The Telecommunications and Data Acquisition Progress Report 42-117

January–March 1994

Joseph H. Yuen
Editor

May 15, 1994

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

This publication was prepared by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Preface

This quarterly publication provides archival reports on developments in programs managed by JPL's Office of Telecommunications and Data Acquisition (TDA). In space communications, radio navigation, radio science, and ground-based radio and radar astronomy, it reports on activities of the Deep Space Network (DSN) 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 other NASA program offices through and with the cooperation of OSC. The first of these is the Orbital Debris Radar Program funded by the Office of Space Systems Development. It exists at Goldstone only and makes use of the planetary radar capability when the antennas are configured as science instruments making direct observations of the planets, their satellites, and asteroids of our solar system. The Office of Space Sciences funds the data reduction and science analyses of data obtained by the Goldstone Solar System Radar. The antennas at all three complexes are also configured for radio astronomy research and, as such, conduct experiments funded by the National Science Foundation in the U.S. and other agencies at the overseas complexes. These experiments are either in microwave spectroscopy or very long baseline interferometry.

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 Telecommunications and Data Acquisition Progress Report* will present material in some, but not necessarily all, of the aforementioned programs.

DSN Systems Implementation
CAPABILITIES FOR NEW PROJECTS

NETWORK UPGRADE AND SUSTAINING

DSN Operations
FLIGHT–GROUND ADVANCED ENGINEERING

Contents

OSC TASKS
DSN Advanced Systems
TRACKING AND GROUND-BASED NAVIGATION

COMMUNICATIONS, SPACECRAFT–GROUND

The Mars Observer Differential One-Way Range Demonstration

P. M. Kroger and J. S. Border
Tracking Systems and Applications Section

S. Nandi
Earth Observations Analysis Systems Section

Current methods of angular spacecraft positioning using station differenced range data require an additional observation of an extragalactic radio source (quasar) to estimate the timing offset between the reference clocks at the two Deep Space Stations. The quasar observation is also used to reduce the effects of instrumental and media delays on the radio metric observable by forming a difference with the spacecraft observation (delta differential one-way range, ΔDOR). An experiment has been completed using data from the Global Positioning System satellites to estimate the station clock offset, eliminating the need for the quasar observation. The requirements for direct measurement of the instrumental delays that must be made in the absence of a quasar observation are assessed. Finally, the results of the "quasar-free" differential one-way range, or DOR, measurements of the Mars Observer spacecraft are compared with those of simultaneous conventional ΔDOR measurements.

I. Introduction

The state of an interplanetary spacecraft can be inferred from Doppler and range data recorded at a single Deep Space Station (DSS) of the DSN. The information content of single-station observations is mainly along the direction of the spacecraft line of sight, although the angular position can be inferred from the diurnal signature in the Doppler data [1]. Direct estimation of the angular position of a spacecraft may be accomplished with the use of station-differenced range data. One such technique has been adopted for operational use by the DSN [2,3]. Referred to as delta differential one-way range (ΔDOR), this method uses the recorded phase (range) of side tones or subcarrier harmonics of the spacecraft's telemetry signal to estimate the difference in the time of arrival, or delay, of these signals at a pair of Deep Space Stations (Fig. 1). As indicated in Fig. 1, an additional observation of an extragalactic radio source or quasar is required to form the final ΔDOR observable: a difference between the spacecraft and quasar delays. This "differential" delay is related to the positions of the spacecraft and quasar on the plane of the sky and the relative positions of the Deep Space Stations by

$$\Delta\tau_{\Delta DOR} = \tau_{SC} - \tau_{egrs} = \frac{\mathbf{B} \cdot (\hat{\mathbf{S}}_{SC} - \hat{\mathbf{S}}_{egrs})}{c} \quad (1)$$

where τ_{SC} is spacecraft delay, τ_{egrs} is quasar delay, c is the speed of light, $\hat{\mathbf{S}}_{egrs}$ and $\hat{\mathbf{S}}_{SC}$ are unit vectors in the direc-

tions of the quasar and spacecraft, respectively, and **B** is the baseline vector between the two Deep Space Stations.

The additional quasar observation is necessary for two reasons. Correlation of the broadband noise signal of the quasar recorded at the two Deep Space Stations allows the timing or "clock" offset between them to be estimated. Taking the difference between the spacecraft and quasar delays produces an observable, $\Delta\tau_{\Delta DOR}$, that is nearly free of media and instrumental delay effects common to both the spacecraft and quasar signals. It is this differencing which in fact allows this technique to measure spacecraft positions with an accuracy approaching 5 nrad in the quasar reference system.

The requirement of an additional quasar observation has several drawbacks that affect the utility of this technique in operational spacecraft tracking. First, the tracking antenna must point away from the spacecraft during the time of the quasar observation. The resulting loss of lock on the spacecraft carrier interrupts the transmission of telemetry data to the ground station during this period. Furthermore, the DSN must have a catalog of quasar positions to ensure that a sufficient number of natural radio sources of known source strength, whose positions are accurately known and near to the spacecraft trajectory, are available throughout the duration of each mission. Because a natural radio source's brightness changes with time, a continuing series of radio interferometric measurements of the source strengths is required to "maintain" this catalog. Processing the quasar signal to estimate the delay and clock offset also requires specialized hardware in the form of the Block 1 very long baseline interferometry (VLBI) correlator [4].

For these reasons, it would be valuable in terms of reduced cost and increased operational utility to be able to measure the angular position of a spacecraft without the need for an accompanying radio source observation, and this is the primary motivation for the work described here. What is required is an alternate means to accomplish what the quasar observation now does: measure the clock offset between the Deep Space Stations and remove the effects of instrumental and media delays from the spacecraft delay observable. A clock offset with an accuracy of 1 nsec, neglecting all other error sources, would yield an angular accuracy of about 50 nrad, sufficient for many future missions, especially during their cruise phase.

In the work described in this article, clock offsets determined by analysis of data from the Global Positioning System (GPS) of satellites recorded by receivers located at the Deep Space Stations have been used in place of the clock offsets normally obtained from quasar observations in conventional ΔDOR measurements. In addition, daily measurements of instrumental timing offsets at the Goldstone, California, and Canberra, Australia, Deep Space Communications Complexes (DSCCs) and the analyses of locally generated calibration tones were made to assess the stability of these delays on a day-to-day basis and determine whether it is feasible to use such measurements to calibrate instrumental effects without a quasar-differenced ΔDOR observable.

II. The Mars Observer Differential One-Way Range Demonstration

During the cruise phase of the Mars Observer (MO) mission, conventional ΔDOR measurements were made on a weekly basis as specified in the overall mission navigation plan.[1] Coincident with these measurements, an attempt was made to perform "quasar-free" differential one-way range (DOR) measurements by recording the phase of several spacecraft tones and combining these with estimates of the station time offsets obtained from analysis of data from the GPS satellites. An important part of this demonstration was the use of the Experimental Tone Tracker (ETT). This remotely controlled device consists of a modified TurboRogue GPS receiver with the capability to acquire and track spacecraft signals at either 2.3 GHz (S-band) or 8.4 GHz (X-band) in two separate baseband channels. ETTs were installed at both the Goldstone and Canberra Deep Space Stations to support this demonstration.

A. Experimental Description

Figure 2 is a schematic diagram of the experimental setup used for recording the Mars Observer spacecraft signals and identifies the instrumental delay measurements that were also made as part of the demonstration. Commands downloaded to the ETTs from JPL contain information on the predicted frequencies of the spacecraft tones that are used by the digital phase-locked loop to acquire the spacecraft signals. Figure 3 is a schematic diagram of the ETT and how it is connected to the radio frequency system of the DSN radio antenna. The radio metric data, consisting of the amplitude, phase, and frequency of the spacecraft signals, are stored in the memory of the ETT and subsequently transferred to a personal computer at JPL through a high-speed modem. The direct transfer of

[1] M. J. Rokey, ed., *Mars Observer Mission Plan*, Mars Observer Project Document 642-311 (internal document), Jet Propulsion Laboratory, Pasadena, California, July 1992.

the data and the ability to remotely control the receiver from JPL permit a very efficient method of acquiring and processing radio metric data, requiring a minimum number of personnel at JPL and the Deep Space Stations.

B. DOR Observable

The spacecraft delay, or DOR, observable is formed from the measured phases of spacecraft tones in the same manner as in a conventional ΔDOR measurement. A group delay is computed by taking the difference of the station-differenced phases of two spacecraft tones at their received frequencies and dividing this quantity by the transmitted frequency difference of the tones [3]:

$$\tau_{DOR}(\nu_i, \nu_j; t) = \frac{(\phi_1(\nu_i; t) - \phi_2(\nu_i; t)) - (\phi_1(\nu_j; t) - \phi_2(\nu_j; t))}{\nu_c(m_i - m_j)} \quad (2)$$

where $\phi_1(\nu_i; t), \phi_1(\nu_j; t), \phi_2(\nu_i; t)$, and $\phi_2(\nu_j; t)$ are the phases of tones i and j at time t at stations 1 and 2, ν_c is the transmitted carrier frequency, m_i and m_j are the multipliers that relate the carrier frequency to the side-tone frequencies, and τ_{DOR} is the spacecraft or DOR delay for tones i and j at time t.

A number of delay observables, with different delay ambiguities, can be formed from combinations of the four tones recorded for the MO measurements. The most precise delay is formed from the tone pair with the widest separation (23.1 MHz), but the delays from the more closely spaced tones are needed to resolve the delay ambiguities.

Knowledge of the instrumental delays introduced by the receiving system at the Deep Space Stations is critical in computing an accurate DOR delay observable. The schematic diagram in Fig. 4 shows how a number of instrumental delay terms and timing offsets contribute to the total DOR delay of Eq. (2) that can be expressed as:

$$\begin{aligned}\tau_{DOR}(\nu_i, \nu_j; t) &= \tau^0_{DOR}(\nu_i, \nu_j; t) \\ &+ \Delta\tau_{FE}(\nu_i, \nu_j; t) + \Delta\tau_I(\nu_i, \nu_j; t) \\ &+ \Delta\tau_C(t) - \Delta\tau_E(t) + \Delta\tau_{\text{misc}}(t) + \eta \end{aligned} \quad (3)$$

where τ_{DOR} is the observed spacecraft delay and τ^0_{DOR} is the model delay computed from a priori knowledge of the geometry of the spacecraft and the locations of Deep Space Stations. The other terms in Eq. (3) are

(1) $\Delta\tau_{FE}$ is the instrumental delay above the phase-calibration tone injection point. This also includes the delay between the antenna reference point used to compute the baseline vector, **B**, and the phase-calibration injection point.

