1990.

[6] J. M. Maciejowski, *Multivariable Feedback Design*, Wokingham, England: Addison–Wesley, 1989.

[7] W. Gawronski and J. A. Mellstrom, "Modeling and Simulations of the DSS 13 Antenna Control System," *The Telecommunications and Data Acquisition Progress Report 42-106, vol. April-June 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 205–248, August 15, 1991.

Table 1. Pointing errors.

Pinion position, deg	Elevation errors, mdeg	Cross-elevation errors, mdeg	Total errors, mdeg
0	1.43	0.14	1.44
60	0.76	0.08	0.77
90	0.35	0.07	0.36

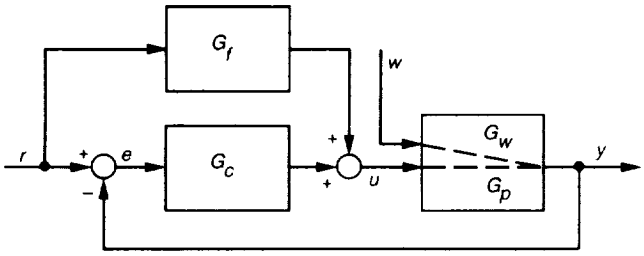


Fig. 1. The PI controller with the feed-forward loop.

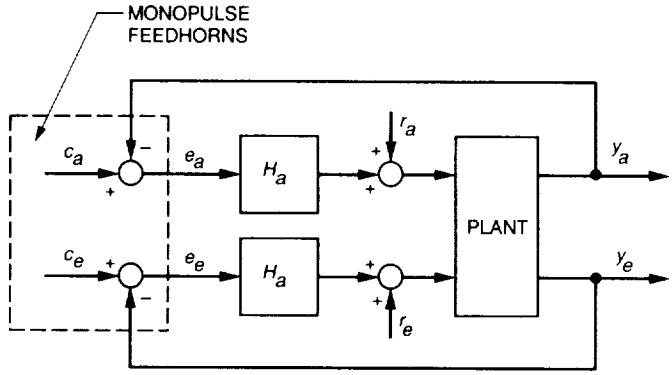


Fig. 4. The monopulse tracking system.

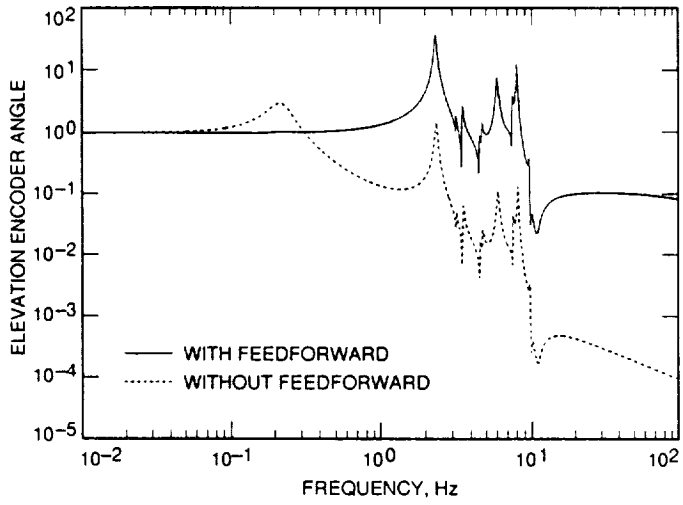


Fig. 2. Magnitudes of the closed-loop transfer function.

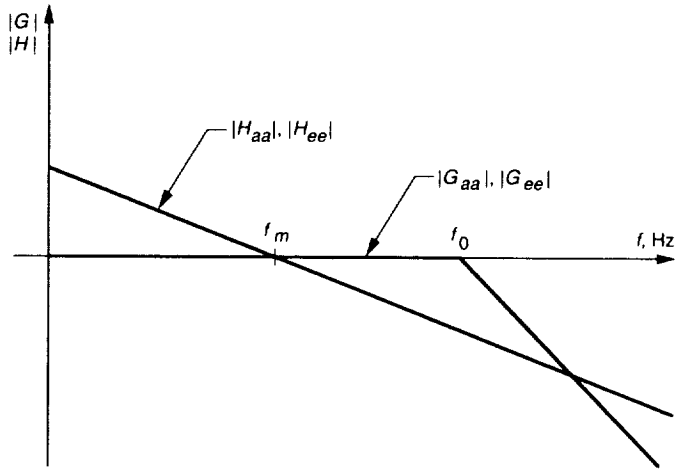


Fig. 5. Magnitudes of G and H.

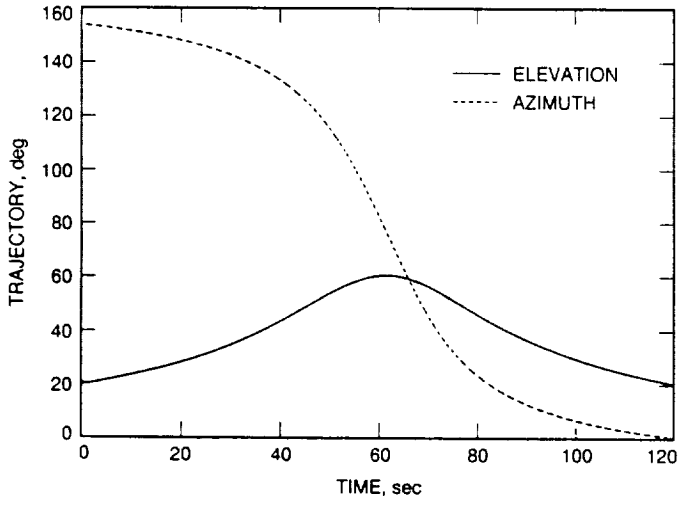


Fig. 3. Elevation and azimuth trajectories.

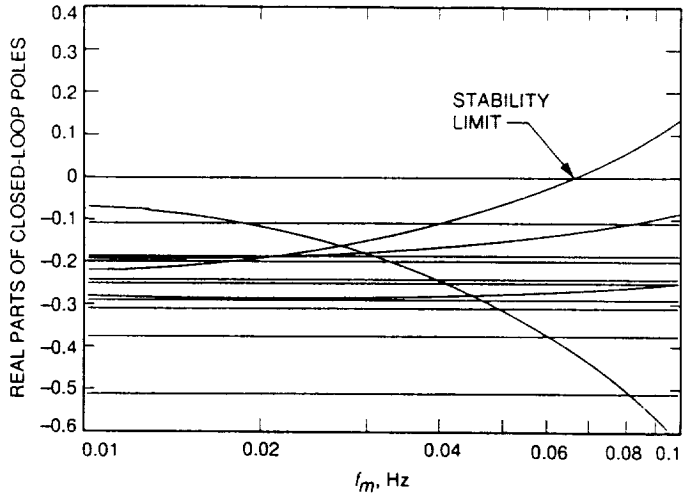


Fig. 6. Real parts of the closed-loop poles versus f_m.

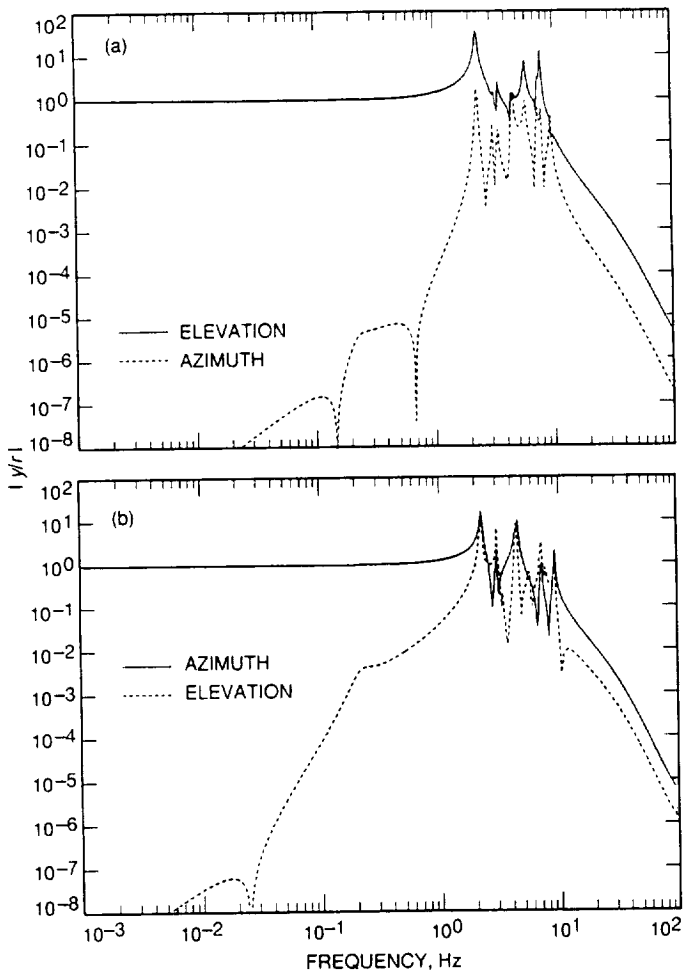


Fig. 7. Magnitudes of the plant transfer function from input *r* to output *y*: (a) elevation angle command and (b) azimuth angle command.

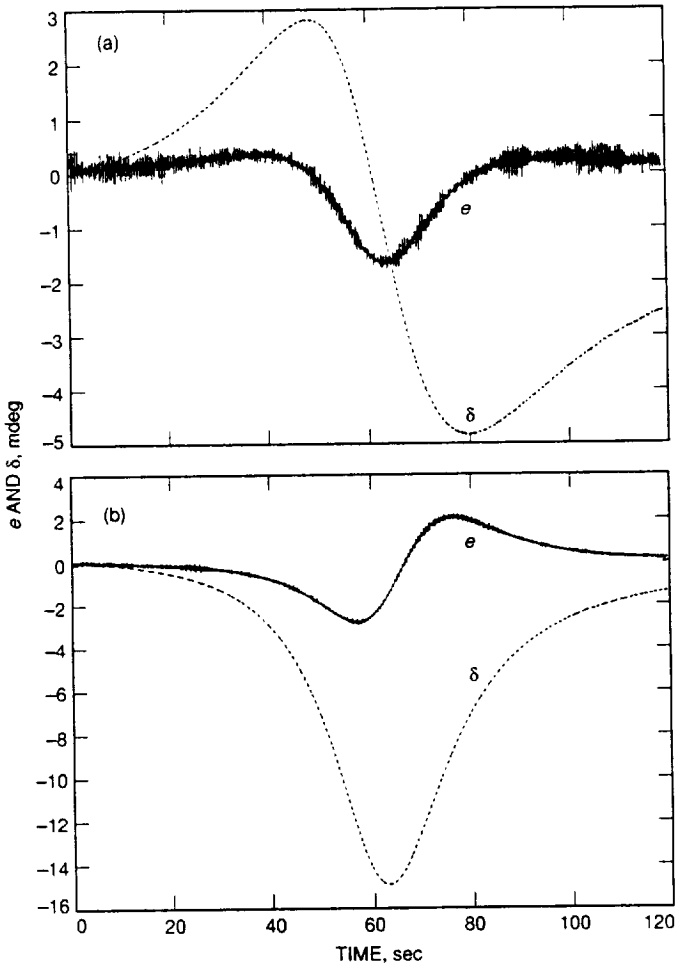


Fig. 8. Tracking error *e* and deviation δ in (a) elevation and (b) azimuth.

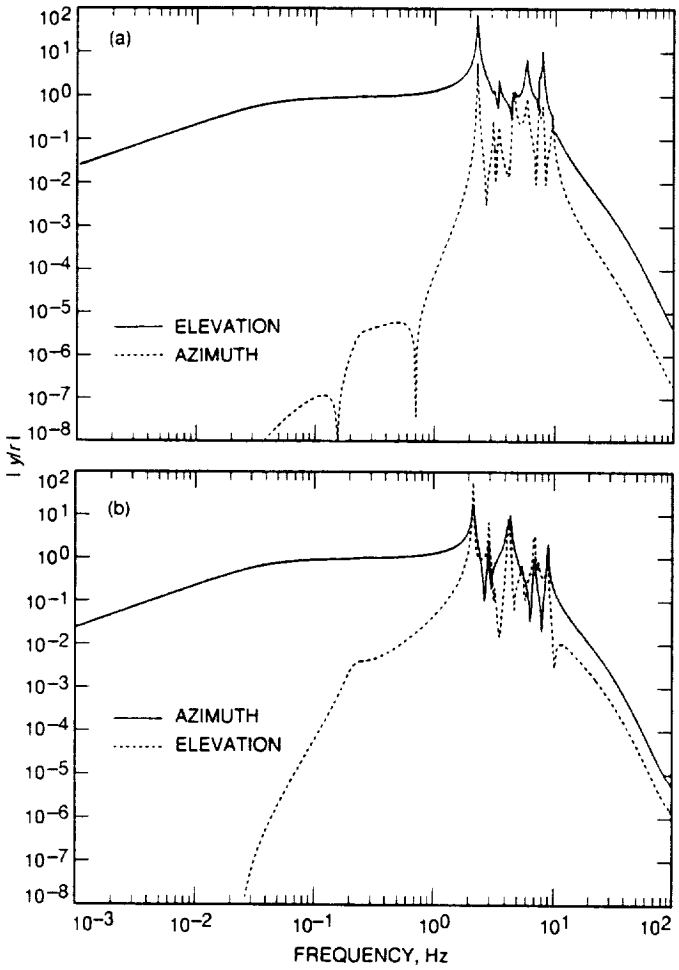


Fig. 9. Magnitudes of the closed-loop transfer function from input *r* to output *y*: a) elevation angle command and b) azimuth angle command.

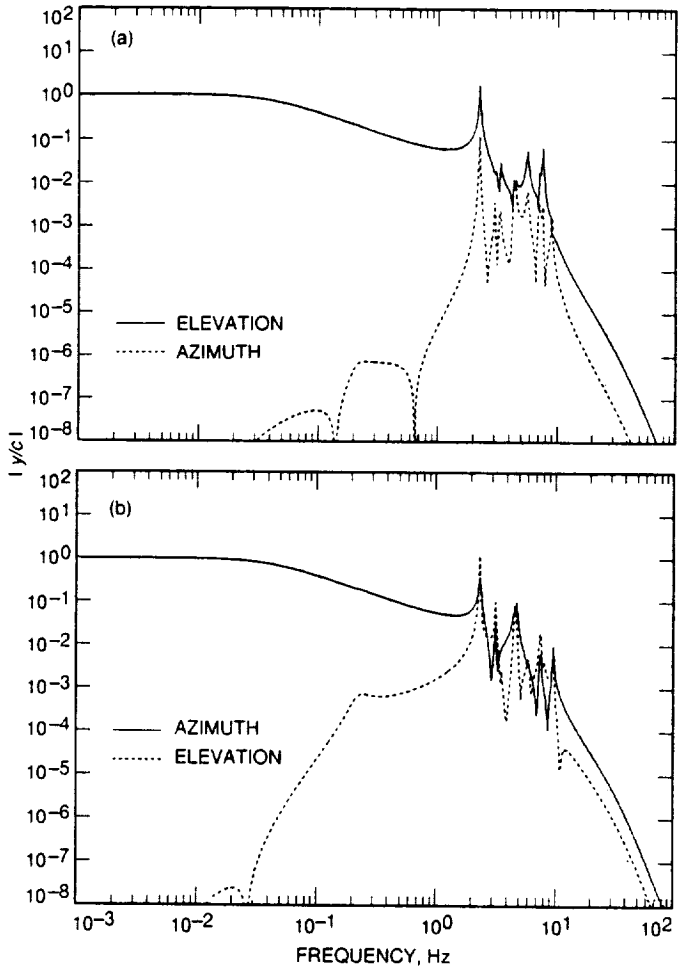


Fig. 10. Magnitudes of the closed-loop transfer function from input *c* to output *y*: (a) elevation angle command and (b) azimuth angle command.

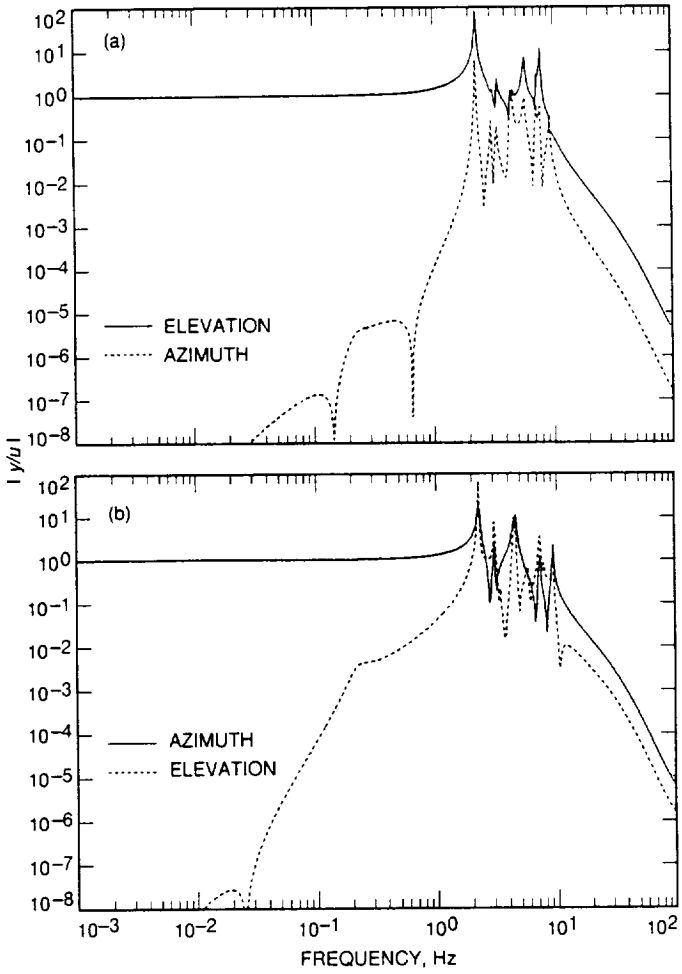


Fig. 11. Magnitudes of the closed-loop transfer function from input $u = c = r$ to output y: (a) elevation angle command and (b) azimuth angle command.

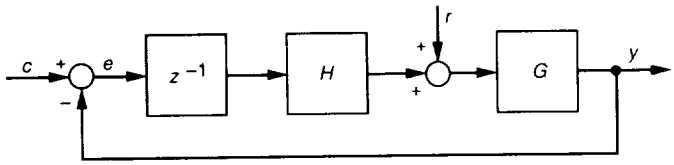


Fig. 12. The discrete-time monopulse tracking system.

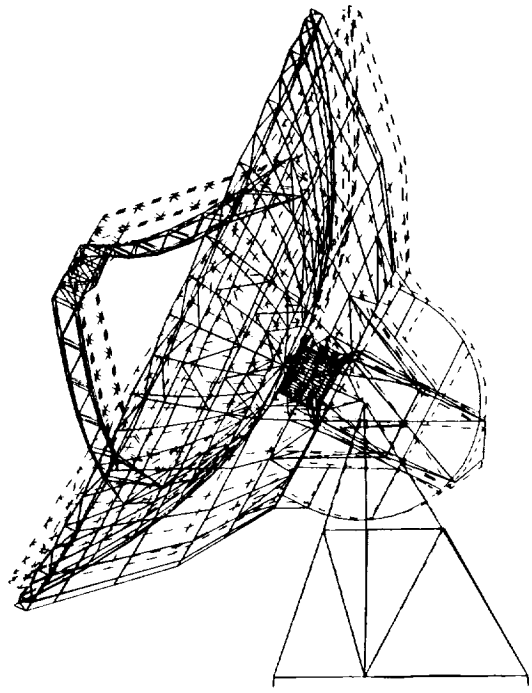


Fig. 13. The bending mode of the antenna.

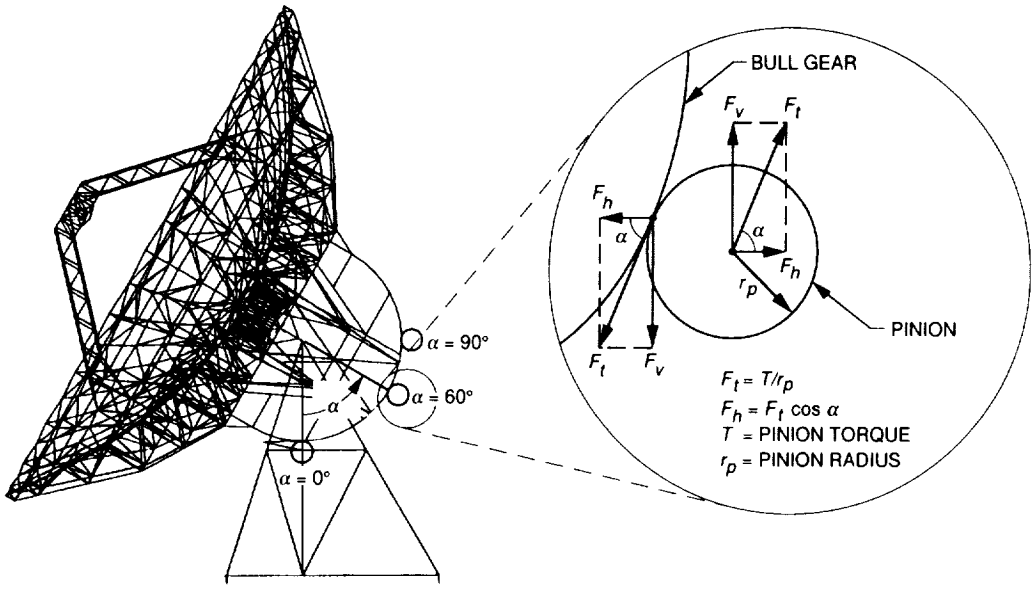


Fig. 14. Elevation pinion locations under investigation.

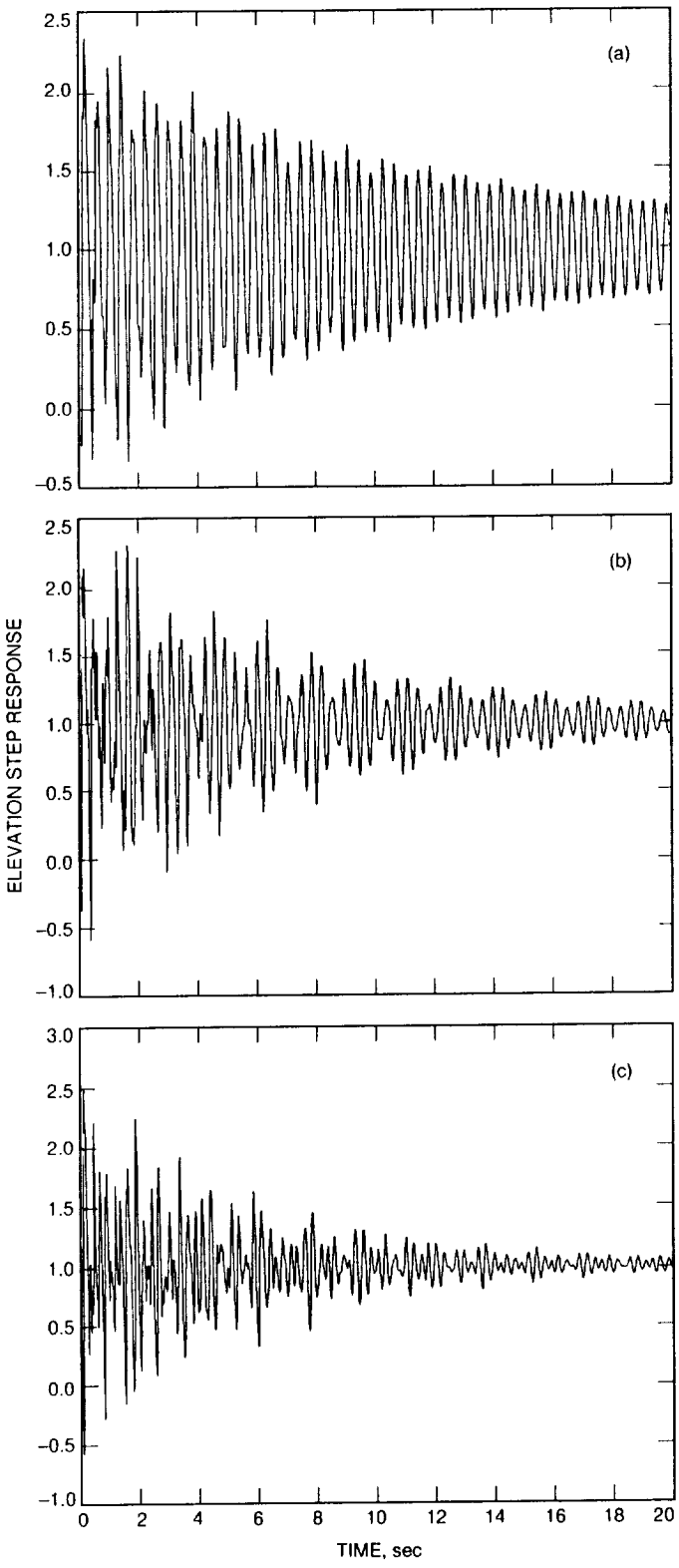


Fig. 15. Step responses (elevation command to elevation encoder) of the closed-loop system, for the elevation pinion at: (a) 0, (b) 60, and (c) 90 deg.

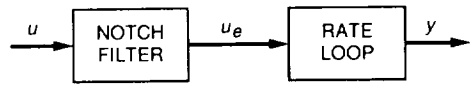


Fig. 16. The rate loop model with a notch filter.

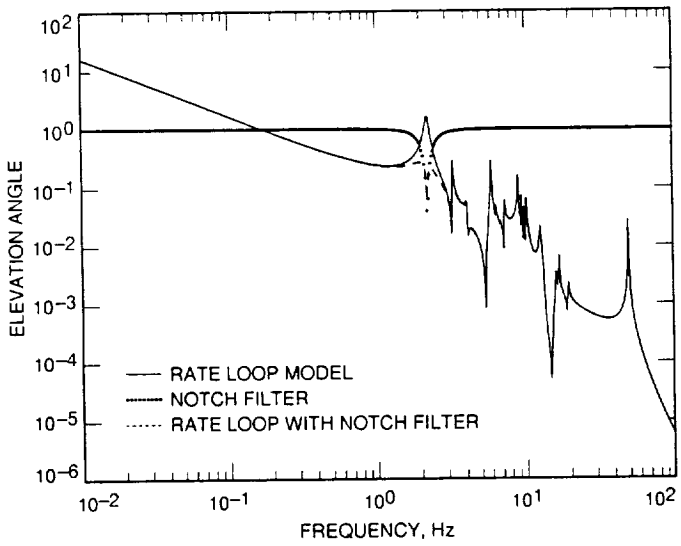


Fig. 17. Transfer functions.

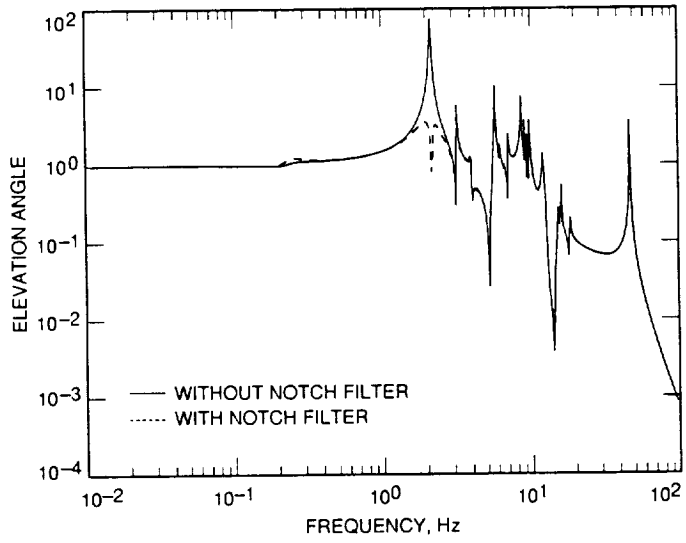


Fig. 18. Transfer functions (elevation command to elevation encoder) of the closed-loop antenna.

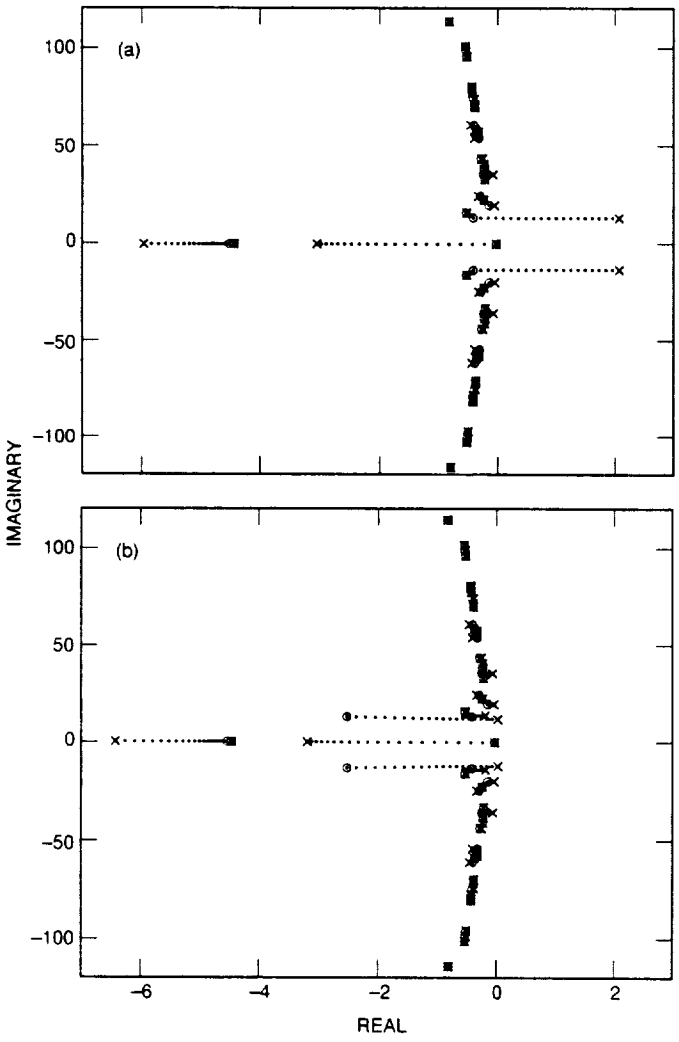


Fig. 19. Root locus of the antenna: (a) without the notch filter and (b) with the notch filter.

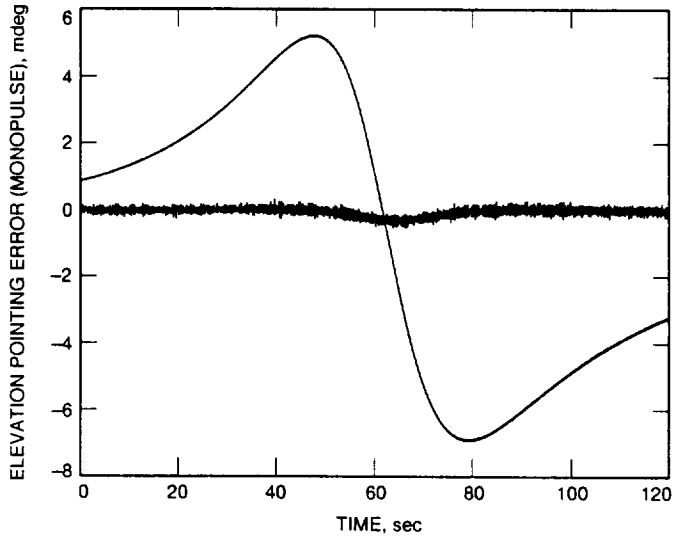


Fig. 20. Elevation pointing error for the antenna with the notch filter.

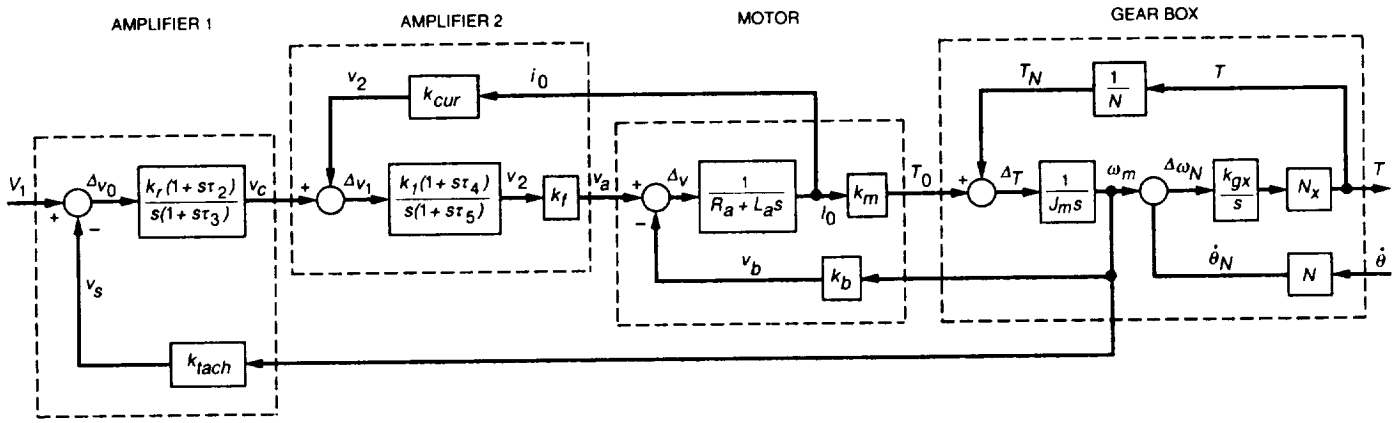


Fig. 21. The elevation drive model.

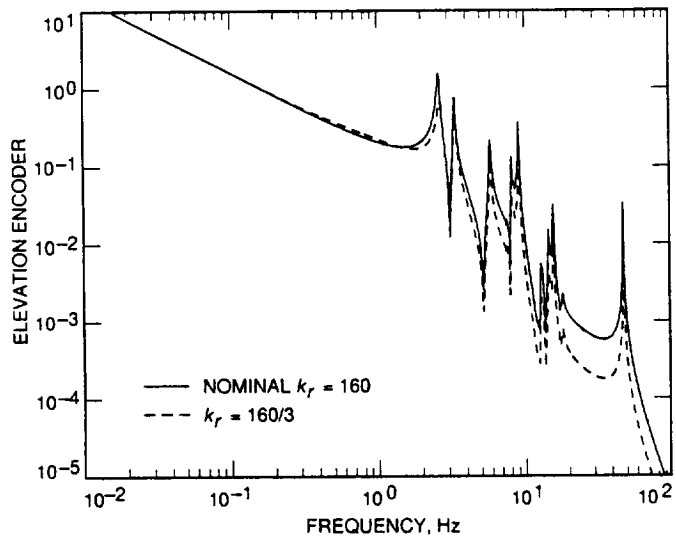


Fig. 22. Transfer functions (elevation rate command to elevation encoder) of the rate loop model.