(2) $\Delta\tau_I$ is the instrumental delay from the phase-calibration tone injection point to the ETT data sampler.

(3) $\Delta\tau_C$ is the timing or clock offset between the master clocks of the Deep Space Stations.

(4) $\Delta\tau_E$ is the timing offset between the ETT data sampler and the station's master clock.

(5) $\Delta\tau_{\text{misc}}$ represents a number of miscellaneous delays including electronic delays within the Rogue GPS receiver and the Experimental Tone Tracker.

(6) η is a random quantity representing all measurement errors.

As indicated in Eq (3), these delays may be both time and frequency dependent. It was far beyond the scope of this first DOR demonstration to measure all the terms listed above. The DOR delay of Eq. (2) measured in this demonstration is therefore "biased" by the uncalibrated delays and timing offsets. It is crucial to the interpretation of the results of this demonstration to understand what terms in Eq. (3) were not measured. These included items (1) and (5) from the above list. The total magnitude of these terms may be on the order of 1 μsec, but their variation is expected to be less than 1 nsec. The DOR bias could easily be estimated by comparing the DOR and ΔDOR measurements, as was done in this demonstration (see Section VI). The bias should be independent of the spacecraft and, as shown in Section VI, remain constant over long periods of time.

The measurement of the instrumental delay, $\Delta\tau_I$, provided by the phase-calibration tones was itself biased by the uncalibrated uplink delay, τ_u, and the timing offset between the station master clock, τ_C, and the clock reference of the calibration tone generator, τ^{ϕ}_C , as shown in Fig. 4. Hence, the delay measured by the phase-calibration tones is actually

$$\tau_\phi = \tau_I + \tau_u - \tau_E - \tau_C^\phi \quad (4)$$

where τ_ϕ is the delay computed from analysis of the phase-calibration tones, τ_u is the uplink delay, τ_E is the timing offset between the ETT and the station master clock, and τ_C^ϕ is the offset between this clock and the reference clock of the tone generator. A complete discussion of the instrumental delays computed from the phase-calibration tones is provided in Section V.B.

It should be noted that the ΔDOR observable is not contaminated by these instrumental delays and timing offsets because they are common to both the spacecraft and quasar signals and are removed when these delays are differenced, as in Eq. (1). A comparison of the ΔDOR delays with the DOR delays will therefore provide an estimate of the magnitude and variability of the uncalibrated delay terms of Eq. (3).

III. Mars Observer DOR Measurements

Table 1 is a summary of the Mars Observer DOR measurements that were completed during the period March 21, 1993, to August 16, 1993. Each measurement consisted of two spacecraft scans separated by the quasar observation required by the concurrent conventional ΔDOR observations.

During the spacecraft observations, the ETT receiver was programmed to record the amplitude and phase of four spacecraft tones, including the 8.4-GHz carrier and three side or "DOR" tones. Figure 5 shows the baseband frequencies (ν_{BB}) of these tones within the 40-MHz intermediate frequency (ν_{IF}) bandpass of the DSN intermediate frequency (IF) system.

IV. GPS Clock Offsets

The $\Delta\tau_C$ term in Eq. (3), representing the timing offset between the Deep Space Stations, is expected to be one of the largest and most time variable of the delays in this equation. Though the other delay terms in Eq. (3) may be of comparable magnitude, they are expected to remain relatively constant over long periods of time and need be measured only infrequently. The time or "clock" offset, $\Delta\tau_C$, however, must be directly measured during the actual DOR measurement since its time variability can be large and it is not predictable at the nanosecond level. Fortunately, the arrival of the GPS system of satellites provides a way to directly measure the time offset between widely separated sites with subnanosecond accuracy [5]. For this demonstration, analysis of data from a global network of GPS receivers [6], which included the DSN Rogue receivers, was used to estimate the timing offset between the Goldstone and Canberra reference clocks at the time of each MO DOR observation. To relate the GPS estimated clock solutions to the station master clock, the time delay between the GPS receiver and the Deep Space Station's reference clock (τ_R in Fig. 4) was also measured.

Figure 6 shows estimates of the clock offset between the Canberra and Goldstone reference clocks obtained by combining the Rogue GPS clock solutions and the measured time delay, τ_R, between the Rogue receivers and the stations' clock reference point (Fig. 2). The Rogue receiver clock offsets have been corrected for the timing offset between the station master clock (Fig. 4) and the Rogue receiver. The abrupt changes in the value of the Rogue GPS clock offsets are most likely due to errors in the measurement of this timing offset and do not reflect changes in the Rogue GPS clock solution. The Rogue clock offsets are determined from the analysis of GPS data from a worldwide network of over 40 receivers. Also shown on this plot are the clock offsets determined by the DSN's own GPS measurement system, which uses the more conventional, but less accurate, "common view" technique.[2] The greater accuracy of the clock offsets estimated from the Rogue receiver data is evident from the much lower scatter in these clock offsets compared to those determined by the frequency and timing subsystem (FTS). An examination of differences between the FTS and Rogue clock offsets shows a scatter of approximately 15 nsec. That this scatter is dominated by the FTS clock estimates was confirmed by examination of the residuals to a polynomial fit of a 3-day span of Rogue clock offsets. The scatter in the Rogue residuals was approximately 0.2 nsec, consistent with the formal errors reported for these clock solutions.

The large jumps in the Rogue receiver clock offsets that are apparent in Fig. 6 are artifacts introduced by the measurement of the τ_R delay shown in Fig. 4 and do not represent true changes in the clock offset. No abrupt changes are seen in the FTS clock offsets, which use a single value for the equivalent τ_R delay term.

V. Instrumental Delay Calibrations

A. Timing Offset Measurements

Although it was not possible to measure all the delays and timing offsets shown in Fig. 4 and Eq. (3), a number of

[2] G. A. Santana, personal communication, TDA Mission Support and DSN Operations Section, Jet Propulsion Laboratory, Pasadena, California, December 13, 1993.

timing offsets were measured on a daily basis throughout the 8-month span of the Mars Observer cruise phase. The primary purpose of these daily measurements was to assess the long- and short-term variability of these and similar timing offsets. Figure 7 shows how the measurements were made using a time-interval counter with an accuracy of 100 psec. It was expected that the timing offsets would remain constant at a level of 1 nsec or less over the course of these measurements.

Figure 8, for example, shows the results of the daily measurement of the τ_E delay of Eq. (3) during the period from February 15, 1993, through August 16, 1993, for both the Canberra and Goldstone Deep Space Stations. The measurements were performed with the experiment setup shown in Fig. 7. The day-to-day variability in this delay, as measured at the Canberra Deep Space Station, is extremely small, probably less than 1 nsec. There does appear to be a slight long-term increase in this delay that is not presently understood. In the case of the Goldstone station, there initially appeared to be large variations in this delay in comparison to the Canberra results. After day of year (DOY) 162, however, this variability greatly decreased for reasons yet unknown. The greater variability in the Goldstone delays is probably a result of the measurement procedure and is not a true measure of the variability of this delay. The timing offset measurements at Canberra were completely automated. At Goldstone, these measurements were made manually each day by different station personnel. There are certainly no a priori reasons to expect the large variations in these timing measurements that were reported at the Goldstone Deep Space Station. It is now believed that the measurement methodology was not subject to strict configuration control.[3] This was caused by noise in the 1-pulse/sec output of the Rogue receiver at Goldstone that necessitated adjustments to the trigger level of the time interval counter, which may have contributed to the large variation seen in the timing offsets measured at Goldstone in the early part of this demonstration.

B. Instrumental Delay Measurements

The delay introduced into the spacecraft signal as it traverses the electronic components of the receiving system before being sampled and recorded in the ETT memory is represented by the term $\Delta\tau_I$ in Eq. (3). This delay cannot be directly measured in the manner of the timing offsets discussed above. Instead, a comb of calibration tones separated by 0.1 MHz and spanning the entire bandwidth of the radio-frequency system is injected near the antenna front end (Fig. 3).

These "phase-calibration" tones experience the same instrumental delays as the spacecraft tones, and their phase and amplitude can be recorded by the ETT receiver in the same manner. The use of these tones to calibrate instrumental delays has been a part of the DSN VLBI system since its inception [7]. Because they are derived from the station's hydrogen maser frequency standard, the tones should exhibit stable amplitude and phase over extended periods of time. As discussed in Section II.B, the delay that is measured through analysis of the phase-calibration tone data is not identically the τ_I term of Eq. (3), but is the biased delay of Eq. (4).

In the MO DOR demonstration, the only information on the instrumental delay term τ_I comes from measurement of the phase-calibration tones. During the actual Mars Observer DOR observations, these tones were turned off to prevent the stations' telemetry receivers from inadvertently locking to one of these tones instead of the spacecraft carrier. Measurements of the phase-calibration tones were restricted to the "precalibration" period, which might precede the spacecraft observations by several hours, under the assumption that the instrumental delays will not undergo significant changes in the period before and during the 40-min ΔDOR measurement that follows. This assumption was tested and confirmed by independent measurements of the phase and amplitude of these tones over periods of several hours, where it was shown that the relative phases of the tones remained constant over periods of several hours.[4]

In spite of this, significant problems were faced in the use of the phase-calibration tones to measure the instrumental delay term of Eq. (3) apart from the bias discussed earlier. Initially, it was believed that the instrumental phase could be calibrated by measuring the phase of calibration tones whose baseband frequencies were close to the spacecraft tone frequencies. This method was discussed in an earlier status report on the MO DOR demonstration.[5] The basic idea, as described there, is to compute a mean spacecraft frequency for each DOR tone and interpolate the instrumental phase correction from the measured phases of calibration tones that are close in fre-

[3] M. Manning, personal communication, TDA Mission Support and DSN Operations Section, Jet Propulsion Laboratory, Pasadena, California, January 21, 1994.

[4] S. Nandi, P. M. Kroger, and J. S. Border, "Mars Observer Differential One-way Ranging Experimental Investigation: Status as of 6-4-93," JPL Interoffice Memorandum 335.1-93-20, Jet Propulsion Laboratory, Pasadena, California, June 28, 1993.

[5] Ibid.

quency to the spacecraft tone frequency. In this case, the instrumental delay correction is given by

$$\Delta\tau_\phi(\nu_i, \nu_j) =$$

$$\frac{[\phi_{cal,1}(\nu_i) - \phi_{cal,1}(\nu_j)] - [\phi_{cal,2}(\nu_i) - \phi_{cal,2}(\nu_j)]}{\nu_c(m_i - m_j)} \quad (5)$$

where $\Delta\tau_\phi$ is the estimated instrumental delay correction, $\phi_{cal,1}$ and $\phi_{cal,2}$ are phases obtained by interpolating the measured phase-calibration tone phases at stations 1 and 2 (after the cycle ambiguities have been resolved) to the mean spacecraft tone frequencies ν_i and ν_j, ν_c is the best estimate of the spacecraft carrier frequency, and m_i and m_j are the multipliers that relate the carrier frequency to the DOR tone frequencies for tones i and j. It was found that delays computed in this manner for a number of tone pairs showed variations that were much larger than expected.[6] That this could be attributed to errors in the ETT tone phase measurement was confirmed to some degree by comparison of calibration tone phases measured by the ETT with those measured simultaneously by the DSN's narrow-band VLBI system during regular VLBI clock-synchronization measurements. This comparison indicated the presence of errors in the ETT's calibration tone-phase measurement on the order of several hundredths of a cycle (Figs. 9 and 10). Phase errors of this magnitude can translate into instrumental delay errors that exceed the 1-nsec goal of this demonstration. Based upon Eq. (5), the relation between the phase error and the delay error is given by

$$\sigma_\tau(\nu_i, \nu_j) = \frac{\sqrt{2}\sqrt{2}\sigma_\phi}{\nu_i - \nu_j} \quad (6)$$

The maximum frequency spacing for the MO DOR tones is approximately 23.1 MHz. This requires a phase accuracy of at least 0.01 cycle to achieve a 1-nsec delay error. The comparison of the ETT phase measurement with the DSN's narrow-channel bandwidth (NCB) VLBI measurements of the same tone phases, however, appears to indicate that ETT tone phases may have errors as large as 0.05 cycle. The reasons for errors of this magnitude are not completely understood, but may be related to the 1-bit data quantization and the 3-level tone stopping function that is used in the ETT, compared to the 128-level stopping function used in the NCB VLBI system. The fact that each ETT channel contains 160 phase-calibration tones (and all of their intermodulation products) while each NCB channel contains only 2 tones may also contribute to errors in the ETT tone-phase measurement. In any case, the above method of instrumental delay calibration did not provide 1-nsec delay accuracy.

Until the problems with the ETT phase measurements are resolved, the use of the phase-calibration tone phases recorded by the ETT to compensate for instrumental group delays remains problematic. For the purposes of this demonstration, all the calibration tone phases recorded in each ETT channel were simply fit to a line whose slope was taken as an estimate of the group delay for that channel. This was done for the phase-calibration tones recorded during the precalibration period of all DOR measurements listed in Table 1. Figure 11 is a typical plot produced from calibration tone phases recorded during the precalibration period of the June 8, 1993, DOR measurement. Each point in this plot represents a group delay computed from a linear fit of five calibration tone phases to their baseband frequencies. Abrupt changes in the group delays of 10 to 15 nsec are seen in most of the DOR measurements and are likely caused by changes to the radio-frequency signal path that occur during the precalibration period as the receiving equipment is being prepared for the Mars Observer tracking pass.

Table 2 contains the estimated group delays computed from linear fits of the calibration tone phases to their baseband frequencies in each ETT channel. Each number represents the mean value taken over the indicated measurement interval. It is evident from these results that the group delays are quite sensitive to the number of calibration tones. In the first half of the MO DOR demonstration, only three tones were recorded in channel 2 of the ETT compared to five tones in channel 1. The delays computed using only three tones (before July 4, 1993) differ in magnitude from the delays computed with five tones by as much as 30 to 50 nsec. After July 4, 1993, when five tones were recorded in both ETT channels, the group delays in each channel are much more consistent with an average difference of approximately 3.2 nsec, on the same order as the variation in the channel group delays between DOR measurements. These problems are not completely understood, but are probably related to the errors in the calibration tone phases measured by the ETT mentioned above. Because of this, it was decided not to use the group delays computed from the phase-calibration tones to calibrate the instrumental delay term, $\Delta\tau_I$ of Eq. (3), but rather to assume that it would contribute a constant bias along with the other uncalibrated delay terms in this equation.

[6] Ibid.

C. ETT Channel Phase Offset

A noninteger phase offset was found to exist between the two baseband channels of the experimental tone tracker at Canberra. This is apparent when spacecraft group delays are computed from the measured phases of the spacecraft tones and is manifest as a systematic difference in the magnitude of the group delays computed from tones within a single channel compared with those computed using tones in different ETT channels. The magnitude of the ETT channel phase offset at Canberra can be estimated from the phase-calibration data recorded immediately before each DOR pass by examination of residuals to linear fits of calibration tone phase to frequency.

Figure 12 shows residuals from a linear fit of calibration tone phase and frequency for data taken on August 15, 1993, at Goldstone and Canberra. These data indicate a phase offset of 0.27 cycle for the ETT at Canberra. There does not appear to be a significant phase offset between the channels of the ETT at Goldstone.

VI. Comparison With Conventional ΔDOR Measurements

Figure 13 shows the uncalibrated delay residuals for all successful DOR measurements that were completed during the first 8 months of 1993 (Table 1). No Rogue GPS clock solutions were available on June 8, 1993. The GPS clock solutions on days June 15, 1993 (day number 166), June 19, 1993 (day number 170), and June 23, 1993 (day number 174), are believed to be corrupted by bad timing offset measurements (see Section IV). As expected, these residuals show a significant signature due to the presence of instrumental and clock delays. That the majority of this signature is due to the clock offset between the DSN antennas is evident by comparing the general trend of the DOR delay residuals with the clock offsets estimated from GPS data that are also shown in this figure. Figure 14 shows the corrected DOR delay residuals produced by subtracting the clock offsets obtained from the GPS data that include the timing offset, τ_R, that relates the GPS clock to the station master clock. All delays in this figure correspond to the most widely separated tone pair ($\Delta\nu = 23.1$ MHz). The large biases in the quasar-free group delays are due to uncalibrated instrumental effects. The parentheses around the DOR points in this figure indicate that the GPS clock offsets for these points were computed by interpolation from nearby clock offsets. This was necessary because of problems in the timing offset measurements on these days (Section IV).

According to Eq. (3), the DOR delay residuals shown in Fig. 14 can be represented by

$$\begin{aligned}\delta\tau_{DOR}(\nu_i,\nu_j;t) &= \tau_{DOR}(\nu_i,\nu_j;t) - \tau^0_{DOR} - \Delta\tau_C(t) \\ &= \Delta\tau_{FE}(\nu_i,\nu_j;t) + \Delta\tau_I(\nu_i,\nu_j;t) \\ &\quad - \Delta\tau_E(t) + \Delta\tau_{\text{misc}}(t) + \eta(t) \qquad (7)\end{aligned}$$

where the δ designates a residual delay. In so far as the quantities on the right side of Eq. (7) are constant, the residual DOR delay, $\delta\tau_{DOR}$, should remain constant, and, in so far as these quantities are antenna pair dependent, so should the residual DOR delays be antenna dependent. That this is so is obvious from Fig. 14, where the residual DOR delays clearly fall into groups corresponding to the antenna pair used in the measurement.

The variations in the magnitudes of the residual delays, expressed as a sample standard deviation about the mean for the three antenna pairs used in this demonstration, are shown in Fig. 14 along with the unbiased ΔDOR delays. The scatter in the DOR delay residuals is about 5 to 7 nsec, albeit for rather small sample sets. The "flatness" of the DOR residuals is encouraging since it confirms our assumption that variations in the instrumental and miscellaneous delays and timing offsets remain relatively constant over long periods of time.

However, a delay measurement with a precision of 5 to 7 nsec, corresponding to an angular precision of approximately 250 to 350 nrad, would not provide a useful spacecraft navigation data type. The origin of this poor precision must lie in one of the terms on the right side of Eq. (7). For example, the formal errors on the Rogue receiver clock offsets, $\Delta\tau_{GPS}$, are less than 1 nsec, but the scatter in the timing offset measurement, τ_R, that relates the GPS clock to the station master clock, τ_C, is about 1.5 to 2.0 nsec. As mentioned earlier (Section V.A), there are also indications that strict configuration control was not applied to the τ_R calibrations by the DSN station personnel at Goldstone,[7] but it is unclear what this might have contributed to the scatter in the timing offsets.

An assumed variation of 2.0 nsec in the $\Delta\tau_C$ term still leaves a delay scatter of 3 to 6.7 nsec. The most likely origin of this scatter lies in variations of the instrumental delay, $\Delta\tau_I$. Indeed, the group delays computed from the

[7] M. Manning, op. cit.

phase calibration tones in Table 2 show variations on this order.

Variations in the phase offset between the ETT channels, discussed in Section V.C, could introduce additional noise into the delays computed from spacecraft tone pairs if each tone were in a different channel. However, a comparison of residual delays computed using the two side tones with 7.4-MHz separation in ETT channel 1 (Fig. 5) with those computed from the most widely spaced tone pair (23.1 MHz) where each tone lies in a different ETT channel shows that the delay variation due to the phase offset is only approximately 0.5 nsec, and cannot be responsible for the observed scatter in the residual delays of Fig. 14.

VII. Summary and Conclusions

The primary goal of the Mars Observer DOR demonstration was to assess the feasibility of measuring spacecraft angular position with 50-nrad accuracy without the need for an additional quasar observation. Spacecraft differential range measurements with approximately 0.2-nsec precision (i.e., DOR measurements) were obtained in near-real time using a new closed-loop receiver (experimental tone tracker) at each Deep Space Station. As expected (Section II.B), comparisons of the DOR measurements with simultaneous conventional ΔDOR measurements of the Mars Observer spacecraft show a large but constant bias due to uncalibrated instrumental delays and timing offsets. The DOR delays fall into groups corresponding to the particular antenna pair used in the DOR measurement. Within each group, the variation in the delay residuals over the course of the 5-month demonstration ranges from 3 to 7 nsec (Fig. 14).

This relatively large scatter in the DOR delay residuals is believed to result primarily from uncalibrated variations in the instrumental delays of the receiving systems at each complex. Although measurements of the phase-calibration tones should have been able to compensate for these variations, the station instrumental delays were not measured well in this demonstration (Section V.B). The calibration tone-phase data that were obtained, however, are consistent with delay variations of this order (Table 2).

The station instrumental delay calibrations could be improved by

(1) Improving the reliability of the station phase-calibration system (calibration data in this demonstration were obtained in less than 60 percent of the scheduled measurements).

(2) Maintaining a constant tone power level during every measurement (configuration control).

(3) Reducing the total number of tones in each channel by using a wider tone spacing in the phase-calibration tone generator [7].

(4) Modifying the ETT to use 8-bit sampling of the signal.

(5) Modifying the ETT to use an 8-bit phase model in the tone stopping function.