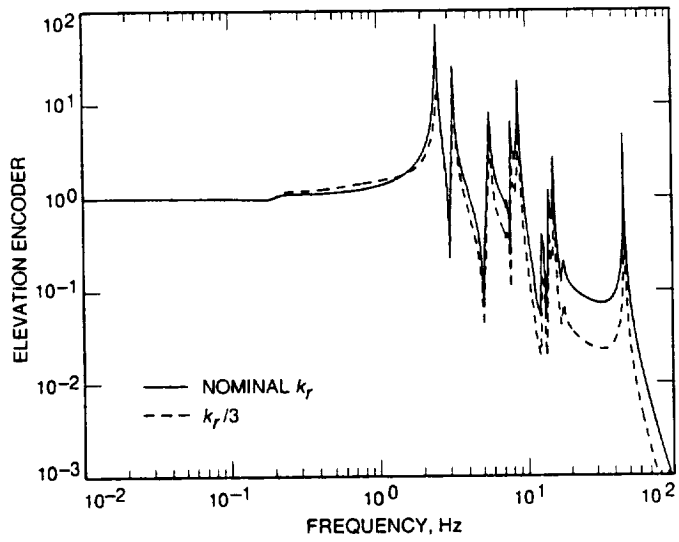


Fig. 23. Transfer functions (elevation command to elevation encoder) of the closed-loop antenna.

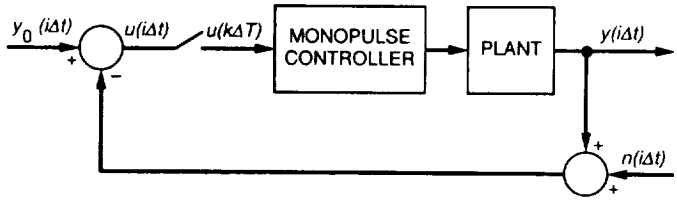


Fig. 24. The monopulse control system.

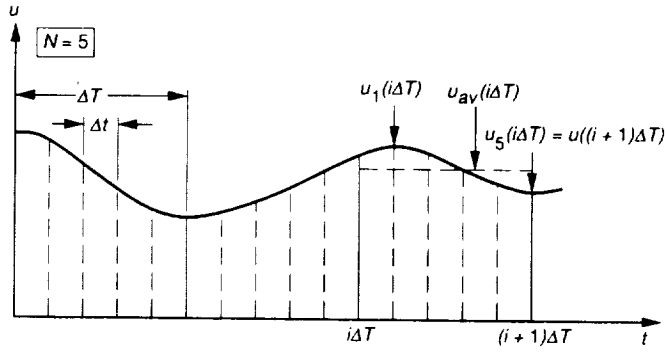


Fig. 26. The monopulse signal.

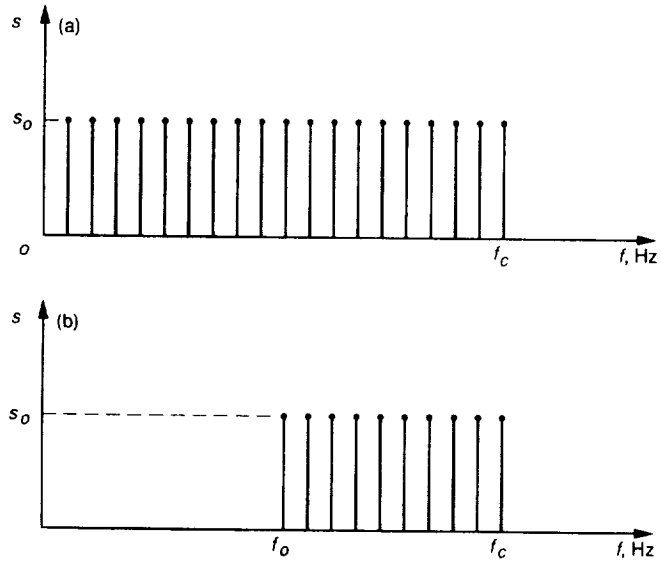


Fig. 25. Spectra of (a) white noise and (b) colored noise.

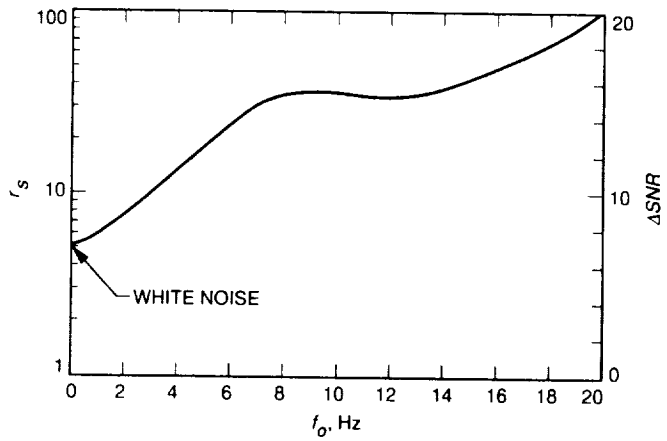


Fig. 27. Variances ratio and SNR increase versus cutoff frequency.

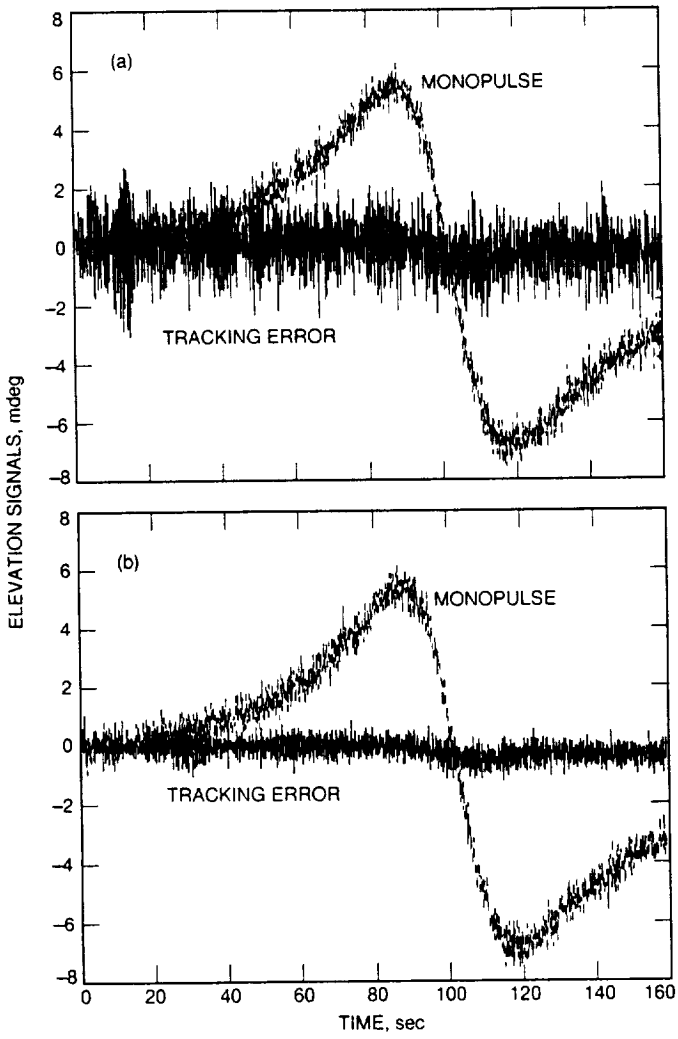


Fig. 28. Elevation tracking error and noisy monopulse signal (a) without averaging and (b) with averaging.

Appendix

Proof of Eq. (17)

The first part of Eq. (17) follows from the definition of the averaged process, Eq. (16), and the equality of mean values in the cluster, Eq. (15). Namely,

$$m_{av} = \frac{1}{N}\sum_{k=1}^{N} E(u_k(i\Delta T)) = \frac{1}{N}\sum_{k=1}^{N} m_k = m_N \tag{A-1}$$

In order to prove the second part of Eq. (17), denote

$$\Delta u_k(i\Delta T) = u_k(i\Delta T) - u_o(N\Delta t) = n_k(i\Delta t) \tag{A-2}$$

$$n_k(i\Delta T) = n(i\Delta T + k\Delta t),\ k = 1, \cdots, N \tag{A-3}$$

and

$$\Delta u_{av}(i\Delta T) = \frac{1}{N}\sum_{k=1}^{N} \Delta u_k(i\Delta T) \tag{A-4}$$

Thus, the variance of the averaged process is

$$\sigma_{av}^2 = E(\Delta u_{av}(i\Delta T))^2 = \frac{1}{N^2}\sum_{k,l=1}^{N} E(\Delta u_k(i\Delta T)\Delta u_l(i\Delta T)) \tag{A-5}$$

Since the white noise is not correlated, that is

$$E(n_k(i\Delta T)n_l(i\Delta T)) = 0 \quad \text{for } k \neq l \tag{A-6}$$

Therefore

$$E(\Delta u_k(i\Delta T)\Delta u_l(i\Delta T)) = 0 \quad \text{for } k \neq l \tag{A-7}$$

Introducing Eq. (A−7) and Eq. (15) to Eq. (A−5),

$$\sigma_{av}^2 = \frac{1}{N^2}\sum_{k=1}^{N} E\left(\Delta u_k(i\Delta T)\right)^2 = \frac{1}{N^2}\sum_{k=1}^{N}\sigma_k^2 = \frac{\sigma_N^2}{N} \tag{A-8}$$

proves the second part of Eq. (17).

Optical Subnet Concepts for the Deep Space Network

K. Shaik and D. Wonica
Communications Systems Research Section

M. Wilhelm
Telecommunications Systems Section

This article describes potential enhancements to the Deep Space Network, based on a subnet of receiving stations that will utilize optical communications technology in the post-2010 era. Two optical subnet concepts are presented that provide full line-of-sight coverage of the ecliptic, 24 hours a day, with high weather availability. The technical characteristics of the optical station and the user terminal are presented, as well as the effects of cloud cover, transmittance through the atmosphere, and background noise during daytime or nighttime operation on the communications link. In addition, this article identifies candidate geographic sites for the two network concepts and includes a link design for a hypothetical Pluto mission in 2015.

I. Introduction

Communications systems are inherently capable of operating at higher antenna gain and modulation bandwidth as carrier frequency increases. Optical frequencies (approximately 10^{14} Hz) are several orders of magnitude higher than the operating carrier frequencies of the conventional RF communication systems (approximately 10^{10} Hz) in use today.

The promise of the large antenna gain and modulation bandwidth that become available at optical frequencies is the basic reason for the interest in the development of optical communication systems.

Optical systems also promise smaller size and mass and lower power consumption as compared to RF systems with similar performance characteristics. For planetary space missions, the advantage of reduced size, mass, and power requirements will allow more room for science instrumentation aboard a spacecraft.

The optical subnet concepts for the DSN reported in this article were developed, and their telemetry performance was estimated, for the Ground Based Advanced Technology Study (GBATS). The GBATS work was performed in conjunction with Deep Space Relay Satellite System (DSRSS) study contracts,[1,2] and its purpose was to initiate exploration of Earth-based alternatives to the

[1] JPL Contract 958733 with TRW, Jet Propulsion Laboratory, Pasadena, California, March 28, 1990.

[2] JPL Contract 958734 with STEL, Jet Propulsion Laboratory, Pasadena, California, March 28, 1990.

DSRSS that would allow significantly higher telemetry rates for future NASA deep-space missions.

The GBATS study of optical subnets draws on previous design studies of the Deep Space Optical Reception Antenna (DSORA) [1] and on a weather model [2,3] of ground-based laser communications. The GBATS study also makes use of work accomplished by TRW, one of the contractors working on the DSRSS study, for the user-terminal design concept.[3] The emphasis of the work was on telemetry support. It is anticipated that future work on the optical subnets will include uplink command, navigation, and optical science.

This article describes initial concept designs for an optical subnet to augment the DSN. In Section II, a description of the ground optical terminal, which forms the basis of the optical subnet, and a description of the user-spacecraft terminal are provided. An overview of the optical subnet concepts is provided in Section III. The propagation and weather models are developed in Section IV to provide a basis for the calculation of network availability and coverage. No specific planetary missions are considered for the optical subnets discussed here, though a hypothetical Pluto mission in 2015 is used as an illustration (see Section IV). Accordingly, the future considerations of mission sets, operational issues, enhancement of a ground station's capabilities, etc., may profoundly affect the performance, configuration, and operation of an optical subnet.

II. Ground and Spacecraft Optical Terminals

A. Ground Optical Terminal

Each optical station operates in the direct detection mode at optical wavelengths between 500 and 2000 nm. All calculations in this study were made using 532 nm as the operating wavelength. The telescope consists of a 10-m, non-diffraction-limited, segmented primary mirror and a secondary mirror, in a Cassegrain configuration, as shown in Fig. 1. The telescope is mounted on azimuth–elevation gimbals and is housed in an environmental enclosure (dome). The receiver subsystem includes the beam-reducer optics, steering mirror, tracking detector, and the communications detector. Facilities for data processing, ground communications, logistics, and security, as well as office space and other uses, are identified in this section for completeness, but are not examined in detail in this article.

[3] TRW briefing, "Deep Space Relay Satellite System Study," Quarterly Progress Review, presented to JPL on February 25, 1993.

The optical terminal as described in this section provides the basic building block of the optical subnets. The performance of the optical subnets, calculated in the following sections of the article, was based on the capabilities of a single ground station. The following assumptions and guidelines were used to arrive at a definition of the ground optical terminal:

(1) A 10-m-diameter primary mirror.

(2) Telemetry reception under both daytime and nighttime conditions.

(3) Telemetry reception within 10 deg of the Sun.

(4) Operating wavelength of 532 nm.

(5) Tracking and slew rates compatible with deep-space probes.

(6) Acquisition of a user signal within 20 minutes at an elevation angle of about 15 deg under all operating conditions.

(7) A 2-mrad field of view (FOV) for the Cassegrain receiver telescope with a coarse pointing accuracy of 0.2 mrad.

(8) A 0.1-mrad FOV for the communications detector (this matches the blur diameter of the telescope).

(9) A fine pointing mechanism with an accuracy of 0.01 mrad.

(10) Station operation at high altitudes to reduce the impact of the atmosphere (up to 4.2 km).

(11) Uplink transmitter, command, emergency command, and navigation requirements were not considered at this time.

B. Block Diagram of the Optical Ground Station

Figure 2 describes the flow of information and control signals for the receive system of the optical station. The telescope with a 10-m fast primary collects optical energy and delivers it to the Cassegrain focus. The wide-FOV sensor provides calibration, removes systematic telescope-mount error, and helps in the acquisition of the user spacecraft within the telescope coarse FOV. From here the incoming beam is further reduced, is controlled, and is delivered to the communications detector. The communications detector demodulates the optical signal, and the resultant data stream is fed to the signal processor for bit/frame synchronization, decoding, error checking, etc. From the signal processor, the data are sent to the Ground Communications Facility (GCF) for transmission to the Network Operations Control Center (NOCC) in real time.

Raw or processed data are also stored in the archival subsystem for playback in case of GCF outage. The executive controller manages station activities automatically or manually through the command console, communicates with the outside world through the ground communications facility, and receives inputs from and sends commands to slave computers which include the pointing controller, the tracking controller, the figure controller, the signal processor, and the facility controller.

C. Ground Terminal Architecture

The system breakdown for the optical station is shown in Fig. 3. Note that subsystems other than the optical terminal are mentioned here for completeness and are not discussed any further. Additionally, the subsystems related to an optical uplink transmitter are not considered at this time.

1. Optical Terminal. An optical terminal consists of the following subsystems:

a. Telescope and Optics. The telescope subsystem provides an aperture to collect necessary photons for direct detection of incoming signals. The telescope employs a 10-m segmented primary mirror. There are 60 hexagonal segments, arranged in four rings, with each segment about 1.1 m in size (see Fig. 1). Other elements of the receiver telescope include a secondary-mirror assembly, a truss support structure, appropriate baffles to avoid the Sun, and other optics as needed. Each of the mirror assemblies includes mounts and the necessary actuators and baffles.

Table 1 provides a representative prescription for a Ritchey–Chretien Cassegrain telescope. The focal ratio for the 10-m segmented and hyperbolic primary is 0.5. The secondary mirror is 4.5 m from the primary mirror and is 1 m in size. The Cassegrain focus, where the optical communication instrument will be placed, is 3.25 m behind the primary. The image size at the Cassegrain focus for the usable diametric FOV (2 mrad) is about 16 cm.

b. Receiver Subsystem. The receiver subsystem consists of the optical communications instrument (OCI), which includes the receive beam-control optics (the beam-reducer optics, steering mirror, spectral filter, etc.), the tracking detector, and the communication detector. Fine pointing and tracking of the spacecraft are achieved by the OCI. Once coarse pointing is established by the acquisition, pointing, and tracking (APT) assembly, the OCI uses the communication signal as a beacon to aid in the fine acquisition, pointing, and tracking process. The communication detector begins telemetry reception and transfers it to the signal-processing subsystem once tracking has been established.