None of the above items presents serious technical challenges, and they are well within the capability of the DSN. The 8-bit data sampling and stopping function have already been implemented in similar hardware developed as part of the Galileo antenna arraying project.[8] The overall station-pair-dependent DOR delay bias could be determined by occasional ΔDOR measurements since that part of the bias is expected to remain constant. ΔDOR measurements on one spacecraft should suffice to determine this bias for DOR measurements on other spacecraft.

Finally, it should be mentioned that one of the results of this work is the demonstration of interstation time synchronization at the subnanosecond level (Section IV). Although this was not the primary goal of the Mars Observer DOR demonstration, it is a necessary part of the DOR concept, and in itself constitutes an important result.

[8] D. H. Rogstad, personal communication, Tracking Systems and Applications Section, Jet Propulsion Laboratory, Pasadena, California, February 18, 1994.

Acknowledgments

It would have been extremely difficult to complete the DOR measurements described here without the experimental tone tracker and the many individuals who contributed to the design, construction, and installation of this equipment at Goldstone and Canberra; these included Elliot Sigman of the Processor Systems Development Group and Jeff Srinivasan and Scott Stephens of the GPS Systems Group. Larry Young participated in helpful discussions during the preparation of this article. Individuals at the Deep Space Stations at Goldstone and Canberra provided the timing offset measurements, and Dave Jefferson of the Earth Orbiter Systems Group provided the Rogue GPS clock solutions from the FLINN analysis.

References

[1] D. W. Curkendall and S. R. McReynolds, "A Simplified Approach for Determining the Information Content of Radio Tracking Data," *Journal of Spacecraft and Rockets*, vol. 6, no. 5, pp. 520–525, May 1969.

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

[3] J. S. Border and J. A. Koukos, "Technical Characteristics and Accuracy Capabilities of Delta Differential One-Way Ranging (ΔDOR) as a Spacecraft Navigation Tool," *Proceedings of the Meeting of the Consultative Committee for Space Data Systems on Radio Frequency and Modulation Standards Approval*, Munich, Germany, September 20, 1993 (in press).

[4] K. M. Liewer, "DSN Very Long Baseline Interferometry System Mark IV-88," *The Telecommunications and Data Acquisition Progress Report 42-93, vol. January–March 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 239–246, May 15, 1988.

[5] 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, July–September 1992*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–9, November 15, 1992.

[6] J. F. Zumberge, D. C. Jefferson, G. Blewitt, M. B. Heflin, and F. H. Webb, "Jet Propulsion Laboratory IGS Analysis Center Report, 1992," *Proceedings of the 1993 IGS Workshop*, G. Beutler and E. Brockmann, eds., Astronomical Institute, University of Berne, pp. 154–163, 1993.

[7] J. B. Thomas, "The Tone Generator and Phase Calibration in VLBI Measurements," *The Deep Space Network Progress Report 42-44, vol. January and February 1978*, Jet Propulsion Laboratory, Pasadena, California, pp. 63–74, April 15, 1978.

Table 1. Summary of Mars Observer DOR measurements.[a]

Experiment times			Antenna		Phase calibration		GPS clock offset (DSCC 10/DSCC 40), nsec	Residual delay ($\Delta\nu$ = 23.1 MHz), nsec[b]
DOY	Start	Stop	DSCC 10	DSCC 40	DSCC 10	DSCC 40		
80	0540	0620	14	45	No	Yes	2198.773	537.9569
93	0500	0540	15	45	Yes	No	2221.496	413.5409
147	0530	0610	14	45	Yes	Yes	1775.489	115.675
150	0510	0550	14	45	Yes	Yes	1772.710	118.385
159	0435	0515	14	45	Yes	Yes	—[c]	121.562
166	0420	0500	15	43	No	Yes	2832.140	90.711
170	0405	0445	14	45	Yes	Yes	2247.283	100.95
174	0200	0240	14	45	Yes	No	2241.790	98.74
178	0425	0505	15	43	No	Yes	1756.806	58.725
185	0215	0255	14	45	No	Yes	1729.440	65.64
199	0320	0400	15	43	Yes	No	1728.658	21.906
206	0255	0335	15	43	Yes	Yes	1694.939	4.499
213	0100	0140	15	43	Yes	Yes	1684.606	−8.735
219	0035	0115	15	43	Yes	Yes	1662.329	−34.839
227	0235	0315	15	45	Yes	Yes	1626.346	67.742
228	0230	0310	15	43	Yes	Yes	1623.552	−75.187

[a] Includes only those measurements where spacecraft signal was acquired at both stations.
[b] Difference between observed and model delay.
[c] No GPS clock offset available on this day.

Table 2. Summary of ETT instrumental group delays.

Phase calibration measurement interval			ETT channel group delay, nsec							
			DSS 14		DSS 15		DSS 43		DSS 45	
DOY	Start	Stop	Ch 1	Ch 2	Ch 1	Ch 2	Ch 1	Ch 2	Ch 1	Ch 2
The following measurements used five phase-calibration tones in ETT channel 1 and three tones in ETT channel 2										
80	0446	0451							−5930.58	−5896.97
92	2318	2323			−5234.54	−5219.14				
147	0518	0519	−3296.37	−3356.80						
147	0431	0436							−5930.59	−5932.58
150	0451	0456	−3288.76	−3277.38						
150	0338	0342							−5293.45	−5970.15
159	0418	0423	−3296.54	−3361.55						
159	0223	0228							−5929.61	−5968.96
166	0358	0403					−3391.83	−3399.62		
170	0336	0341	−3291.58	−3273.43						
170	0346	0349							−5927.88	−5976.65
174	0127	0131	−3296.27	−3233.94						
178	0400	0403					−3393.01	−3402.87		
The following measurements used five phase-calibration tones in each ETT channel										
185	0146	0148							−5930.56	−5925.18
199	2101	2104			−5234.62	−5237.89				
206	1756	1759			−5222.15	−5232.48				
206	0234	0239					−3395.34	−3396.64		
213	1600	1605			−5235.38	−5236.14				
213	0128	0133					−3393.21	−3392.58		
219	1647	1652			−5234.23	−5234.25				
219	0026	0031					−3400.97	−3402.43		
227	1706	1708			−5222.97	−5234.52				
227	0204	0209							−5930.71	−5930.11
228	1901	1906			−5238.23	−5237.86				

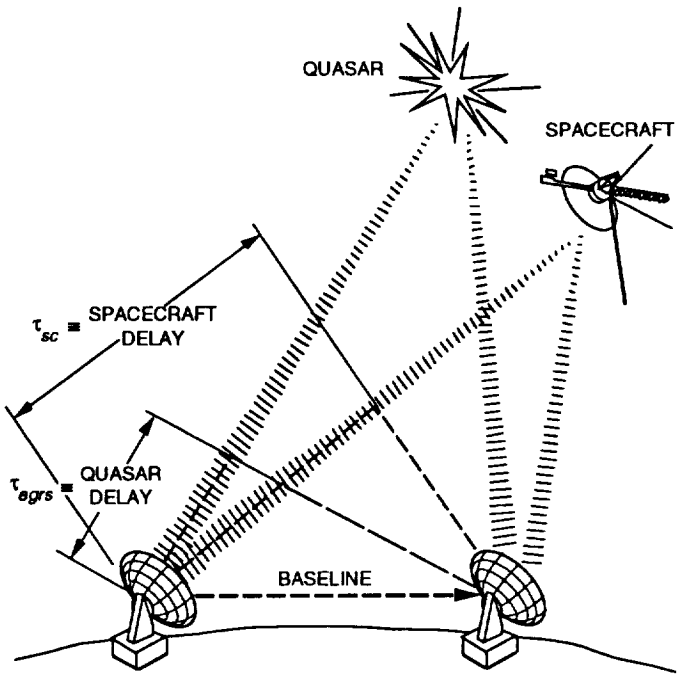


Fig. 1. Geometry of a typical ΔDOR observation.

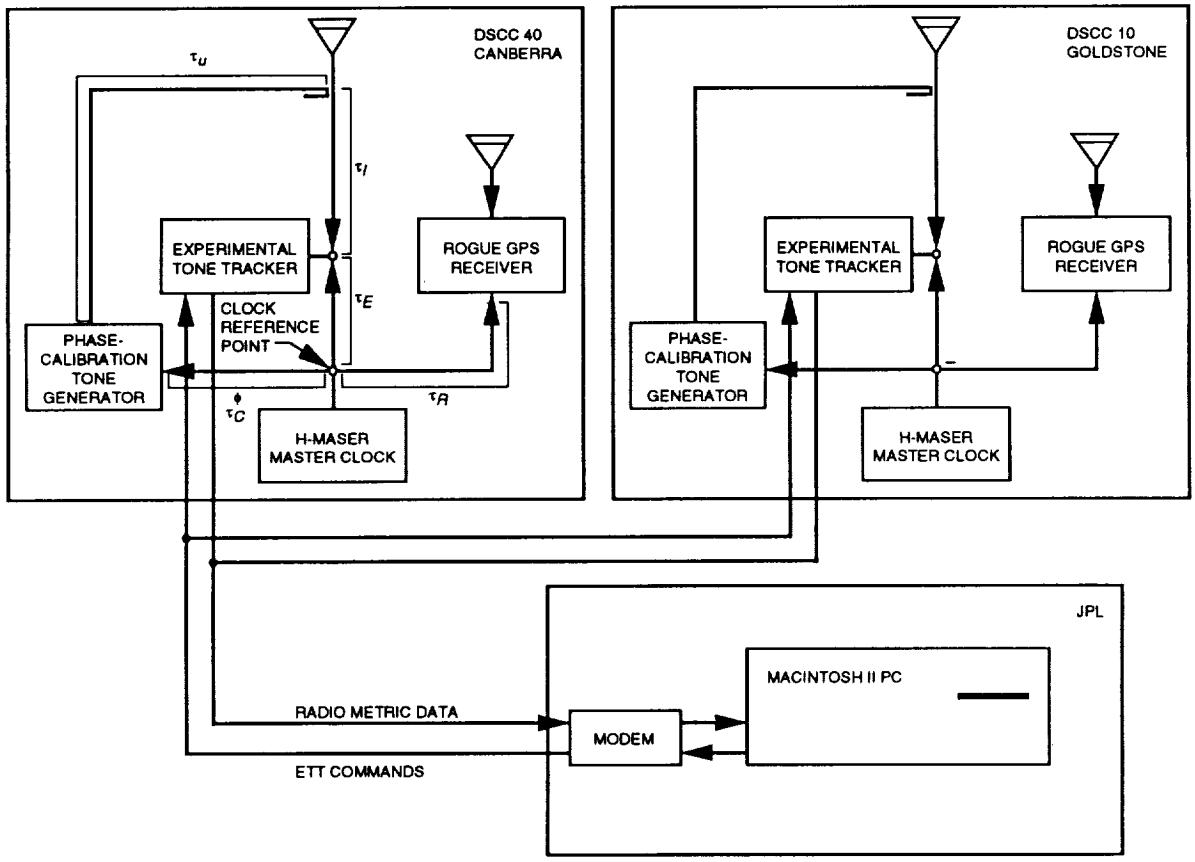


Fig. 2. Experimental setup used at DSCC 10 and DSCC 40 during the Mars Observer DOR demonstration. Various instrumental timing offsets are defined in Section V.