Figure 4 shows a conceptual drawing of the OCI with its optics, spatial and spectral filters, steering mirror, and detectors. The received beam at the Cassegrain focus is corrected by a field corrector, spatially filtered by the field lens, and reduced and collimated by the reducer optics. The beam is spectrally filtered and steered by a two-axis steering mirror for fine pointing. A tracking detector is used to acquire, track, and center the received beam on the communications detector. The diametric FOV of the communications detector is restricted to 0.1 mrad.

c. Acquisition, Pointing, and Tracking. The APT assembly uses computer controlled azimuth–elevation gimbals. The telescope is mounted on the gimbals, and this mounting provides coarse pointing to and tracking of the user spacecraft. Initial coarse pointing coordinates, which will be used to bring the spacecraft within the telescope FOV, will be provided by the DSN. The network configurations studied here allow roughly 20 minutes to acquire the spacecraft and establish tracking.

Table 2 provides estimates of the pointing and tracking requirements. The coarse pointing requirement (0.2 mrad) is chosen to be an order of magnitude less than the useful Cassegrain FOV. The fine pointing requirement (0.01 mrad) is an order of magnitude less than the communication detector's FOV. The tracking rate is consistent with sidereal tracking requirements for deep space spacecraft. If the ability to track highly elliptical orbits (HEO's) is considered necessary, the tracking and slew rates must be revised upward as needed.

d. Environmental Housing. The environmental housing will consist of a protective dome over the telescope structure. Figure 5(a) shows a conceptual diagram for the dome when the dome is closed. It is similar to the dome built for the Air Force Starfire Optical Range's 3.5-m facility in New Mexico. The dome protects the telescope from catastrophic failure due to severe weather and protects optical coatings on the primary and the secondary from premature degradation. Figure 5(b) shows the telescope fully exposed under normal operating conditions when the dome is folded down to the pier.

D. Configuration of the User-Spacecraft Terminal

The user spacecraft terminal configuration used in this article is based on a TRW concept for a future optical

terminal.[4] A block diagram for the user terminal is shown in Fig. 6. The telemetry data are encoded via a pulse position modulation (PPM) scheme, and provided to the modulator drive electronics. A modulated diode-pumped neodymium-doped: yttrium aluminum garnet (Nd:YAG) laser beam travels through a frequency doubler, a beam expander, a two-axis point-ahead mirror, a two-axis fine-steering mirror, and then exits from the telescope. The transmitter employs a 0.75-m telescope with no obscuration. The transmit wavelength is 532 nm. A small fraction of the transmit laser energy is directed toward acquisition and tracking electro-optics that use a beam splitter and a corner cube to determine the point ahead.

Table 3 shows a list of important transmitter parameters and the values used to estimate telemetry capability. See Appendix A for further details on communications link calculations.

III. Subnet Overview

A. Operations Concept

Like the current DSN, link geometry drives the major characteristics of the optical subnet. DSN user spacecraft with interplanetary trajectories will require multiple stations located about the equatorial region to provide continuous telemetry support to any point near the ecliptic plane. As the Earth rotates, continuous telemetry coverage is provided to any given user spacecraft via a hand-off strategy between the stations. As each station comes within the line of sight (LOS) of a user and good link geometry is established, telemetry reception begins. As the Earth continues to rotate and the user passes into the LOS of the next optical station, a hand-off occurs. Initial acquisition and tracking of a user spacecraft begin with the reception of the user ephemeris data provided by the NOCC. The user ephemeris provides coarse pointing information for acquiring the user transmit signal within the FOV of the telescope. Once coarse pointing is established by identifying the received beam on the acquisition and tracking detector, the receiver subsystem uses a fast steering mirror for fine pointing and centering of the signal beam on the communications detector. User spacecraft tracking is maintained throughout the pass by the combined action of the coarse pointing mechanism of the telescope and the fine steering mirror. The acquisition sequence followed by telemetry reception is repeated with down-line stations for the duration of the user spacecraft's need.

[4] Ibid.

User spacecraft pointing is established by detection of an uplink beacon, detection of the crescent Earth, or detection of the Sun with the point ahead off-set to the Earth (user spacecraft pointing is not part of GBATS). Coarse pointing is provided by the spacecraft attitude-control system from data provided by an onboard star tracker. Once the target (Earth) is acquired within the FOV of the user-spacecraft telescope, a fast steering mirror fine points and centers the target on a charge-coupled device (CCD) array. Data transmission begins once user pointing is established.

Based upon a 30-AU Pluto mission, and a 0.75-m user aperture, the footprint of the beam transmitted by the user terminal is smaller than the Earth's diameter; therefore, it is necessary to point the beam to the designated receiving station accurately. This can be accomplished without difficulty since the pointing bias and jitter errors, as shown in the earlier section on the user-terminal design, are much smaller than the signal beam diameter. A station is designated to receive telemetry when (1) it is within the LOS of the user terminal and (2) it has cloud-free weather. The need to predict weather availability for some subnet configurations is addressed in appropriate sections below. (Weather availability is a measure of station outage due to weather effects such as clouds, rain, and dense fog.)

The baseline for this study provides for one receive aperture per geographic location. This places some restrictions on simultaneous support of multiple missions. For example, user spacecraft with simultaneous-coverage requirements must be located nominally 180 deg apart.

B. Subnet Configurations

The presence of opaque clouds generally limits the availability of a single ground station for optical communications to less than 70 percent. This problem can be handled by employing spatial diversity.

There are two fundamentally different methods to provide the necessary spatial diversity to improve network weather availability for optical communications. The two concepts use different strategies in the location of optical stations to provide station diversity. These two approaches are referred to as the clustered optical subnet (COS) concept and the linearly dispersed optical subnet (LDOS) concept. In this article, two specific configurations, based on the COS and the LDOS concepts, were developed in detail. They were a COS network with nine stations and an LDOS network with six stations. Both configurations were developed based upon site-specific weather statistics, site surveys (accomplished by literature searches), coverage analysis, and projected telemetry performance. While

they use the same 10-m optical station and the same basic operations concept, each subnet offers unique advantages and disadvantages. Each subnet is designed to provide high weather availability. A detailed characterization of the two concepts and the reasons for selecting the number of stations in each case are provided in Section IV.

It is assumed that each station will require less than 20 minutes to acquire, track, and lock onto the incoming optical beam for both the LDOS and the COS concepts.

Figure 7(a) depicts network geometry for an LDOS showing three ground stations, and Fig. 7(b) depicts geometry for a COS network showing two of the clusters, each with three stations. Telemetry received by the available station for each subnet concept is demodulated and sent to the station data processing subsystem for one of three purposes: processing and formatting, storage in the archival subsystem, or for transmission in raw form to JPL's NOCC for distribution to end users. The stations are connected to the existing DSN infrastructure via the GCF.

IV. Performance Analysis

To develop optical network configurations that meet certain performance goals, several analyses were performed to identify a preferred approach. These efforts included the development of a propagation model, a weather model, an ideal-coverage model for the COS and the LDOS concepts, and availability assessments for various network configurations. For illustrative purposes, two network configurations, one from among the COS concepts and one from among the LDOS concepts, were selected for detailed study. For these two configurations, an LDOS with six stations and a COS with three clusters of three stations (COS 3 × 3), a coverage analysis was made for ideal conditions, as was a telemetry performance projection for a Pluto mission in the year 2015.

A. Propagation Model

Earth's atmosphere has a dominating impact on the propagation model for ground-based optical communications. Propagation loss and sky background radiance are two significant factors. Propagation loss, that is, loss due to transmission through the atmosphere, can be predicted using semiempirical models under various operating conditions. The problem of opaque cloud cover is studied in Section IV.B, where a weather model is produced.

The U.S. Standard Atmosphere 1976 model was used in this study to evaluate the effects of station altitude, meteorological range (i.e., visibility), and zenith angle. Section IV.A.1 shows that the impact of using atmospheric models other than the U.S. Standard Atmosphere 1976 is very small.

It is also important to study the impact of sky background noise on optical communications, especially the impact during daytime operations. This is addressed in Section IV.A.5. The results are used to develop average telemetry rates for daytime operations in Section IV.F.

1. Atmospheric Transmittance Model. LOWTRAN7, a transmittance model developed by the Air Force Geophysics Laboratory (AFGL) for visible and infrared wavelengths, was used to calculate propagation effects on wavelengths of interest, including 532 nm. The results of using the U.S. Standard Atmosphere 1976, mid-latitude winter, and mid-latitude summer atmospheric models on the transmittance, which was supplied by LOWTRAN7, are shown in Fig. 8(a). The curves shown for all the models assume the presence of high cirrus clouds, a 2.3-km altitude for the ground station, a 17-km meteorological range (visibility), and a zenith path through the atmosphere. Since the atmospheric transmittance models do not differ significantly from each other, the U.S. Standard Atmosphere 1976 model was used to calculate nominal spectral transmittance under all operating conditions.

2. Spectral Transmittance Versus Altitude. Figure 8(b) shows the transmittance for selected altitudes as predicted by LOWTRAN7. In the ideal-coverage model, the station altitude (2.3 km) of the Table Mountain Facility (TMF) was used as the baseline for the optical stations. Altitudes for the actual locations were used once specific LDOS and COS configurations were developed.

3. Spectral Transmittance Versus Meteorological Range. Varying meteorological range (visibility) will have an impact on the transmittance of the optical beam. Figure 8(c) shows the spectral transmittance for selected visibilities for wavelengths between 0.4 and 2.0 μm. A meteorological range of 17 km (defined as clear) was used as the basis for all calculations in this article.

4. Spectral Transmittance Versus Zenith Angle. The most dominant factor influencing the transmittance of the optical beam through the atmosphere is the operational zenith angle during telemetry reception. Figure 8(d) is a LOWTRAN7 plot of spectral transmittance for selected zenith angles for wavelengths between 0.4 and 2.0 μm. At a 70-deg zenith angle, the air mass through which the signal must propagate is about three times larger than the air mass at zenith. This is equivalent to about 17 dB of loss. In this article, the telemetry reception of the

optical station down to a zenith angle of 70 deg is included in the coverage analysis and link calculations.

5. Optical Background. Optical communications system performance in terms of data rate varies significantly between night and day. For a ground-based receiver, the sky radiance is a major source of optical noise, especially for daytime operation. This information was factored in when data volume over a 24-hour period was calculated for the GBATS work.

a. Nighttime. The sky brightness at night is about 50 nW/(m^2·nm·sr). This brightness is equivalent to a star of visual magnitude of 21.25 per square arcsec [4].

b. Daytime. Figure 9 shows sky radiance as a function of solar elongation. Sky radiance decreases by an order of magnitude for solar elongation (Sun–Earth–spacecraft angle) of 180 deg from a high of about 0.6 W/(m^2·nm·sr) when one is looking about 10 deg from the Sun. The graph is derived from LOWTRAN7 calculations for normal weather (17-km visibility) for a TMF-like receiver site. An average daytime data rate was calculated using six representative daytime sky radiances, specifically at 10-, 40-, 70-, 100-, 130-, and 160-deg solar elongation.

B. Weather Model and Availability Analysis

Besides geometry, the largest driver in terms of network performance is weather availability. With optical communications, the effects of weather on station availability are significantly more severe than they are at microwave frequencies. Unlike microwave frequencies, practically no communications can take place when the propagation path for an optical link is blocked by clouds. In this article, a weather model developed by Shaik [2] was used to model weather effects on link availability for optical stations in spatially independent weather cells. A total network availability of 90 percent was chosen as the performance goal.

1. Weather Model. For potential optical station sites, rough estimates of pertinent weather statistics can be obtained from existing sources, which include weather satellites. Figure 10 shows a contour diagram for the probability of a clear sky over the United States obtained from two years of GOES satellite data [5]. As can be seen, the probability of cloud-free skies over Southern California is about 66 percent. This means that 34 percent of the time, this area has partial to full cloud cover. To provide high weather availability (approximately 90 percent) requires that multiple stations be within the line of sight of the user spacecraft but located in different uncorrelated cells.

Based upon empirical information obtained from the AFGL, the cloud-system correlation coefficient between sites was expressed as[5]

$$\rho = \exp\left[-\Delta x^2/2\sigma^2\right] \quad (1)$$

where Δx is the distance between sites and $\sigma = 50$ km. This empirical result is then used to obtain the extent of cloud-system correlation for any two sites. An intersite distance of at least 3–4 σ, (or about 150–200 km) for $\rho \leq 0.01$ is found adequate to ensure spatially independent weather cells.

Given ground stations in spatially independent weather cells, a parametric weather model [2] can be used to compute link availability statistics. The model may be used to predict the joint probability (the percentage of time) that the extinction loss due to the atmosphere is less than some threshold for at least one of the ground stations. We define $\omega_n(L)$ as the cumulative distribution function (CDF)—that is, as the fraction of time when the propagation loss due to the atmosphere is less than or equal to L dB for at least one of the n sites with a line of sight to the user spacecraft. The weather availability can then be expressed as the CDF

$$\omega_n(L) = 1 - \{q\exp[-0.23b(L - L_0)]\}^n; \quad (L \geq L_0) \quad (2)$$

where L_0 is the acceptable loss through the atmosphere in dB and defines the operational telemetry line for the optical subnet. The minimum loss through the atmosphere is given by $\eta_a \sec(\zeta)$ in dB, where ζ is the zenith angle and η_a represents a site-altitude-dependent empirically derived propagation loss through the atmosphere under normal clear conditions. Since $\eta_a \sec(\zeta)$ estimates the minimum possible loss through the atmosphere, $L_0 \geq \eta_a \sec(\zeta)$. Parameter b is a site-dependent parameter and is derived empirically to model the CDF curve [2]. In this study, $b = 0.11$ and is derived from the assumption that $\omega_1(L = 30) = 0.8$ at zenith.[6] The equation assumes that the probability of cloudy skies, q, is the same for all sites, but it can easily be extended to site dependent q.

Equation (2) provides a simple model to compute the weather availability of an optical subnet. For example, under normal weather conditions for the Table Mountain

[5] K. Shaik, "Spatial Correlation of Cloud Systems," JPL Interoffice Memorandum 331-88.6-564 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 7, 1988.

[6] The probability of opaque clouds occurring in the Southwestern United States is less than 20 percent [2].

Facility, the minimum propagation loss at $\zeta = 60$ deg is -4.7 dB. Choosing this as the acceptable propagation loss, $L_0 = -4.7$ dB, and with $q = 0.34$ at TMF, the availability of a single site for $L = L_0$ is found to be $\omega_1(L_0) = 0.66$. If there are three such independent and identical sites in a subnet within the LOS of the user spacecraft, then from Eq. (2), the subnet availability is found to be $\omega_3(L_0) = 0.96$.

2. Weather Availability. As previously mentioned, weather availability is a measure of station outage due to weather effects such as clouds, rain, and dense fog. Individual sites for an optical subnet were chosen for their good cloud-free statistics, and are located far enough apart, as determined by Eq. (1), to ensure independent weather from station to station. The availability of a single station is expected to be at least 66 percent. The availability of a given network configuration is discussed in Section IV.D.

C. Coverage Analysis

LOS coverage (or, more simply, coverage) is defined as the percent of time during a 24-hour period when an unobstructed path, excluding weather conditions, exists between one or more stations on Earth and the user spacecraft. The performance goal for all networks is to provide 100 percent coverage.

A ground-based network consists of Earth stations strategically placed around the globe to provide full coverage, 24 hours a day. Ideally, only two stations located near the equator and placed exactly 180 deg apart would be required to provide full coverage. However, the number of stations quickly increases due to the constraint of the minimum operational elevation angle of 15 or 30 deg, the fact that the stations cannot always be placed at the equator, and the need to have more than one station in the spacecraft LOS to provide high weather availability. Specific network configurations and the coverage they provide are presented in the following paragraphs.

D. Network Analysis

The most promising network concepts which provide high weather availability and full coverage of the ecliptic were introduced in Section III.B earlier. In this section, subnet concepts are described in greater detail under idealized conditions to provide a rationale for the selection of promising configurations. The selected configurations, an LDOS with six stations and a COS configuration with nine stations, were then studied under realistic conditions with reference to a Pluto mission in 2015. The coverage curves and the telemetry rates are derived using actual site parameters, including longitude, latitude, altitude, and cloud-cover statistics, obtained from satellite data or in situ observations, and compared to the results obtained under ideal conditions.

1. LDOS Analysis. In this study, LDOS configurations were designed with six to eight ground stations spaced roughly equidistant from each other and placed around the globe near the equatorial region. An LDOS with five stations was not considered since the availability of this configuration is considerably below 90 percent (the percent required by the GBATS guidelines), and because the optical subnet would need to operate at very low elevation angles for a large fraction of the time.

Since the characteristic cloud systems calculated according to Eq. (1) are of the order of a few hundred kilometers in size, which is much smaller than the interstation distance, the adjacent stations will lie in different climatic regions and thus have uncorrelated cloud-cover statistics. Once specific sites were chosen, single as well as joint cloud-cover statistics for two or more consecutive sites were evaluated and used to predict link availability.