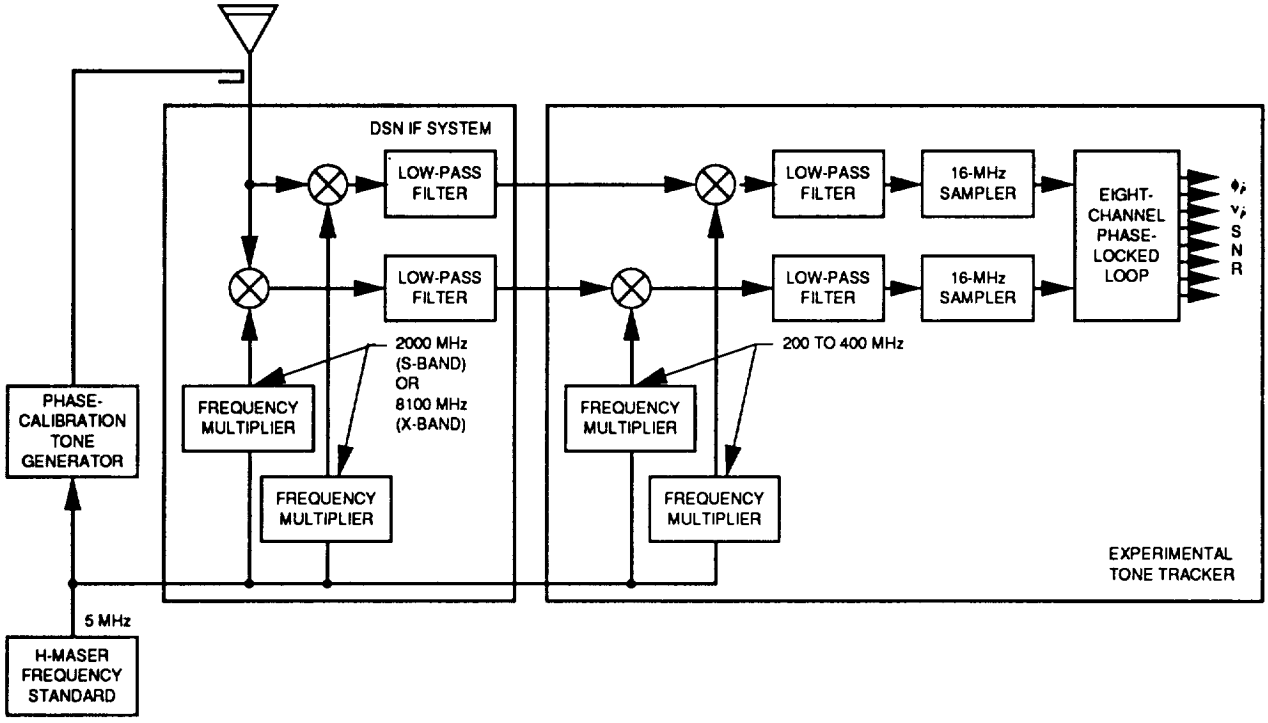


Fig. 3. The ETT tone tracker and its connection to the DSN antenna receiving system.

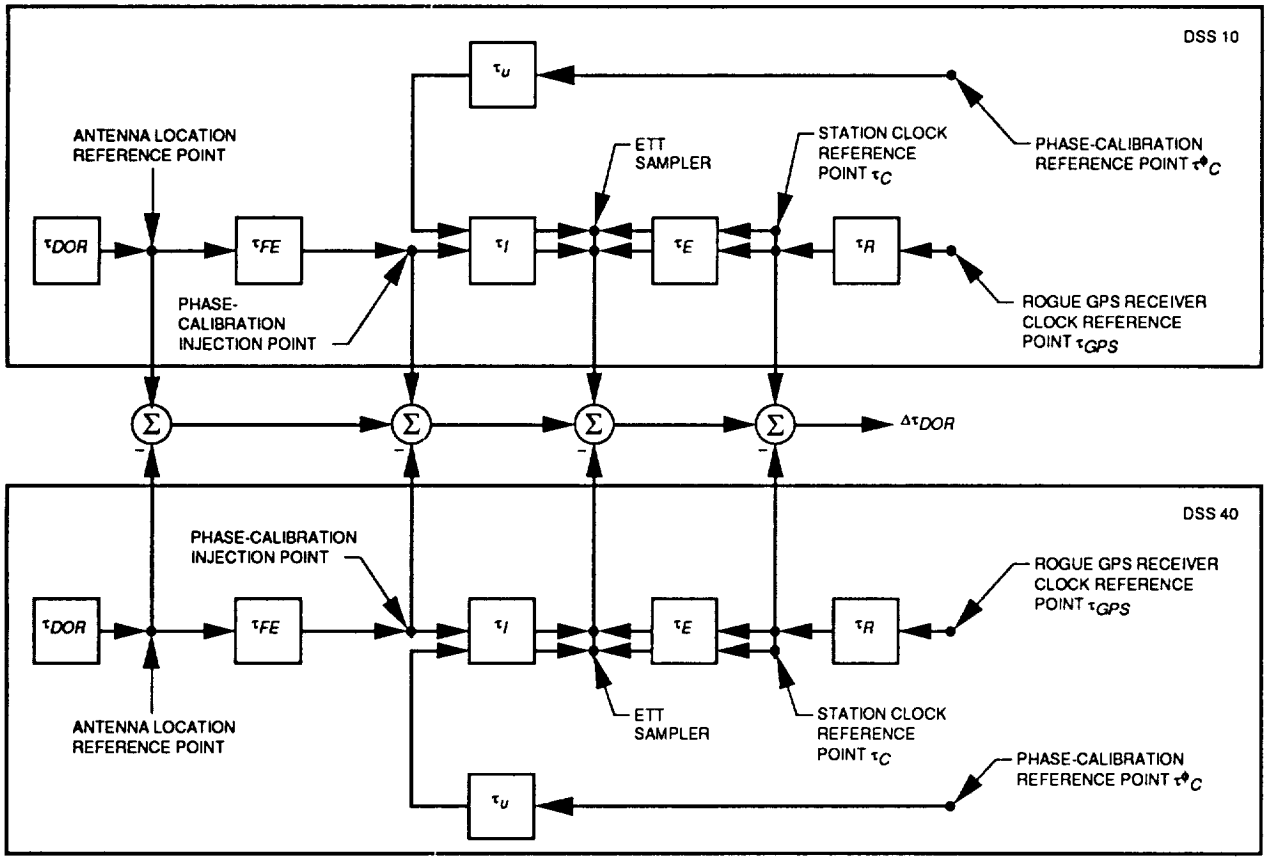


Fig. 4. Instrumental delay terms and timing offsets that contribute to the DOR delay observable of Eq. (3).

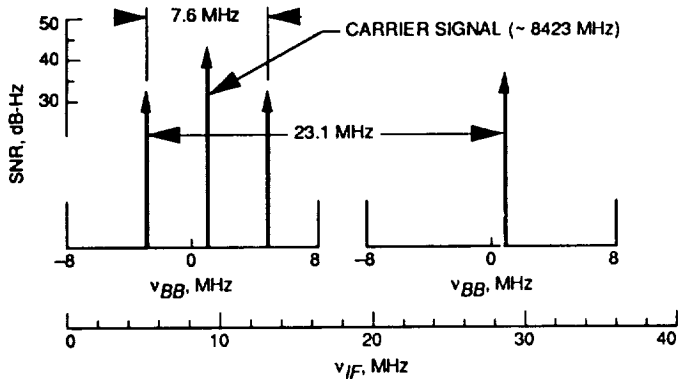


Fig. 5. Location of carrier signal and DOR tones recorded by the ETT during the Mars Observer DOR demonstration. The height of the arrows represents the relative signal-to-noise ratio of the tones.

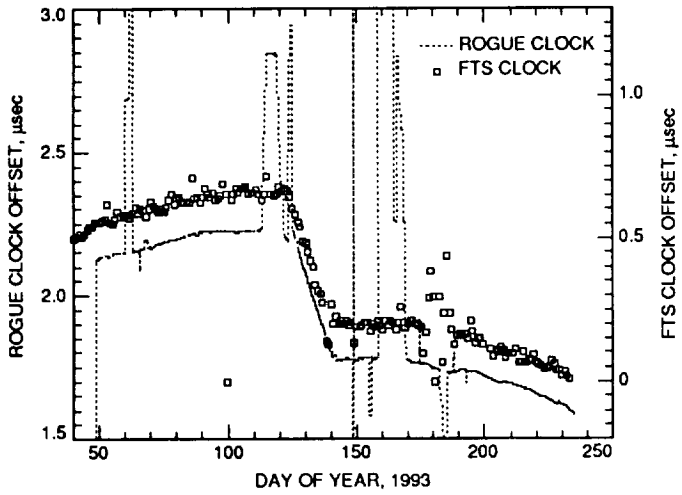


Fig. 6. Comparison of station clock offsets determined from the DSN's FTS system and those determined from the Rogue GPS receivers.

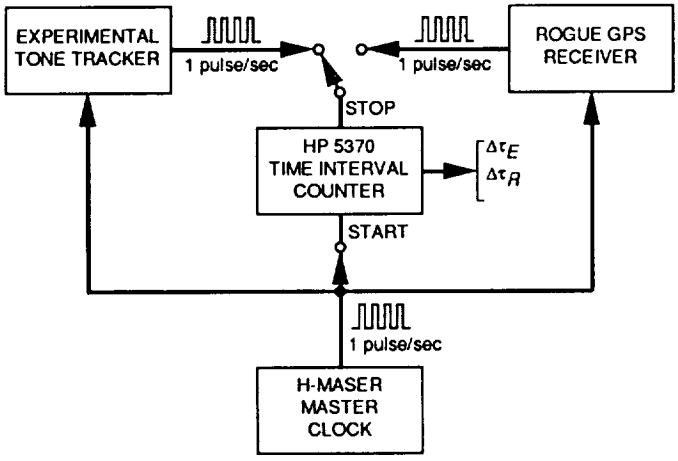


Fig. 7. Timing measurements performed at DSCC 10 and DSCC 40 to determine the time difference between the station master clock and the ETT and Rogue GPS receiver.