The probability of link outage for the LDOS configuration is low because (a) several stations are within the LOS of the user spacecraft, and (b) the stations lie in different climatic zones and hence their weather patterns are uncorrelated. Since the receiving sites are far apart, data with high spatial resolution on cloud-cover statistics are not needed. Existing data with a resolution of about 100 km are sufficient. However, further site surveys are needed to provide weather data with high temporal resolution. The weather data with high temporal resolution are needed to compute and predict short-term outage statistics accurately. Weather data with hourly or better temporal resolution will probably be needed to finalize site selection.

The distance between the receiving stations in the LDOS concept is very large; therefore, the full benefit of using optical wavelengths can be realized only when the user spacecraft points accurately to the designated receiving station in the subnet. Since the spacecraft can be 4–5 light hours from the Earth for some planetary missions, the weather availability of the subnet has to be predicted several hours in advance to designate the receiving station, and the location of the designated station must be uplinked to the user spacecraft terminal for pointing purposes.

a. LDOS With Six Ground Stations. The LDOS which consists of six optical stations located approximately 60 deg apart in longitude about the equatorial region is

shown in Fig. 11. Each optical station is located in a different climatic region (approximately 7000 km apart), and thus they have statistically uncorrelated cloud cover.

Figure 12 shows ideal coverage curves for six stations 60 deg apart in longitude. To calculate propagation loss, the model assumes that all station sites have normal visibility (17 km) and are as high as the Table Mountain Facility (2.3 km). It is also assumed that each site has cloud-free days at least 66 percent of the time (i.e., $q = 0.34$). For this configuration, only two stations will have LOS coverage of the spacecraft at all times when the telemetry line (acceptable zenith angle loss through the atmosphere) is consistent with a 60-deg zenith angle. The availability for this optical subnet is calculated to be $\omega_2(L_0) = 0.88$. The availability of the subnet can be increased to about 92 percent if a telemetry line consistent with a 75-deg zenith angle can be used.

Consider the situation when station 3 is receiving from a spacecraft on an equatorial path. The natural point to hand-off telemetry to station 4 is when zenith angle $\zeta_3 = \zeta_4 = 30$ deg (the subscript refers to the station number). Note that while ζ_3 is increasing, ζ_4 is decreasing. As calculated from the weather model described above, about 12 percent of the time, station 4 will be unavailable due to weather. In this case, station 3 continues to receive up to the point when $\zeta_3 = 60$ deg, at which point station 5 is activated at $\zeta_5 = 60$ deg. For this configuration, the telemetry line is placed at $\zeta = 60$ deg. The necessary trade-offs to optimize the position of the telemetry line have not been made. This leaves about one hour for acquisition and overlap between stations, as the stations are required to operate down to $\zeta = 75$ deg in zenith.

Table 4(a) provides a list of possible geographical sites for this LDOS configuration as an example. Appendix B describes the guidelines and the procedures used to select geographical sites in Table 4(a) and the following site tables. Weather statistics for all locations, except for the Hawaii and Chile sites, were obtained using satellite data [6] and are shown in Table 4(a). The data used for Hawaii and Cerro Pachan in Chile were based upon in situ observations [7]. Table B-1 in Appendix B lists possible additional sites.

Using specific sites given in Table 4(a) and assuming a hypothetical mission to Pluto in 2015 for illustrative purposes, a set of coverage curves was derived for a realistic LDOS with six stations. Figure 13 shows the coverage curves when data on actual geographical sites are used for the Pluto mission. The site-specific information used to obtain these curves includes altitude, longitude, and latitude, as well as Pluto's trajectory across the sky. Note that Pluto does not pass through the zenith for any of the sites. As can be seen in the figure, coverage will last from 2.5 to 4 hours, depending on the specific latitude of the optical station. For example, at the site in Siding Spring Mt., Australia, a telemetry pass will last approximately 4 hours.

A close examination of Fig. 13 shows that the telemetry line has been placed a little lower compared to Fig. 12. The acceptable atmospheric loss for the realistic Pluto mission is about −6.2 dB, instead of −4.7 dB for the ideal case, and corresponds to a 70-deg zenith angle rather than the 60-deg angle used in the ideal case. The introduction of actual geographical parameters has reduced the network availability for an LDOS with 6 stations from 88 percent for the ideal case to 81 percent. Also note that the acquisition time is about 20 minutes for the Pluto mission case instead of 1 hour for the ideal case.

b. LDOS With Seven Ground Stations. The interstation distance in this case will be roughly 51 deg in longitude (approximately 6000 km). Here, 35 percent of the time, three stations will be 30 or more deg above the horizon. The rest of the time, only two stations will be available. Availability for this configuration, when two or three stations are above 30 deg in elevation, is calculated to be $\omega_{2/3}(L_0) = 0.65\omega_2(L_0) + 0.35\omega_3(L_0) = 0.91$. The telemetry line for this configuration is at a 60-deg zenith angle. Table 4(b) provides a list of possible geographical sites. Table B-1 in Appendix B lists possible additional sites.

c. LDOS With Eight Ground Stations. The interstation distance for this configuration will be roughly 45 deg in longitude (approximately 5000 km). This configuration will ensure that three stations are 30 or more deg above the horizon about 66 percent of the time in a 24-hour period. An LDOS with 8 stations will provide 94 percent availability. The telemetry line is at a 60-deg zenith angle as before, providing considerably long overlap between stations. Table 4(c) shows a list of possible geographical sites for this configuration.

2. Analysis for the Clustered Optical Subnet. For geopolitical or operational reasons, the stations of an optical subnet may be required to be located within three or four locations around the globe that were chosen for their optimally cloud-free skies. In this concept, a cluster of three stations no more than a few hundred kilometers apart is envisioned for each of the selected regions. This distance is necessary to ensure that each station is located in a unique weather cell. For a major portion of the time, the spacecraft points to only one of these clusters; the

spacecraft hands over the signal beam to the next cluster as the spacecraft rises sufficiently above the horizon. Since the intracluster distance between stations is of the order of a few hundred kilometers, cloud-cover data with much finer spatial resolution (a few tens of kilometers) than for the LDOS configuration are required. In addition, the requirements for obtaining site-specific cloud-cover data with sufficient temporal resolution, which were discussed previously, apply here as well.

An advantage of the COS concept over the LDOS is that there is no need to predict weather availability several hours in advance. All stations within a cluster monitor the user-spacecraft's transmitted beam jointly with little pointing loss. Additionally, there is no need to designate a receiving station and, therefore, no need to uplink such information to the user spacecraft.

a. COS With 3 × 3 Stations. The clustered optical subnet to be discussed in detail consists of nine stations located in three clusters of three stations (COS 3 × 3); the clusters are approximately 120 deg apart in longitude (approximately 14,000 km). This configuration provides 96 percent weather availability since the stations are located within a cluster at distances no more than a few hundred kilometers apart.

Ideal coverage curves to model a COS 3 × 3, with the clusters located 120 deg apart in longitude, are seen as a subset of the curves for the LDOS configuration with six stations, which is shown in Fig. 12. (Consider curves 1(a), 3, 5, and 1(b) only.) The assumptions about the sites are the same as those described for the LDOS with six stations (see above); however, it is assumed that only one site in the cluster is receiving telemetry. The weather availability of this configuration is 96 percent, and the telemetry line is at ζ = 60-deg zenith angle, which is where the handing over to the following cluster takes place.

The geographical cluster locations chosen for the COS 3 × 3 are shown in Fig. 14. Table 5(a) provides a list of the specific geographical sites and their weather statistics. Like the sites chosen for the LDOS subnet, each COS 3 × 3 site has cloud-free days at least 66 percent of the time. In this configuration, each cluster is dedicated to a single user pass, resulting in a 96 percent probability that at least one optical station will have a clear LOS to the user.

Figure 15 shows the coverage curves for the COS 3 × 3 stations when data on one of the three actual geographical sites in each cluster are used for a Pluto mission in 2015. The actual sites used to obtain the coverage curves are TMF in California, Siding Spring Mt. in Australia, and Calar Alto in Spain. The site-specific information used to obtain these curves includes altitude, longitude, and latitude, as well as Pluto's trajectory across the sky. Note that Pluto does not pass through the zenith for any of the sites.

Like the LDOS configuration discussed above, the characteristic performance of the optical channel at approximately 70 deg off zenith (hand-over) is the determining factor for telemetry performance. The telemetry curve for the Pluto mission is placed at −6.2 dB, compared to −4.7 dB for the ideal case. However, even with this change, two gaps exist in the LOS coverage, totaling about 4 hours per day. The LOS coverage provided by the COS 3 × 3 for a Pluto mission in 2015 is about 79 percent. As is the case with the LDOS concept, each optical terminal has about 20 minutes to acquire, track, and lock onto the incoming optical beam. The total network availability has not changed, since each cluster contains three sites in independent weather cells.

Although this configuration provides the same telemetry rate as the LDOS network with six stations and better weather availability, the gaps in coverage and the significantly larger number of stations required for the clustered concept are distinct disadvantages.

b. COS With 3 × 4 Stations. A total of 12 optical stations will be necessary in this subnet configuration (COS 3 × 4). The distance between clusters will be roughly 90 deg in longitude (approximately 10,000 km).

Table 5(b) shows a list of probable geographical sites for COS 3 × 4. Each cluster (numbered 1 to 4) contains three optical station sites to satisfy the ground rules for the COS concept discussed above.

3. Network Availability. Weather-related availabilities for the idealized network configurations are shown in the second column of Table 6. The probabilities have been calculated using the model described above, with q = 0.34 for each individual site. Additionally, the acceptable zenith angle loss, or the telemetry line, used to calculate availabilities for the ideal LDOS networks is consistent with a 60-deg zenith angle, and the link calculations shown in Sections IV.E and IV.F below are based on this assumption. The telemetry line, however, can be made consistent with a 75-deg zenith angle to increase network availability to 92, 95, and 96 percent for an LDOS with 6, 7, or 8 stations, respectively. The trade-offs to identify optimum position for the telemetry line were not performed.

For an actual LDOS with six stations for the Pluto mission, a telemetry line at a 70-deg zenith angle was used

to calculate the network availability and the data rates shown in Section IV.F. The weather availability for the specific Pluto mission for an LDOS with six stations and for a COS with three clusters of three stations each is shown in the third column in Table 6.

4. Network Coverage. Table 7 shows that the LOS coverage for all idealized optical subnet configurations considered here is 100 percent. The coverage numbers for the actual geographical sites chosen for an LDOS with six stations and a COS 3 × 3 for a Pluto mission in 2015 are shown in the third column of the same table. Note that the coverage for the COS 3 × 3 for this specific case drops to 79 percent. The LOS coverage for COS 3 × 4 and LDOS with seven or eight stations with actual sites considered was not calculated but is expected to be 100 percent.

E. Link Calculations

Link analysis for a 30-AU Pluto mission at night was performed using OPTI 4.0, a software package developed in-house at JPL (see Appendix A).

1. Operational Considerations. The operational parameters used to estimate the telemetry capacity in this study are shown in Table 8. Details on other parameters used in the communication link budget are shown in Appendix A.

The modulation format used with the OPTI software was PPM. The alphabet size, as shown in Table 8, is 256.

A nominal raw link bit error rate (BER) of 0.013 was used. This was reduced to 10^{-5} by applying 7/8 Reed–Solomon coding. The 7/8 correction was applied to the data rate calculated by OPTI.

F. Telemetry

The telemetry return capability was used as the primary measure of the subnet performance. The benchmark established in the study for telemetry was 240 kb/sec for a future 70-m Ka-band (32 GHz) receiver, averaged over a 24-hour period. The user spacecraft antenna for this benchmark is 5 m in diameter.

The following assumptions and procedures were followed to calculate telemetry return capability for optical communications:

(1) The user spacecraft employs a transmitter proposed by TRW for its DSRSS study.[7] It is based on a 0.75-m telescope and a 7-W laser operating at 532-nm wavelength. See Appendix A for a list of transmitter parameters used.

(2) The optical terminal is based on a 10-m telescope. See Appendix A for a list of receiver parameters.

(3) Data rates for night and day were calculated separately. For the daytime calculation, an average data rate was computed over a number of daytime sky-radiance values.

(4) Data rates were computed for an ideal optical subnet and a realistic network for a 30-AU mission to Pluto in 2015.

(5) Data rates were computed for a conventional filter with a spectral bandwidth of 0.1 nm and for an atomic resonance filter with a spectral bandwidth of 0.001 nm.

(6) Daytime and nighttime data rates were averaged over a period of 24 hours for both optical filters, and telemetry improvement over the baseline was calculated.

1. Telemetry for the 30-AU Pluto Mission. Table 9 summarizes the data rates, which have been corrected for coding as discussed below, expected for an optical communications link between a 0.75-m user spacecraft transmitter at 30 AU and a 10-m ground station. Data rates using an atomic resonance filter (ARF) as well as a conventional filter were calculated for both an ideal configuration and a specific mission to Pluto in 2015. The acceptable zenith angle losses through the atmosphere for the ideal case and the actual Pluto mission were −4.7 dB (consistent with a 60-deg zenith angle) and −6.2 dB (consistent with a 70-deg zenith angle) respectively. The daytime data rate was obtained by averaging data rates calculated for six representative day-sky radiances between 10 and 180 deg solar elongation.

The data rates were first calculated using OPTI 4.0 for a 0.013 BER without coding. This raw data rate was then multiplied by 0.877 to obtain a 7/8 Reed–Solomon (R–S) coded data rate with a 10^{-5} BER for PPM modulation with an alphabet size of $M = 256$.[8] The dB gain, shown in parentheses with each data rate, was calculated over the agreed baseline telemetry rate of 240 kb/sec. Note that the data rates shown in Table 9 were not corrected for weather availability or the LOS coverage.

[7] TRW briefing, "Deep Space Relay Satellite System Study," Quarterly Progress Review, presented to JPL on February 25, 1993.

[8] W. Marshall, "Using the link analysis program with R-S encoded links," JPL Interoffice Memorandum 331-86.6-202 (internal document), Jet Propulsion Laboratory, Pasadena, California, August 1, 1986.

Table 9 shows that a ground-based optical subnet can provide very high data rates. For the Pluto mission at 30 AU, the telemetry rate can be as high as 1716 kb/sec, about 8.5 dB higher than the baseline rate of 240 kb/sec. Daytime data rates are lower, as expected, but still provide improvement over the baseline performance.

The telemetry rate can be further improved by employing 12- to 15-m receiver apertures. The technology for photon buckets up to 15 m in size is within reach with low technical risk. Use of a larger aperture, for a given data rate, is expected to have a favorable impact on the user-spacecraft design. It will usually mean a user-spacecraft optical terminal with smaller mass, size, and power consumption.

V. Conclusion

Several alternative optical subnet configurations were considered in this article. It is seen that an LDOS with six stations can provide nearly full LOS coverage of the ecliptic and 81 percent weather availability. If higher availabilities are needed, an LDOS with seven or eight stations can be used.

COS 3 × 3 under realistic conditions fails to provide full coverage (it provides approximately 79 percent). If the clustered concept for the optical subnet is desirable, a COS 3 × 4 with 12 ground stations will be required to provide full coverage, at least for the Pluto mission in 2015. The availability of both COS configurations is expected to be 96 percent. The COS configuration imposes an additional requirement over the LDOS configuration for locating appropriate specific sites. The clusters must be about 90 deg apart in longitude for COS 3 × 4, and intracluster station distances must be at least 150 km to ensure decorrelation of weather statistics. This may make it more difficult to find three specific sites within a given cluster when other requirements such as high altitude and reasonable accessibility are included.

A linearly dispersed optical subnet with six to eight stations is recommended, since it accomplishes the task with fewer ground stations than any other configuration considered here.

References

[1] K. Shaik, "Progress on ten-meter optical receiver telescope," *SPIE*, vol. 1635, edited by D. L. Begley and B. D. Seery, pp. 109–117, 1992.

[2] K. Shaik, "A Preliminary Weather Model for Optical Communications Through the Atmosphere," *The Telecommunications and Data Acquisition Progress Report 42-95, vol. July-September 1988*, pp. 212–218, November 15, 1988.

[3] K. Shaik and J. Churnside, "Laser Communication Through the Atmosphere," *Proceedings of the Twelfth NASA Propagation Experimenters Meeting (Napex XII)*, Syracuse, New York, pp. 126–131, June 9–10, 1988.

[4] J. Nelson, T. Mast, and S. Faber, "The Design of the Keck Observatory and Telescope," Keck Observatory Report No. 90, Keck Observatory, Mauna Kea, Hawaii, pp. 2–11, 1985.

[5] D. P. Wylie and W. P. Menzel, "Cloud cover statistics using VAS," *Nonlinear Optical Beam Manipulation, Beam Combining, and Atmospheric Propagation, Proceedings of the Meeting*, Los Angeles, California, January 11–14, 1988, pp. 253–259, 1988.

[6] *Greenhouse Effect Detection Experiment (GEDEX), 1992 Update*, CD-ROM, World Data Center for Rockets and Satellites, Code 902.2 DAAC, NASA Goddard Space Flight Center, Greenbelt, Maryland, 1992.

[7] The Association of Universities for Research in Astronomy (AURA), *The NOAO 8 m Telescopes* (now called Gemini 8-m telescopes), proposal to the National Science Foundation, AURA, Washington, DC, 1989.

Table 1. Telescope description.

Parameter	Value
System focal length, m	77.5
System focal ratio	7.75
System scale at Cassegrain focus, μrad/mm	12.9
System blur diameter, μrad	$\leq$100
System FOV at Cassegrain focus, mrad	2.0
Primary mirror size, m	10
Primary radius of curvature, m	−10.0
Primary focal ratio	0.5
Primary aspheric deformation	−1.0009
Primary to secondary distance, m	−4.5
Secondary mirror size, m	1.0
Secondary radius of curvature, m	−1.069
Secondary aspheric deformation	−1.3062
Secondary to Cassegrain focus distance, m	7.75

Table 2. Pointing, tracking, and slewing capability.

Parameter	Value
Coarse blind pointing, mrad	0.2
Fine pointing, mrad	0.01
Tracking rate, both axes, deg/sec	0.005
Slew rate for both axes, deg/sec	1.0
Acceleration/deceleration for both axes, deg/sec^2	3.0

Table 3. Transmitter parameters.

Transmitter parameter	Value
Average power, W	7
Wavelength, nm	532
Aperture size, m	0.75
Obscuration, m	0.0
Optics efficiency	0.8
Pointing bias error, μrad	0.1
RMS pointing jitter, μrad	0.1

Table 4(a). Linearly dispersed optical subnet with six ground optical stations.

Location	Altitude, km	Longitude, deg	Latitude, deg	Time zone	Cloud-free days/weather	Preexisting facilities and infrastructure
Southwest United States Table Mountain Facility, Calif.	2.3	118 W	34 N	−8	66%/arid[a]	Yes
Hawaii, United States Mauna Kea	4.2	155 W	20 N	−10	>69%/dry [7]	Yes
Australia Siding Spring Mountain	1.1	149 E	31 S	+10	67%/dry	Yes
Pakistan Ziarat	2.0	68 E	30 N	+5	69%/arid	Information NA
Spain/Northwest Africa Calar Alto, Spain	2.2	2 W	37 N	−1	67%/arid	Yes
South America Cerro Pachan, Chile	2.7	71 W	30 S	−4	77%/arid [7]	Yes

[a] ISCCP satellite data, obtained from [6].

Table 4(b). Linearly dispersed optical subnet with seven locations.

Location	Altitude, km	Longitude, deg	Latitude, deg	Time zone	Cloud-free days/weather	Preexisting facilities and infrastructure
Southwest United States Table Mountain Facility, Calif.	2.3	118 W	34 N	−8	66%/arid[a]	Yes
Hawaii, United States Mauna Kea	4.2	155 W	20 N	−10	>69%/dry [7]	Yes
Australia Siding Spring Mountain	1.1	149 E	31 S	+10	67%/dry	Yes
Nepal/South India	NA	NA	NA	+6	NA	Information NA
Saudi Arabia Jabal Ibrahim	2.6	41 E	21 N	+3	NA	Information NA
Spain/Northwest Africa Calar Alto, Spain	2.2	2 W	37 N	−1	67%/arid	Yes
South America Cerro Pachan, Chile	2.7	71 W	30 S	−4	77%/arid [7]	Yes

[a] ISCCP satellite data, obtained from [6].

Table 4(c). Linearly dispersed optical subnet with eight locations. Each of the eight listed locations will have a ground optical receiving station.

Location	Altitude, km	Longitude, deg	Latitude, deg	Time zone	Cloud-free days/weather	Preexisting facilities and infrastructure
Southwest United States						
Table Mountain Facility, Calif.	2.3	118 W	34 N	−8	66%/arid[a]	Yes
Hawaii, United States						
Mauna Kea	4.2	155 W	20 N	−10	>69%/dry [7]	Yes
Australia						
Siding Spring Mountain	1.1	149 E	31 S	+10	67%/dry	Yes
Australia						
Mt. Bruce	1.2	118 E	23 S	+8	NA/dry	Information NA
Pakistan						
Ziarat	2.0	68 E	30 N	+5	69%/arid	Information NA
Saudi Arabia						
Jabal Ibrahim	2.6	41 E	21 N	+3	NA	Information NA
Spain/Northwest Africa						
Calar Alto, Spain	2.2	2 W	37 N	−1	67%/arid	Yes
South America						
Cerro Pachan, Chile	2.7	71 W	30 S	−4	77%/arid [7]	Yes

[a] ISCCP satellite data, obtained from [6].

Table 5(a). Clustered optical subnet locations. The network consists of three ground optical receiving stations in each of the three locations.

Location	Altitude, km	Longitude, deg	Latitude, deg	Time zone	Cloud-free days/weather	Preexisting facilities and infrastructure
Southwest United States						
Table Mountain Facility, Calif.	2.3	118 W	34 N	−8	66%/dry[a]	Yes
Mt. Lemmon, Arizona	2.1	111 W	31 N	−7	>60%/dry [7]	Yes
Sacramento Peak, New Mexico	3.0	106 W	35 N	−7	>60%/dry [7]	Yes
Australia						
Mt. Bruce	1.2	118 E	23 S	+8	NA	Information NA
Mt. Round	1.6	153 E	30 S	+10	NA	Information NA
Siding Spring Mountain	1.1	149 E	31 S	+10	67%/dry	Yes
Spain/Northwest Africa						
Arin Ayachi, Morocco	3.7	5 W	33 N	0	NA	Information NA
Tahat, Algeria	2.9	5 W	22 N	−1	NA	Information NA
Calar Alto, Spain	2.2	2 W	37 N	−1	67%/dry[a]	Yes

[a] ISCCP satellite data, obtained from [6].

Table 5(b). Clustered optical subnet locations. The network consists of three ground optical receiving stations in each of the four locations.

Location	Altitude, km	Longitude, deg	Latitude, deg	Time zone	Cloud-free days/weather	Preexisting facilities and infrastructure
Southwest United States						
Table Mountain Facility, Calif.	2.3	118 W	34 N	−8	66%/dry[a]	Yes
Mt. Lemmon, Arizona	2.1	111 W	31 N	−7	>60%/dry [7]	Yes
Sacramento Peak, New Mexico	3.0	106 W	35 N	−7	>60%/dry [7]	Yes
Australia						
Mt. Bruce	1.2	118 E	23 S	+8	NA	Information NA
Mt. Round	1.6	153 E	30 S	+10	NA	Information NA
Siding Spring Mountain	1.1	149 E	31 S	+10	67%/dry[a]	Yes
Pakistan						
Ziarat	2.0	68 E	30 N	+5	69%/arid	Information NA
Site not determined	—	—	—	—	—	—
Site not determined	—	—	—	—	—	—
Spain/Northwest Africa						
Arin Ayachi, Morocco	3.7	5 W	33 N	0	NA	Information NA
Tahat, Algeria	2.9	5 W	22 N	−1	NA	Information NA
Calar Alto, Spain	2.2	2 W	37 N	−1	67%/dry[a]	Yes

[a] ISCCP satellite data, obtained from [6].

Table 6. Network availability.

Network	Availability with ideal sites, percent	Availability with actual sites, percent
COS 3x3	96	96
COS 3x4	96	96
LDOS: six stations	88	81
LDOS: seven stations	91	—
LDOS: eight stations	94	—

Table 7. Network coverage.

Network	Coverage with ideal sites, percent	Coverage with actual sites, percent
COS 3x3	100	79
COS 3x4	100	—
LDOS: six stations	100	95
LDOS: seven stations	100	—
LDOS: eight stations	100	—

Table 8. Operational parameters for link calculations.

Parameter	Value
PPM alphabet size	256
Link distance, AU	30
Raw bit-error rate	0.013
Slot width, nsec	10

Table 9. Nighttime, daytime average, and day/night average data rates (kb/sec) for a 10-m ground receiver, and average gain (dB) over baseline telemetry for the receiver with an atomic resonance filter (ARF) and with a conventional filter. The user spacecraft transmitter is at a distance of 30 AU and has a telescope 0.75 m in size. The data rates were not corrected for weather availability and LOS coverage.

Period	Ideal LDOS with six stations				Actual LDOS with six stations for a Pluto mission in 2015			
	ARF filter[a]		Conventional filter[b]		ARF filter[a]		Conventional filter[b]	
	kb/sec	dB gain[c]	kb/sec	dB gain[c]	kb/sec	dB gain[c]	kb/sec	dB gain[c]
Nighttime	1716	8.5	1716	8.5	1215	7.0	1215	7.0
Daytime average[d]	1056	6.4	377	2.0	774	5.1	298	0.94
Day/night average	1386	7.6	1047	6.4	994	6.2	757	5.0

[a] The ARF filter has a bandwidth of 0.001 nm.
[b] The conventional filter has a bandwidth of 0.1 nm.
[c] The dB gain is calculated over a baseline telemetry rate of 240 kb/sec.
[d] The daytime average is obtained by averaging data rates calculated for six day-sky radiances between 10- and 180-deg solar elongation.

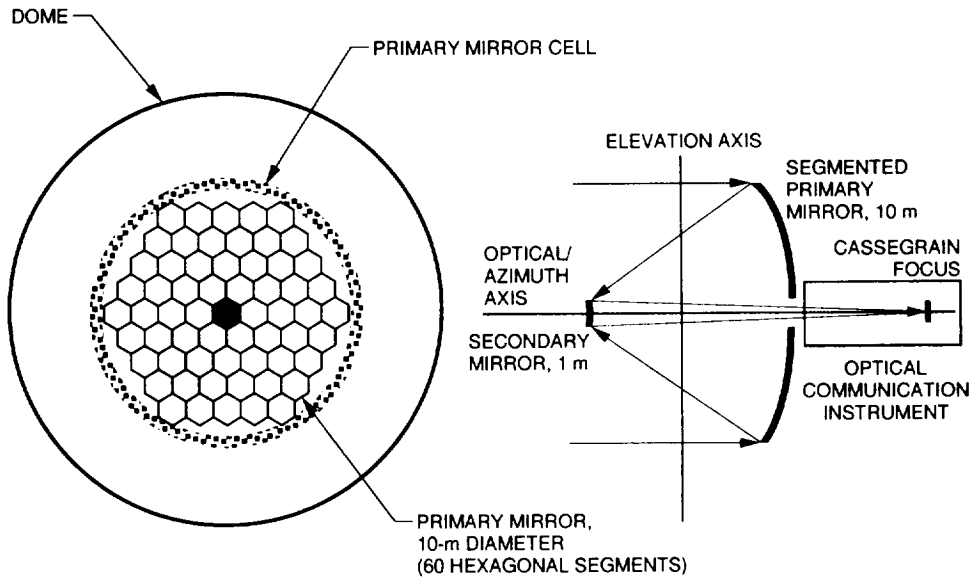


Fig. 1. Conceptual diagram of the 10-m telescope for the ground optical terminal (not drawn to scale). Primary diameter = 10 m; FOV at Cassegrain focus = 2.0 mrad; coarse pointing accuracy = 0.2 mrad; FOV at detector = 0.1 mrad; fine pointing accuracy = 0.01 mrad; and the blur diameter at focus $\leq$ 0.1 mrad.

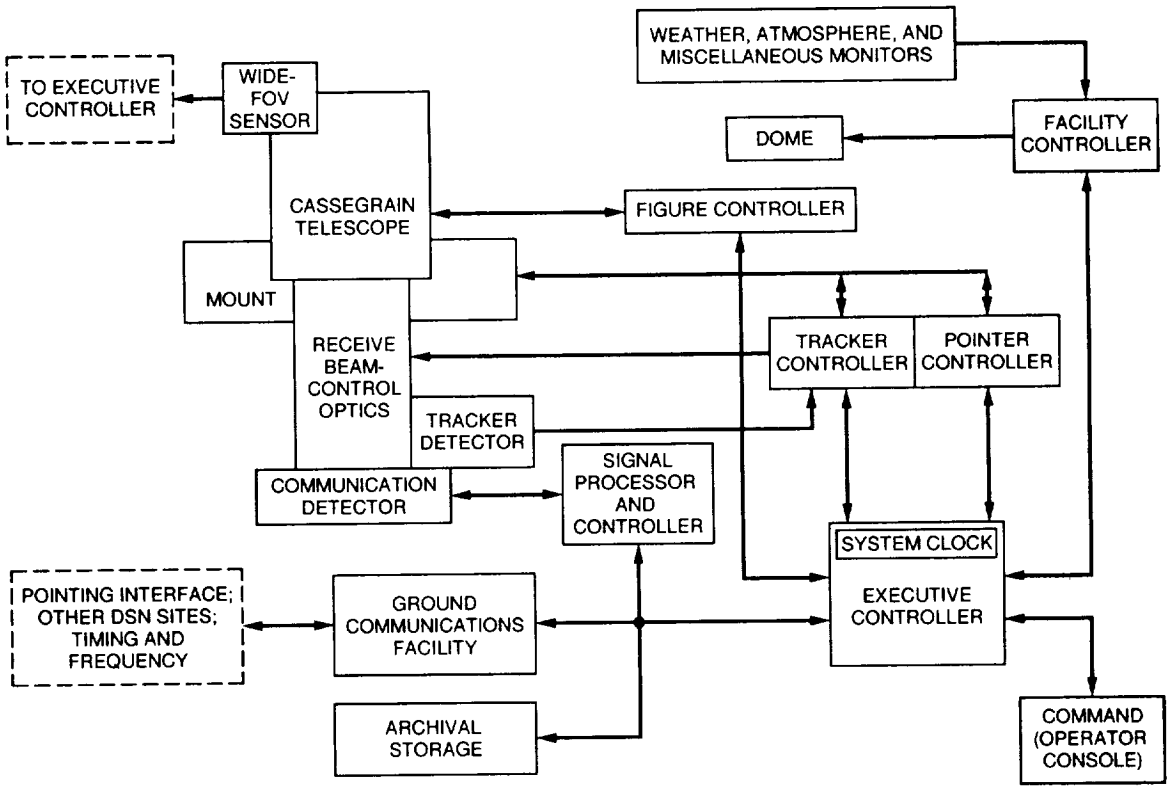


Fig. 2. Ground optical station block diagram.

I. OPTICAL TERMINAL:
   A. TELESCOPE AND OPTICS
   B. RECEIVER SUBSYSTEM
   C. ACQUISITION POINTING AND TRACKING
   D. ENVIRONMENTAL HOUSING

II. DATA PROCESSING:
   A. TIMING AND FREQUENCY
   B. SIGNAL PROCESSING
   C. CONTROL AND MONITOR
   D. GROUND COMMUNICATIONS
   E. ARCHIVAL STORAGE
   F. SIMULATION, TEST AND TRAINING

III. FACILITIES:
   A. BUILDINGS
   B. ROADS/WATER/SEWAGE
   C. POWER/GROUND SUBSYSTEM/UNINTERRUPTIBLE POWER SUPPLY
   D. HEAT, VENTILATION, AND AIR CONDITIONING/ FIRE SUPPRESSION/SECURITY
   E. WEATHER SUBSYSTEM
   F. LEASED TELECOMMUNICATIONS SERVICES

IV. LOGISTICS:
   A. ON-SITE EQUIPMENT SPARES
   B. DIAGNOSTICS, TOOLING, AND TEST EQUIPMENT
   C. TRANSPORTATION

V. OPERATIONS:
   A. WORKFORCE
   B. ADMINISTRATION/MATERIAL/TRAINING

Fig. 3. Optical station system breakdown.

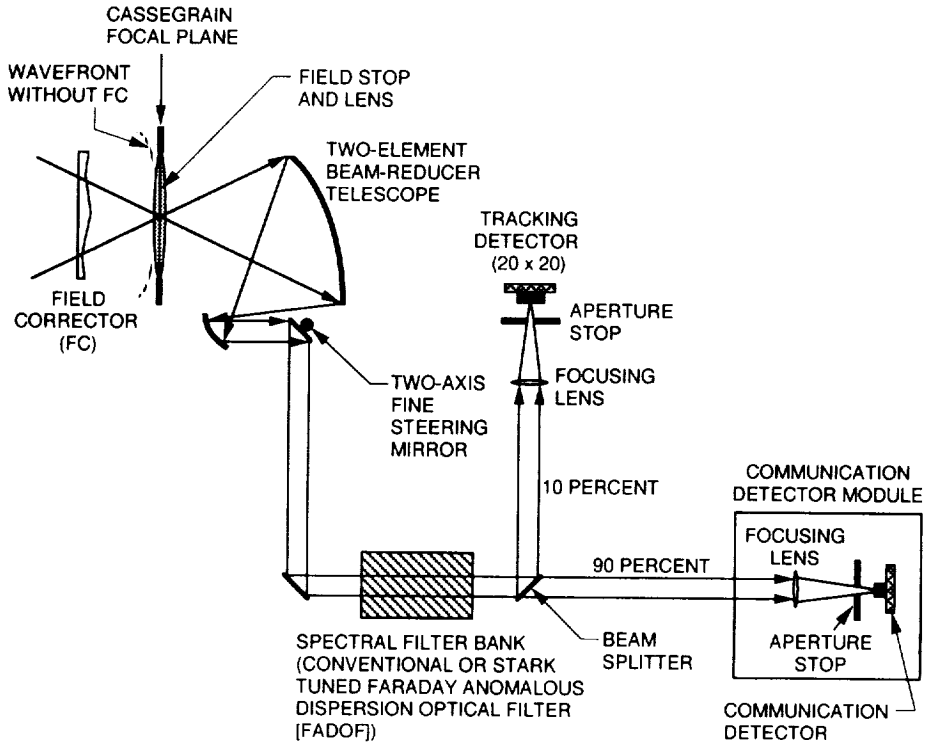


Fig. 4. Conceptual diagram for an optical communications instrument (not drawn to scale).