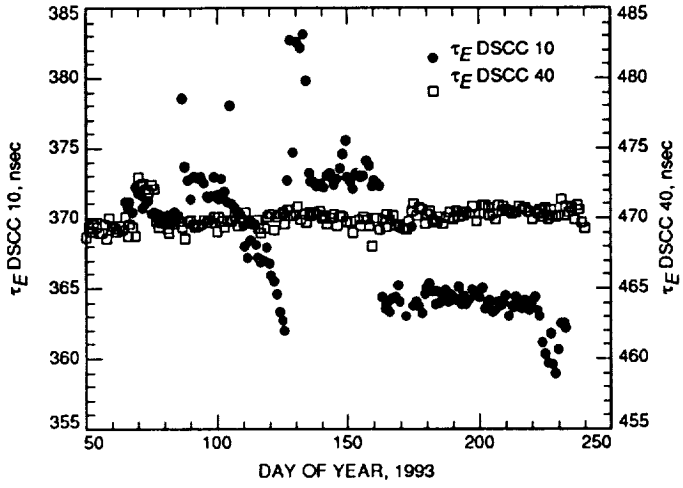


Fig. 8. Results of timing offset measurements for the τ_E delay of Eq. (3).

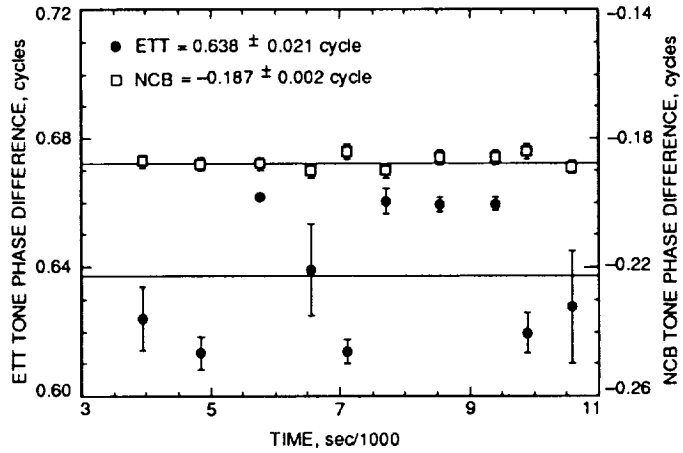


Fig. 9. Tone phase differences for the same pair of calibration tones recorded on May 19, 1993, at DSS 45 by the ETT and the DSN narrow-channel bandwidth VLBI system.

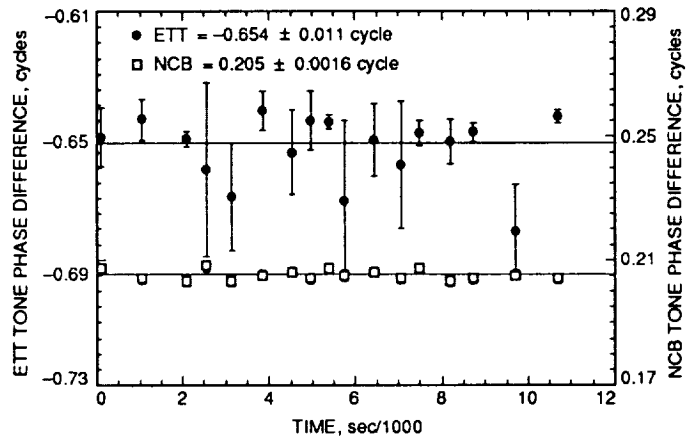


Fig. 10. Tone phase differences for the same pair of calibration tones recorded on March 24, 1993, at DSS 15 by the ETT and the DSN narrow-channel bandwidth VLBI system.

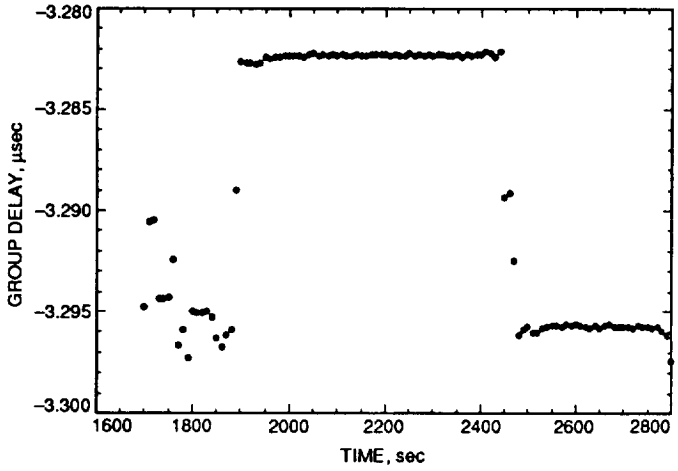


Fig. 11. Instrumental group delays computed from linear fits of tone phase versus frequency plots.

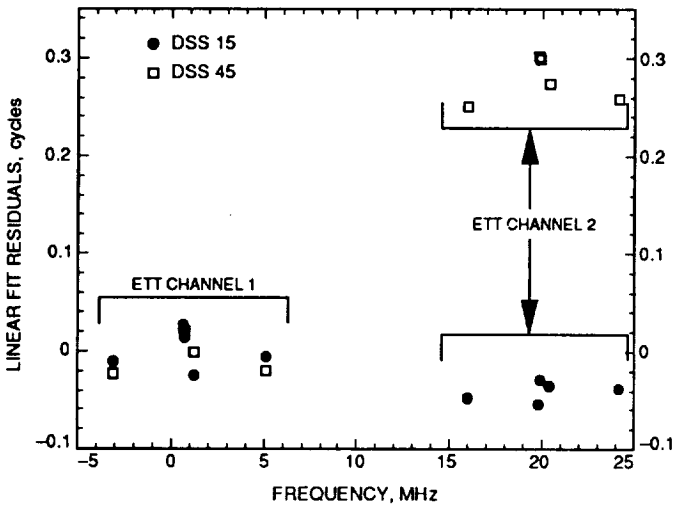


Fig. 12. Typical residuals to linear fit of phase calibration tone phases to frequency for both channels of the ETT.

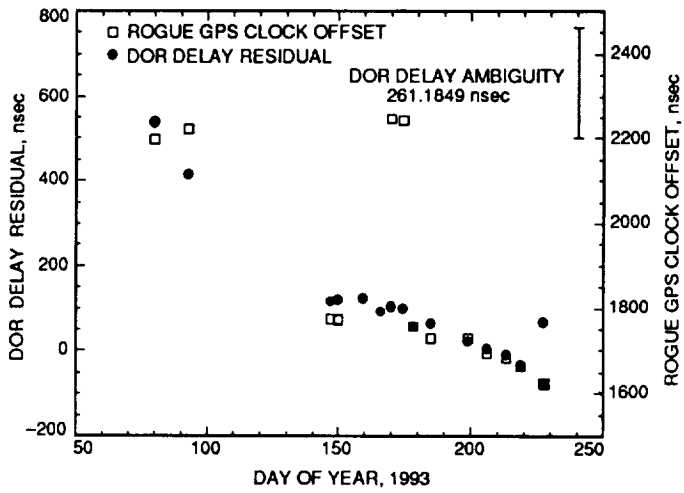


Fig. 13. Comparison of quasar-free DOR group delay residuals corresponding to the most widely separated spacecraft tone pair and station timing offsets estimated from the Rogue receiver GPS data.

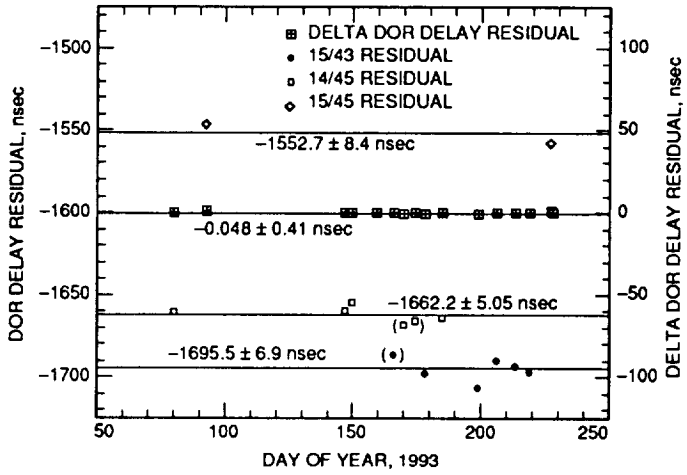


Fig. 14. Comparison of group delay residuals from the conventional ΔDOR measurements of Mars Observer and the group delay residuals from the quasar-free DOR measurements.

N94-35544

TDA Progress Report 42-117

May 15, 1994

Influence of an Externally Modulated Photonic Link on a Microwave Communications System

X. S. Yao and L. Maleki
Communications Systems Research Section

We analyze the influence of an externally modulated photonic link on the performance of a microwave communications system. From the analysis, we deduce limitations on the photocurrent, magnitude of the relaxation oscillation noise of the laser, third-order intercept point of the preamplifier, and other parameters in order for the photonic link to function according to the system specifications. Based on this, we outline a procedure for designing a photonic link that can be integrated in a system with minimal performance degradation.

I. Introduction

Photonic technology has become increasingly important in analog communications systems. For systems with high frequency and high dynamic range, externally modulated photonic links generally have better performance compared to directly modulated links [1,2]. The performance of such links has been analyzed by many authors [3–6]; however, in these analyses the links were assumed to be isolated from the microwave system and, therefore, their effect on the system was not adequately apparent. In addition, many parameters in these analyses were given for component engineers and, thus, are difficult to use for a system designer, who may not be familiar with the photonic technology. Finally, none of the analyses considered the influence of the laser's relaxation oscillation noise amplitude on the microwave system, which, as will be discussed later, may be critical in many applications.

We present here an analysis that emphasizes the integration of the link in an analog system. We pay special attention to the laser's relaxation oscillation noise and determine quantitatively its effect on the system. With parameters and equations intentionally written in system engineering terms, we hope that the results can be readily used by microwave engineers in their system designs.

An analog communications system can be considered as many subsystems that are cascaded together. Each subsystem i has a characteristic gain G_i, noise factor F_i, 1-dB compression P_i^{1dB}, third-order intercept point IP_i, and bandwidth Δf_i. Grouping the subsystems is somewhat arbitrary; for convenience, we group the system under consideration into three subsystems, as shown in Fig. 1. All the components before the optical link are included in subsystem 1, and all the components after the optical link are included in subsystem 3. The optical link itself is subsystem 2. For example, in an antenna remote system where the optical link is inserted between the low-noise amplifier (LNA) of the antenna and the downconverter, subsystem 1 is the LNA and subsystem 3 includes the downconverter

and all the components following the downconverter. With such a grouping, the effect of the optical link on the system's performance can easily be evaluated.