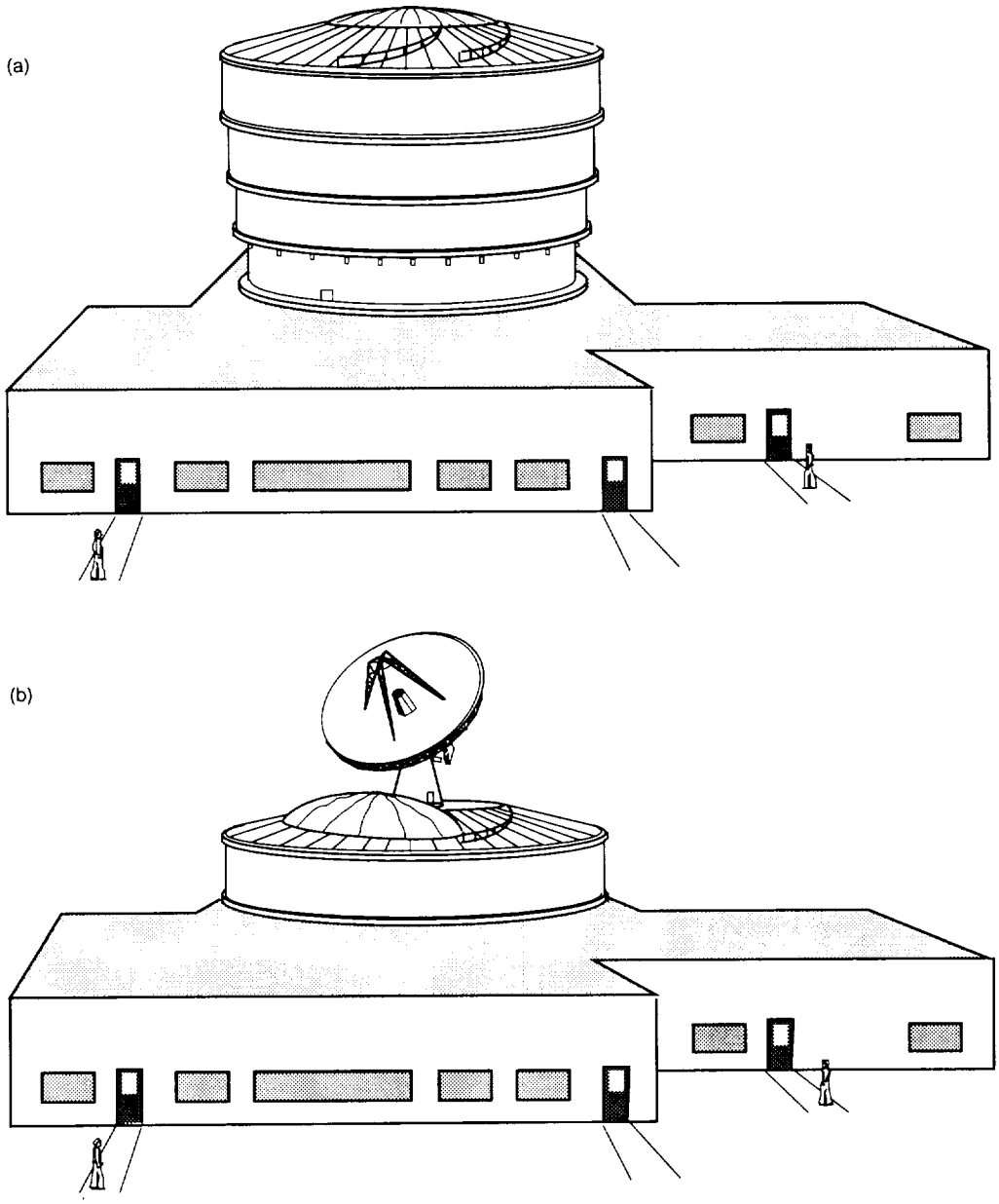


Fig. 5. Typical protective dome for the receiver telescope: (a) closed and (b) open (not drawn to scale).

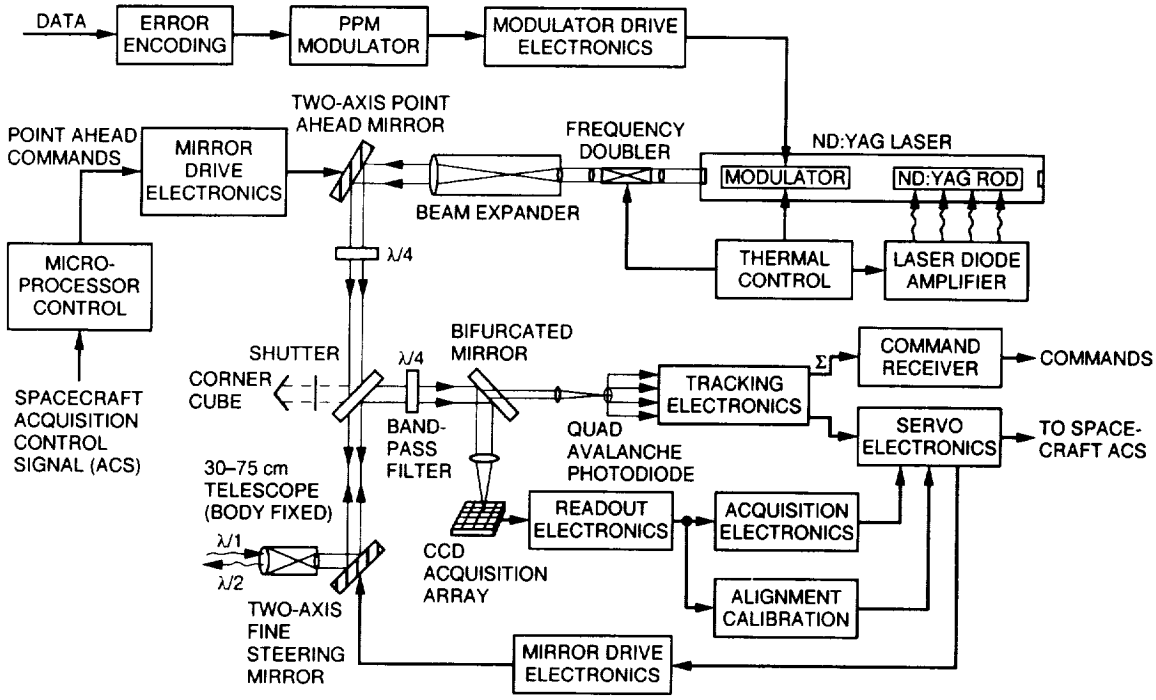


Fig. 6. Block diagram for the user terminal. (The terminal is based on TRW's conceptual design of a user terminal, done as part of a DSRSS study. The conceptual design was presented to JPL on February 25, 1993, as part of a quarterly progress review on the study.)

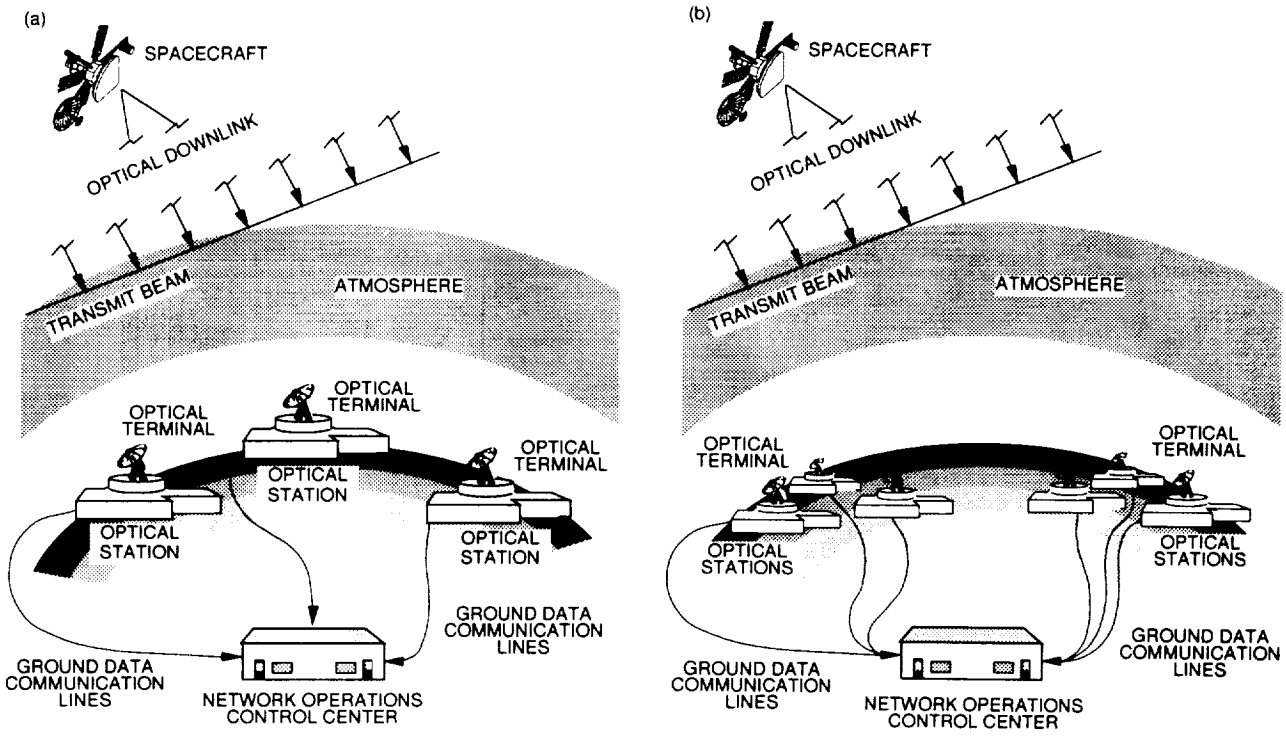


Fig. 7. Network geometry (not drawn to scale): (a) LDOS and (b) COS.

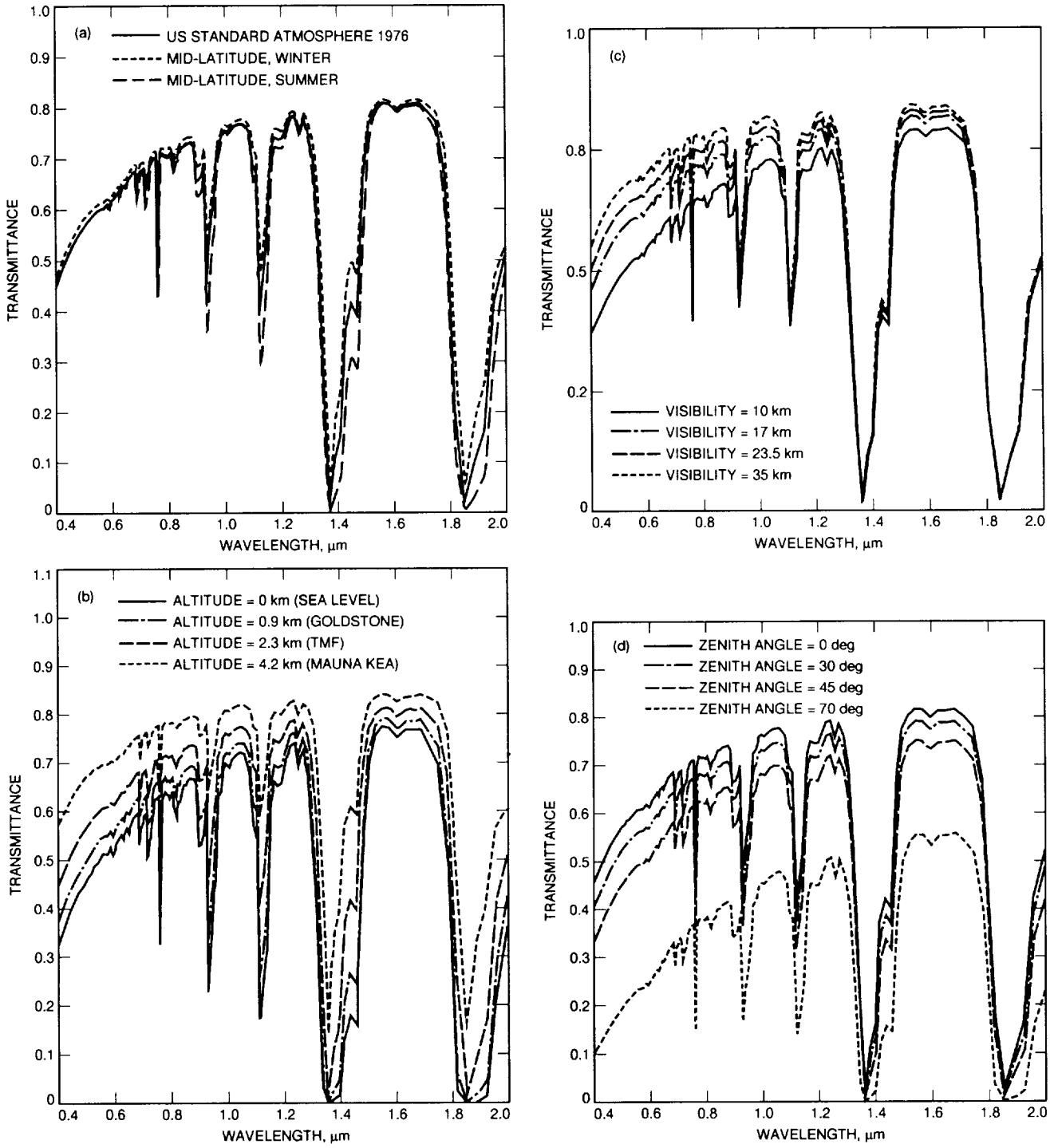


Fig. 8. Spectral transmittance data. All four diagrams assume high cirrus clouds. (a) Spectral transmittance over visible and near-infrared wavelengths for three LOWTRAN atmospheric models. (The diagram assumes a 2.3-km altitude, a 17-km meteorological range [clear], and a zenith path through the atmosphere). (b) Spectral transmittance for selected altitudes over visible and near-infrared wavelengths. (The diagram assumes the use of the U.S. Standard Atmosphere 1976 model, a 17-km meteorological range [clear], and a zenith path through the atmosphere). (c) Spectral transmittance for selected meteorological ranges (visibilities) over visible and near-infrared wavelengths. (The diagram assumes the U.S. Standard Atmosphere 1976 model, a 2.3-km altitude, and a zenith path through the atmosphere.) (d) Spectral transmittance for selected zenith angles over visible and near-infrared wavelengths. (The diagram assumes the U.S. Standard Atmosphere 1976 model, a 2.3-km altitude, a 17-km meteorological range [clear], and a slant path through the atmosphere).

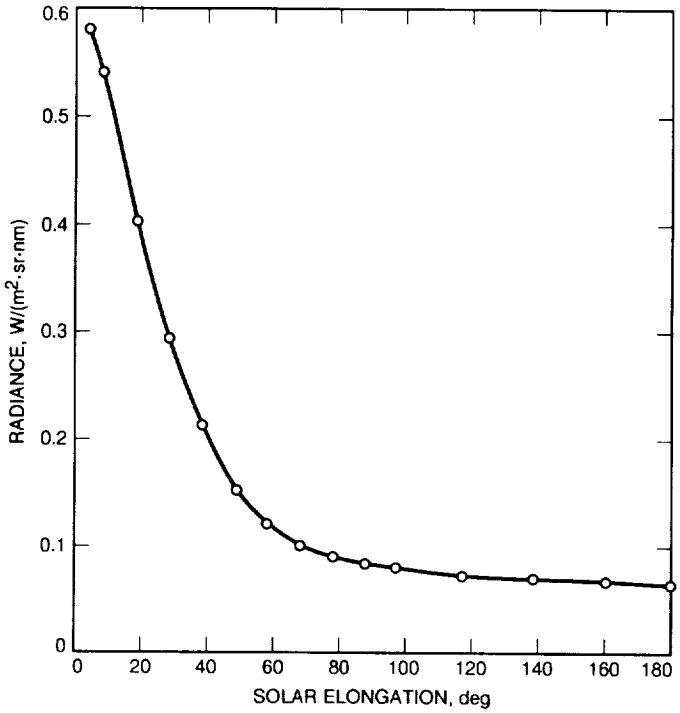


Fig. 9. Daytime sky radiance as a function of solar elongation (Sun–Earth–probe angle).

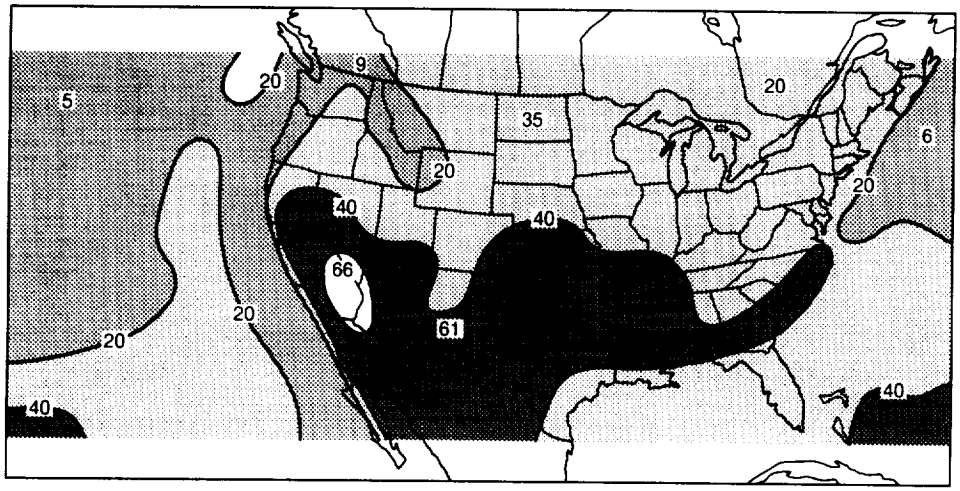


Fig. 10. Contour diagram obtained from two years of GOES satellite data; the diagram shows the probability of clear skies over the United States.