In this article, we first determine the gain, noise figure, 1-dB compression, and intercept point of an isolated optical link (or subsystem 2). We then determine quantitatively the changes in noise figure, dynamic range, and gain profile of the system caused by the insertion of the optical link. From the analysis, we deduce limitations on the magnitude of the relaxation oscillation peak, photocurrent, and other parameters of the optical link in order for the link to function according to the system's specifications. Finally, based on the analysis, we outline a procedure for designing a photonic link that can be integrated in a system with minimal performance degradation.

II. Parameters of an Isolated Optical Link

A. Gain of the Photonic Link

The photonic link consists of an electro-optic (E/O) modulator to convert the RF signal into an optical signal, a length of optical fiber to transmit the optical signal, and an optical receiver to convert the optical signal back to RF. Because two signal conversion processes are involved, the signal loss is generally high. The attenuation of the optical signal in the fiber also produces additional RF loss. The total RF gain (or loss) of the photonic link using the Mach–Zehnder modulator [5] is (see Appendix A)

$$G_{op} = \pi^2 \frac{I_{ph}^2 R_L}{V_\pi^2 / R_m} \tag{1}$$

where V_π is the half-wave voltage of the modulator, R_m is the input impedance of the modulator, R_L is the load impedance of the receiver, and I_{ph} is the average photocurrent in the load resistor of the receiver. In Eq. (1), the numerator is the electrical power generated by the photocurrent in the receiver, and the denominator is the input electrical power to the modulator corresponding to an applied voltage of V_π. The photocurrent in the receiver is related to the received optical power W_o by $I_{ph} = \eta W_o$, where η is the responsivity of the receiver.

B. The 1-dB Compression and Third-Order Intercept Point of the Optical Link

Referring to Appendix A, one can see that the input 1-dB compression $P_m^{1\text{dB}}$ and third-order intercept point IP_m of a Mach–Zehnder modulator are

$$IP_m = 10P_m^{1\text{dB}} = \frac{4}{\pi^2}\frac{V_\pi^2}{R_m} = \frac{4I_{ph}^2 R_L}{G_{op}} \tag{2}$$

They can be converted to the results obtained by Kolner and Dolfi in [7]. From Eq. (2), one can see that for a Mach–Zehnder modulator the third-order intercept point is always 10 times (or 10-dB) higher than the 1-dB compression. For a modulator with V_π of 8 V and R_m of 50 Ω, the 1-dB compression is 17 dBm, and the third-order intercept point is 27 dBm.

The output intercept IP_{op} is simply the product of the input intercept and the gain of the link. From Eq. (2), we can readily obtain

$$IP_{op} = 10P_{op}^{1\text{dB}} = 4I_{ph}^2 R_L \tag{3}$$

where $P_{op}^{1\text{dB}}$ is the output 1-dB compression of the link. It is important to notice that the output 1-dB compression and the third-order intercept are independent of the characteristics of the modulator and are proportional to the photo-electric power in the receiver.

C. Preamplifier Requirements

Typically, the loss of the basic optical link is −20 to −60 dB. To compensate for the signal loss, an amplifier with a gain of

$$G_{pr} = \frac{1}{G_{op}} \tag{4}$$

may be placed either before the modulator or after the receiver. However, a preamplifier is preferred because it also serves to reduce the noise figure of the optical link.

In order for the preamplifier not to limit the dynamic range of the photonic link, its output intercept IP_{pr} and 1-dB compression $P_{pr}^{1\text{dB}}$ must be much larger than the input intercept IP_m and 1-dB compression $P_m^{1\text{dB}}$ of the modulator, respectively. Using Eq. (2), we therefore obtain

$$IP_{pr} >> \frac{4}{\pi^2}\frac{V_\pi^2}{R_m} \tag{5}$$

$$P_{pr}^{1\text{dB}} >> \frac{2}{5\pi^2}\frac{V_\pi^2}{R_m} \tag{6}$$

D. Third-Order Intercept Point of the Loss-Compensated Optical Link

Using the cascading formula given by Norton [8], we obtain the output third-order intercept IP_2 of the loss-compensated photonic link (subsystem 2):

$$\frac{1}{IP_2} = \frac{1}{G_{op} IP_{pr}} + \frac{1}{IP_{op}} \tag{7}$$

Because the loss-compensated optical link has a gain of unity, $G_2 \equiv G_{pr} G_{op} = 1, IP_2$ is also the input intercept of the link. When Eq. (5) is satisfied, $IP_2 = IP_{op}$ and both the input and output intercept points of the loss-compensated link are equal to the output intercept point of the isolated optical link. Because $IP_{op} = 4I_{ph}^2 R_L$, again the intercept of the loss-compensated link is independent of the characteristics of the modulator and depends on only the photo-electric power in the receiver.

Similarly, the input and output 1-dB compression $P_2^{1\text{dB}}$ of the loss-compensated link is

$$\frac{1}{P_2^{1\text{dB}}} = \frac{1}{G_{op} P_{pr}^{1\text{dB}}} + \frac{1}{P_{op}^{1\text{dB}}} \tag{8}$$

E. Noise of the Optical Link

The total noise density (per hertz) at the output of the fiber optical link is

$$\mathcal{P}_{op} = G_{op} \mathcal{P}_m + \mathcal{P}_{thermal} + \mathcal{P}_{shot} + \mathcal{P}_{RIN} \tag{9}$$

where $\mathcal{P}_m$ is the input noise density to the modulator, $\mathcal{P}_{thermal}$ is the thermal noise density generated in the receiver, $\mathcal{P}_{shot}$ is the shot noise density, and $\mathcal{P}_{RIN}$ is the relative intensity noise density generated by the laser around the modulation frequency f_m. In Eq. (9), we neglected the dark current noise generated by the photodetector because the anticipated photocurrent (approximately 1 mA) is much larger than the dark current (approximately 1 nA). The thermal noise and the shot noise are white noise processes, and their expressions are well known, as follows:

$$\mathcal{P}_{thermal} = kT_{op} \tag{10}$$

$$\mathcal{P}_{shot} = 2eI_{ph} R_L \tag{11}$$

where T_{op} is the ambient temperature of the optical link, k is Boltzmann's constant, and e is the charge of the electron.

The relative intensity noise (RIN) of a YAG laser is frequency dependent and has a relaxation oscillation peak around a few hundred kHz [9]. However, this low frequency noise peak can be multiplied up to the modulation frequency by the modulator and contributes to the total relative intensity noise around f_m, as shown in Fig. 2(a). To find the relationship between the amplitudes of the multiplied RIN peak and the modulation, we performed a simple experiment, as shown in Fig. 2(b). We first added a small single-tone modulation to the laser light with a modulator (modulator 1) at a frequency (250 kHz) close to that of the relaxation oscillation peak (187 kHz). A second modulator (modulator 2) was then used to impose a strong modulation at a higher frequency. As one can see in Fig. 2(c), both the relaxation oscillation peak and the single tone were multiplied up and their relative amplitudes remained unchanged. This result indicates that the amplitude of the multiplied RIN peak can be calculated by treating the relaxation oscillation peak of the RIN as a single-tone modulation, as is done in Appendix B.

Consequently, the total RIN, $\mathcal{P}_{RIN}$, at the frequency of interest, f, is the sum of the baseband relaxation oscillation, $\mathcal{P}_{oRIN}$, at f and the multiplied relaxation oscillation peak, $\mathcal{P}_{mRIN}$:

$$\mathcal{P}_{RIN} = \mathcal{P}_{oRIN} + \mathcal{P}_{mRIN} \tag{12}$$

where

$$\mathcal{P}_{oRIN} = I_{ph}^2 R_L RIN\,(f_m) \tag{13}$$

$$\mathcal{P}_{mRIN} = \frac{I}{4} G_{op} P_m RIN\,(f - f_m) \tag{14}$$

In Eqs. (13) and (14), $RIN(f)$ is the laser RIN fluctuation at f and $RIN(f - f_m)$ is the RIN fluctuation around relaxation oscillation frequency f_{RLX}. They have a unit of 1/Hz. Note that $RIN(f - f_m)$ is a strong function of frequency around its peaks at $f - f_m = \pm f_{RLX}$, as shown in Fig. 2(a). From Eq. (14), one can see that the larger the driving signal, the more the low-frequency relaxation oscillation noise contributes to the noise at the modulation frequency, f_m.

From Eqs. (10), (11), and (13), one can see that the thermal noise is independent of photocurrent I_{ph}, the shot noise is proportional to I_{ph}, and the RIN is proportional to I_{ph} squared. Comparing $\mathcal{P}_{thermal}$, $\mathcal{P}_{shot}$, and $\mathcal{P}_{oRIN}$, one can see that at low photocurrent ($I_{ph} \leq 0.25$ mA), the

thermal noise $\mathcal{P}_{thermal}$ dominates. At moderate photocurrent (0.25 mA $\leq I_{ph} \leq$ 10 mA), the shot noise $\mathcal{P}_{shot}$ dominates, and at high photocurrent ($I_{ph} \geq$ 10 mA), the RIN $\mathcal{P}_{oRIN}$ dominates. In the calculation above, a relative intensity fluctuation $RIN(f)$ of -165 dB/Hz and $R_L = 50\ \Omega$ are assumed. Most laser manufacturers use this number to conservatively specify the RIN for diode-pumped YAG lasers at above 10 MHz; no definitive measurement has been performed so far to accurately determine it. It is generally believed that the actual value of the RIN for diode-pumped YAG lasers can be much smaller than -165 dB/Hz.