C-3

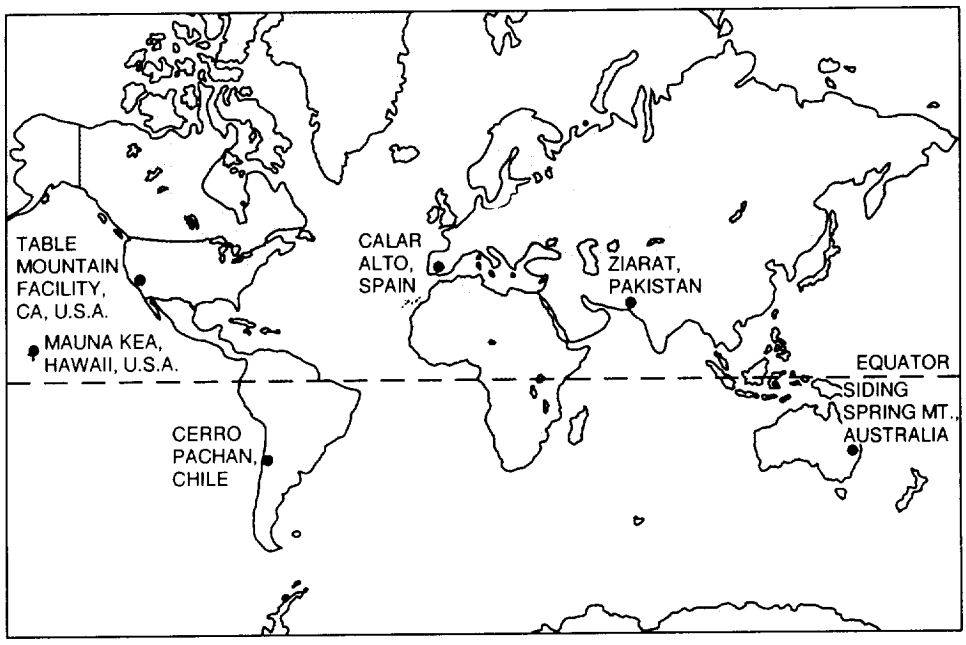


Fig. 11. Geographical sites for an LDOS with six stations.

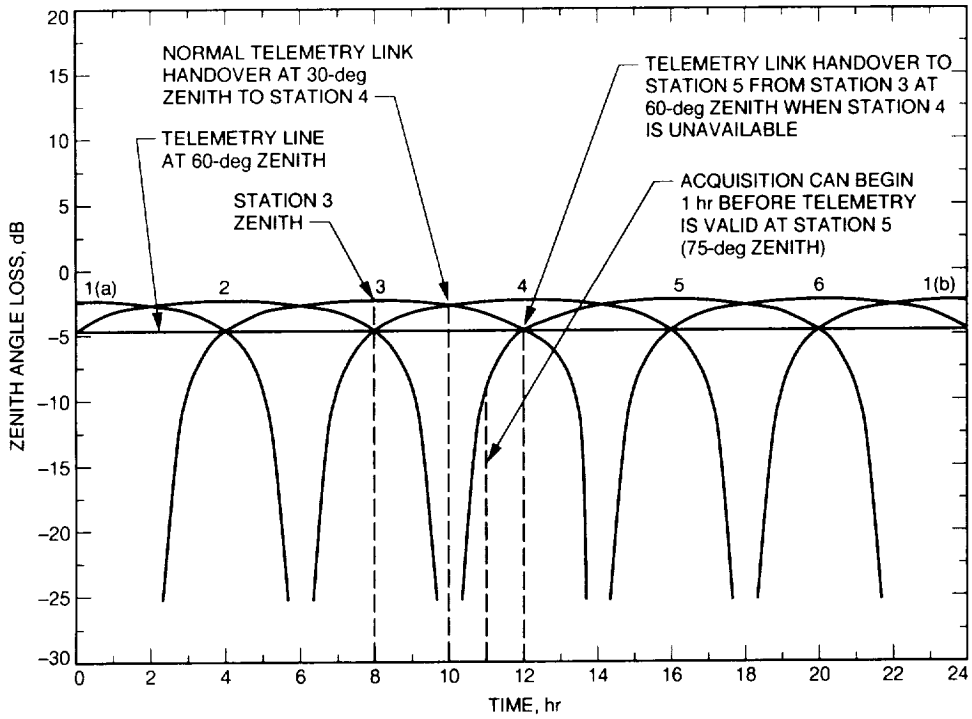


Fig. 12. Ideal-coverage curves over one day for an LDOS subnet with six stations 60 deg apart in longitude in an equatorial belt.

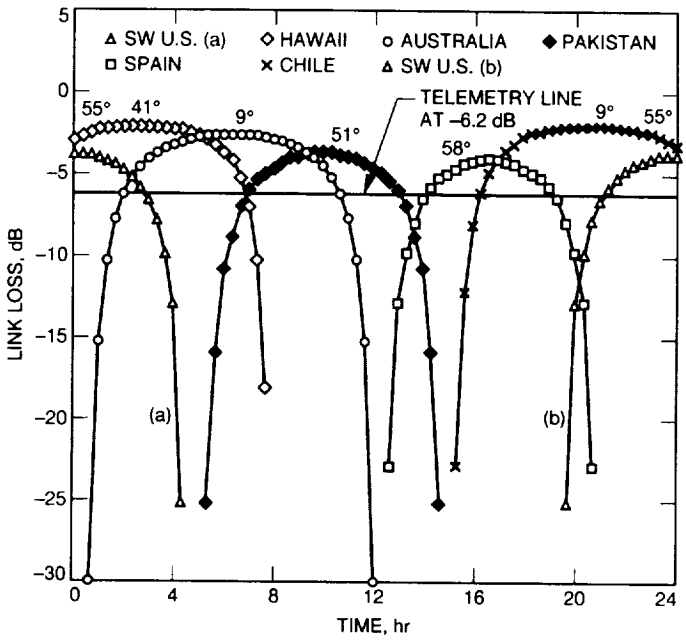


Fig. 13. Coverage curves for six actual sites for a Pluto link in 2015. Zenith angles at local meridian for Pluto in 2015 are shown at the top of each curve. The sites used are shown in Table 3. The coverage curve for the southwestern United States is shown in two halves: SW U.S. (a) and SW U.S. (b).

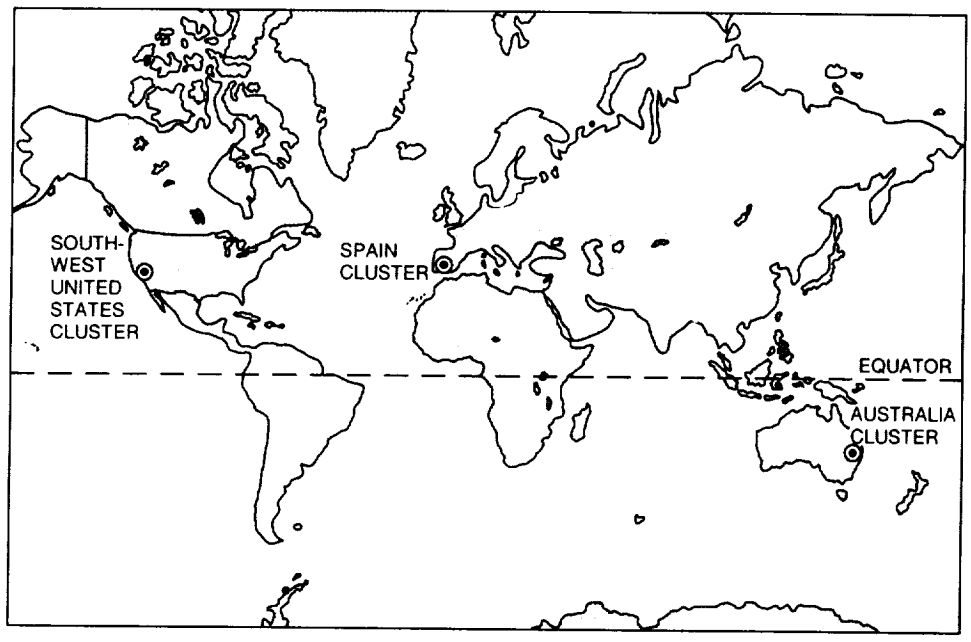


Fig. 14. Geographical sites for the COS 3 × 3 subnet.

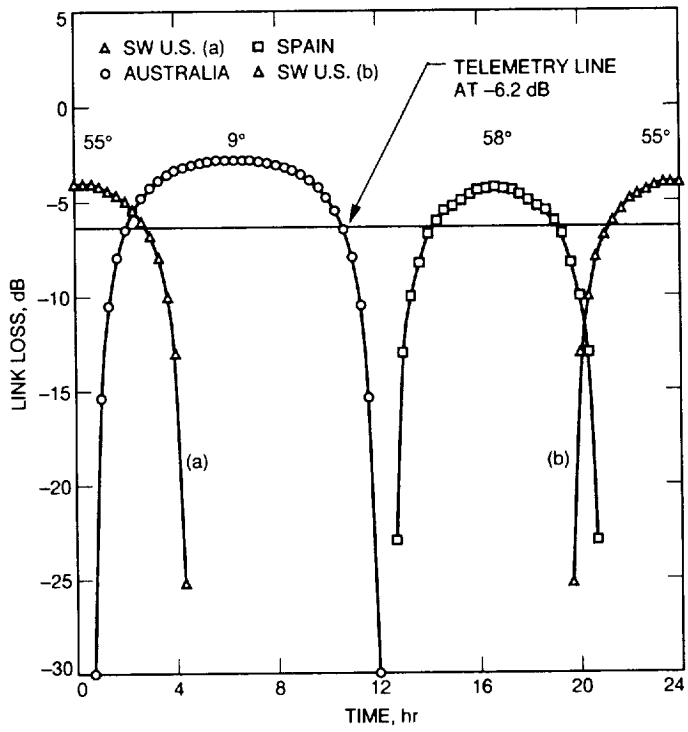


Fig. 15. Coverage curves for a COS 3 × 3 subnet with nine stations for a Pluto mission in 2015. Zenith angles at local meridian for Pluto in 2015 are shown at the top of each curve. The sites used to calculate the coverage curves are TMF in California, Siding Spring Mt. in Australia, and Calar Alto in Spain (see Table 3). The coverage curve for the southwestern United States is shown in two halves: SW U.S. (a) and SW U.S. (b).

Appendix A

OPTI Sample Output

OPTICAL COMMUNICATIONS LINK ANALYSIS PROGRAM
VERSION 4.02
GBATS, 30 AU, nighttime, 70° zenith angle, ARF spectral filter
PPM Direct Detection PMT detector

The transmitter parameters are (user spacecraft):

Transmitter average power (watts)	= 7.0000
Wavelength of laser light (micrometers)	= 0.53200
Transmitter antenna diameter (meters)	= 0.75000
Transmitter obscuration diameter (meters)	= 0.00000
Transmitter optics efficiency	= 0.80000
Transmitter pointing bias error (microrad.)	= 0.10000
Transmitter rms pointing jitter (microrad.)	= 0.10000
Modulation extinction ratio	= 0.10000E+06

The receiver parameters are (ground station):

Diameter of receiver aperture (meters)	= 10.000
Obscuration diameter of receiver (meters)	= 3.0000
Receiver optics efficiency	= 0.70000
Detector quantum efficiency	= 0.21000
Narrowband filter transmission factor	= 0.60000
Filter spectral bandwidth (angstroms)	= 0.10000E-01
Detector dia. field of view (microrad.)	= 100.00

The operational parameters are:

Alphabet size (M = ?)	= 256.00
Data rate (kb/s)	= 1387.8
Link distance (A.U.)	= 30.000
Required link bit error rate	= 0.13000E-01
Atmospheric transmission factor	= 0.24000
Dead time (microseconds)	= 3.2046
Slot width (nanoseconds)	= 10.000

Noise sources

Pluto	RCVR to source distance (AU)	= 30.000

Additional noise sources

nightsky	radiance(W/M**2/SR/A)	= .50000E-08

		Factor	dB
Laser output power (watts)		7.00	38.5 dBm
Min Req'd peak power (watts)	= .40E+04		
Transmitter antenna gain		0.160E+14	132.0
Antenna dia. (meters)	= 0.750		
Obscuration dia.(meters)	= 0.000		
Beam width (microrad)	= 1.121		
Transmitter optics efficiency		0.800	-1.0
Transmitter pointing efficiency		0.893	-0.5
Bias error (microrad)	= 0.100		
RMS jitter (microrad)	= 0.100		
Space loss (30.00 AU)		0.890E-40	-400.5
Receiver antenna gain		0.446E+16	156.5
Antenna dia. (meters)	= 10.000		
Obscuration dia. (meters)	= 3.000		
Field of view (microrad.)	= 100.000		
Receiver optics efficiency		0.700	-1.5
Narrowband filter transmission		0.600	-2.2
Bandwidth (angstroms)	= 0.010		
Detector Quantum efficiency		0.210	-6.8
Atmospheric transmission factor		0.240	-6.2
Received signal power (watts)		0.228E-11	-86.4 dBm
Recv'd background power (watts)	= 0.323E-17		
Photons/joule		0.268E+19	154.3 dB/mJ
Detected signal PE/second		0.255E+07	64.1 dBHz
Symbol time (seconds)		0.290E-05	-55.4 dB/Hz
Detected signal PE/symbol		7.36	8.7
Required signal PE/symbol		3.69	5.7
Detected background PE/slot	= 0.736E-04		
Margin		2.00	3.0

Appendix B
Site-Selection Guidelines and Procedures

I. Selection Guidelines

The following guidelines were used to identify probable sites for the Earth-based optical communication terminals:

(1) Locations as close to the equator as possible.

(2) High altitudes, preferably mountaintops.

(3) Good astronomical seeing.

(4) A large number of cloud-free days per year.

(5) Accessible locations with existing infrastructure, if possible.

II. Selection Procedure

To start, large geographical regions with an appropriate distance in longitude between them for the network configuration under consideration, and as close to the equator as possible, were identified on a map. A detailed literature search was then performed to locate sites at high altitudes in each region, thus generating a large list of likely station sites. Sites with good astronomical seeing, a large number of cloud-free days, and a preexisting infrastructure were favored. Inaccessible sites with wet weather were dropped from consideration when better alternates were available.

III. List of Additional Possible Sites

Table B-1 provides a list of geographical sites in addition to those already listed in the main text of this article. Each possible site in this table, and in the site tables shown elsewhere in this article, is followed by its altitude, longitude, latitude, and the time zone. The next column provides information on the number of cloud-free days and the weather of the site. The cloud-cover data on most sites were obtained from the International Satellite Cloud Climatology Project (ISCCP) as managed by the NASA Climate Data System (NCDS) and are available on CD-ROM [6]. The data provide monthly averages over an eight-year period ending in December 1990, for the entire globe, with a resolution of 250 km.[1] Data on other sites, like Mauna Kea in Hawaii, were obtained in situ for astronomical purposes. The last column indicates if there is a preexisting infrastructure at the site.

The lists of actual sites presented in this article should be treated as tentative and preliminary.

---

[1] K. Shaik and D. Wonica, "Cloud cover data for GBATS," JPL Interoffice Memorandum 331.6-93-098 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 6, 1993.

Table B-1. Additional sites of interest for an optical communications network.

Location	Altitude, km	Longitude, deg	Latitude, deg	Time zone	Cloud-free days/weather	Preexisting facilities and infrastructure
Roque de los Muchachos Observatory, Canary Islands, Spain	NA	16 W	29 N	−2	NA/dry	Yes
Fuente Nueva, La Palma Canary Islands, Spain	NA	16 W	29 N	−2	NA/dry	Yes
Jabal Toukal, Morocco	4.1	8 W	31 N	0	NA/dry	Information NA
Mulhecen, Spain	3.4	3 W	37 N	−1	67%/dry[a]	Information NA
Inaña, Tenerife, Canary Islands, Spain	NA	16 W	29 N	−2	NA/dry	Yes
Cerro Tololo, Chile	2.2	71 W	30 S	−4	77%/arid [7]	Yes
Llano del Hato, Venezuela	3.6	71 W	9 N	−4	NA/dry	Yes
Mt. Ziel, Australia	1.5	133 E	23 S	10	NA/dry	Information NA
Freeling Heights, Australia	1.1	139 E	30 S	10	NA/dry[b]	Information NA

[a] ISCCP satellite data, obtained from [6].
[b] A. Rogers, personal communication, Australian National University, Mount Stromolo and Siding Spring Observatories, Canberra, Australia, June 1993.