F. RIN Noise Peak

Since the multiplied relaxation-oscillation noise peaks are a few hundred kHz away from the modulation frequency, with their amplitudes increasing as the driving signal increases, it is very easy to mistake them for the signal [see Fig. 2(a)]. Thus, for a practical system, these peaks have to be suppressed below the system noise floor. The RIN fluctuation at relaxation oscillation frequency $RIN(f_{RLX})$ of a diode-pumped YAG laser without a noise reduction circuit is typically as high as -100 dB/Hz. Assuming that V_π of the modulator is 8 V, R_L and R_m are 50 Ω; then, with a moderate driving signal level of 1 mW (corresponding to a modulation depth of 1.54 percent) and a photocurrent of 0.02 mA, the multiplied relaxation oscillation peaks will be above the noise floor,set by the sum of the thermal, the shot, and the original RIN noise terms. The maximum multiplied RIN noise density at the peak, $f - f_m = \pm f_{RLX}$, can be obtained by replacing P_m in Eq. (14) with $P_m^{\max}$, the maximum allowed driving power of the system at the modulator:

$$\mathcal{P}_{mRIN}^{\max} = \frac{1}{4} G_{op} P_m^{\max} RIN\,(f_{RLX}) \tag{15}$$

If the input noise $\mathcal{P}_m$ to the optical link is small compared to the other noise terms, and the maximum driving power of the system at the modulator is $P_m^{\max} = P_m^{1\text{dB}}$, then in order for this multiplied RIN to be below the noise floor of the link, $RIN\,(f_{RLX})$ should satisfy

$$RIN\,(f_{RLX}) \leq \frac{10kT_{op}}{I_{ph}^2 R_L} + \frac{20e}{I_{ph}} + 10RIN\,(f_m) \tag{16}$$

However, in an actual system, the input noise $\mathcal{P}_m$ to the optical link [the first noise term in Eq. (9)] is much larger than the rest of the noise terms and sets the noise floor of the system. In order for the multiplied RIN to be 10 dB lower than the input noise floor, we must have

$$RIN\,(f_{RLX}) \leq \frac{0.4}{D_{sys}} \tag{17}$$

where $D_{sys} = P_m^{\max}/\mathcal{P}_m$ is the maximum signal-to-noise ratio of the system, or simply the dynamic range. For example, if the dynamic range of the system is 131 dB-Hz, the RIN peak must be smaller than -135 dB/Hz. Diode-pumped YAG lasers with relaxation oscillation peaks at this level may now be obtained commercially. These lasers reduce the RIN peak amplitude by incorporating a feedback loop in the pump diode circuit [9]. For systems requiring higher dynamic range, further noise reduction with an external circuit [10] may be necessary.

G. Noise Factor of the Loss-Compensated Optical Link

The noise factor of the basic optical link may be expressed as

$$F_{op} = \frac{\mathcal{P}_{op}}{G_{op}\mathcal{P}_m} = 1 + \frac{F_{th} + F_{shot} + F_{oRIN} + F_{mRIN}}{G_{op}} \tag{18}$$

where F_{th}, F_{shot}, F_{oRIN}, and F_{mRIN} are the noise factor contributions from the thermal, shot, RIN, and the multiplied RIN noise, respectively. They are given by the following expressions:

$$F_{th} = \frac{\mathcal{P}_{thermal}}{\mathcal{P}_m} = 1 \tag{19a}$$

$$F_{shot} = \frac{\mathcal{P}_{shot}}{\mathcal{P}_m} = \frac{2eI_{ph}R_L}{kT_{op}} \tag{19b}$$

$$F_{oRIN} = \frac{\mathcal{P}_{oRIN}}{\mathcal{P}_m} = \frac{I_{ph}^2 R_L}{kT_{op}} RIN(f_m) \tag{19c}$$

$$F_{mRIN}(f) = \frac{\mathcal{P}_{mRIN}}{\mathcal{P}_m} = \frac{G_{op}}{4kT_{op}} P_m RIN(f - f_m) \tag{19d}$$

where $\mathcal{P}_m = kT_{op}$ is used throughout. Because G_{op} is small (approximately $10^{-2} - 10^{-6}$), the resulting noise factor, F_{op}, is large; thus, to reduce the noise factor, a preamplifier is required. With the preamplifier gain of G_{pr} and

noise factor of F_{pr}, the noise factor of the loss-compensated optical link is

$$F_2(f) = F_{pr} + \frac{F_{op} - 1}{G_{pr}} = F_{2o} + F_{mRIN}(f) \qquad (20)$$

where

$$F_{2o} = F_{pr} + F_{th} + F_{shot} + F_{oRIN} \qquad (21)$$

is the frequency-independent part of the noise factor and $F_{mRIN}(f)$ is the frequency-dependent part.

H. Compression Dynamic Range of the Optical Link

The dynamic range of a system is defined as the maximum output signal power divided by the total output noise of the system: $D_{op} = P_{op}^{1dB}/\mathcal{P}_{op}$, or

$$\frac{1}{D_{op}} = \frac{1}{D_m} + \frac{1}{D_{thermal}} + \frac{1}{D_{shot}} + \frac{1}{D_{oRIN}} + \frac{1}{D_{mRIN}} \qquad (22)$$

where

$$D_m = \frac{P_{op}^{1dB}}{G_{op}\mathcal{P}_m} = \frac{P_m^{1dB}}{\mathcal{P}_m} \qquad (23a)$$

$$D_{th} = \frac{P_{op}^{1dB}}{\mathcal{P}_{thermal}} = \frac{2}{5}\frac{I_{ph}^2 R_L}{kT_{op}} \qquad (23b)$$

$$D_{shot} = \frac{P_{op}^{1dB}}{\mathcal{P}_{shot}} = \frac{I_{ph}}{5e} \qquad (23c)$$

$$D_{oRIN} = \frac{P_{op}^{1dB}}{\mathcal{P}_{oRIN}} = \frac{2}{5RIN(f_m)} \qquad (23d)$$

$$D_{mRIN} = \frac{P_{op}^{1dB}}{\mathcal{P}_{mRIN}^{max}} = \frac{4}{RIN(f_{RLX})} \qquad (23e)$$

Equations (9) through (14) are used in deriving these equations. To express in dB-Hz, the above equations can be rewritten as

$$D_{th} = 157 + 20 \log I_{ph}(\text{mA}) \qquad (24a)$$

$$D_{shot} = 151 + 10 \log I_{ph}(\text{mA}) \qquad (24b)$$

$$D_{oRIN} = -4 - 10 \log RIN(f_m) \qquad (24c)$$

$$D_{mRIN} = 6 - 10 \log RIN(f_{FLX}) \qquad (24d)$$

In the above calculations, we assume that $T_{op} = 290$ K and $R_L = 50\ \Omega$.

I. Spur-Free Dynamic Range of an Optical Link

The spur-free dynamic range SFD_{op} [in units of $(\text{Hz})^{2/3}$] of the optical link is

$$SFD_{op} = \left(\frac{IP_{op}}{\mathcal{P}_{op}}\right)^{2/3} \qquad (25)$$

where IP_{op} is the output third-order intercept of the modulator and is defined in Eq. (3). Since $IP_{op} = 10P_{op}^{1dB}$, we have the following simple relation between the compression dynamic range and spur-free dynamic range of an optical link:

$$SFD_{op} = (10D_{op})^{2/3} \text{ in units of } (\text{Hz})^{2/3} \qquad (26)$$

or

$$SFD_{op} = 6.7 + \frac{2}{3}D_{op} \text{ in units of dB-}(\text{Hz})^{2/3} \qquad (27)$$

III. Influence of the Optical Link on the System

The ultimate optical link for antenna remoting does not degrade the performance of the existing system and will remain essentially "transparent" to the system. Thus, the insertion of such an optical link in the existing system does not change the system's noise temperature, dynamic range, gain profile, and phase noise. Such an optical link is the basis of our analysis below, where the influence of various parameters is considered in order to determine the effect of each component of the link separately and to specify the required parameters to achieve "transparent" operation upon insertion in the system.

A. Influence on the Noise Factor

The noise factor of the system without the optical link is

$$F_{sys} = F_1 + \frac{T_3}{T_1}\frac{F_3 - 1}{G_1} \tag{28}$$

where G_1, F_1, and T_1 are the gain, noise factor, and input noise temperature of subsystem 1, and where F_3 and T_3 are the noise factor and input noise temperature of subsystem 3.

With the optical link inserted, the total noise factor is

$$F'_{sys} = F_1 + \frac{F_2 - 1}{G_1}\frac{T_2}{T_1} + \frac{F_3 - 1}{G_1 G_2}\frac{T_3}{T_1} \tag{29}$$

Because $G_2 = 1$ for a loss-compensated optical link, the total noise factor of the system is, therefore,

$$F'_{sys} = F_{sys} + \Delta F_2 \tag{30}$$

where

$$\Delta F_2 \equiv \Delta F_{2o} + \Delta F_{mRIN} \tag{31}$$

is the total noise factor increase caused by the insertion of the optical link. In Eq. (31), $\Delta F_{2o} \equiv (T_2/T_1) \times (F_{2o} - 1)/G_1$ is the frequency-independent part of the noise factor increase caused by the preamplifier, thermal, shot, and baseband RIN. It can be expressed as

$$\Delta F_{2o} = \frac{T_2 F_{pr} + 2eI_{ph}R_L/k + I_{ph}^2 R_L RIN(f_m)/k}{G_1 T_1} \tag{32}$$

where $T_{op} = T_2$ is used. Taking $R_L = 50\ \Omega$, $RIN(f_m) = -165$ dB/Hz, and I_{ph} in units of mA, Eq. (32) becomes

$$\Delta F_{2o} \approx \frac{T_2 F_{pr} + 1160 I_{ph} + 116 I_{ph}^2}{G_1 T_1} \tag{33}$$

On the other hand, ΔF_{mRIN} is the frequency-dependent part of the noise factor and is induced by the multiplied RIN noise peak. It is defined as

$$\Delta F_{mRIN} \equiv \frac{1}{4}\frac{P_m/G_1 G_{pr}}{kT_1} RIN(f - f_m) \tag{34}$$

In Eq. (34), $P_m/G_1 G_{pr}$ is the input signal power of the system, and kT_1 is the input noise density to the system. The ratio of the two terms is just the input signal-to-noise ratio (SNR) in a 1-Hz bandwidth. Note that $P_m/G_1 G_{pr}$ can be either the signal of interest or an interference signal that falls into the pass band of the signal of interest. When there is no interference signal present, this term is not critical because, when the signal of interest is small, this term is small. Although this term becomes very large at high signal levels, the resulting noise factor increase is not damaging because the corresponding SNR is so high that a small noise increase will have little effect on the performance of the system.

On the other hand, when there is a strong interference signal present, $P_m/G_1 G_{pr}$ will be large, even though the signal of interest is small. Consequently, the SNR of the signal of interest is reduced, and the performance of the system is degraded.

The largest ΔF_{mRIN} occurs when $f - f_m = \pm f_{RLX}$ and when the SNR of the system is a maximum:

$$\Delta F_{mRIN}^{\max} = \frac{1}{4} F_{sys} D_{sys} RIN(f_{RLX}) \tag{35}$$

Here, F_{sys} is the noise figure of the system, and D_{sys} is the dynamic range of the system or the maximum SNR.

B. Influence on the Dynamic Range of the System

Without the optical link, the original input third-order intercept point IP_{sys} of the system is

$$\frac{1}{IP_{sys}} = \frac{1}{IP_1} + \frac{G_1}{IP_3} \tag{36}$$

where IP_3 is the input intercept point of subsystem 3.

After the insertion of the optical link, the third-order intercept point of the whole system is

$$\frac{1}{IP'_{sys}} = \frac{1}{IP_1} + \frac{G_1}{IP_2} + \frac{G_1 G_2}{IP_3} \tag{37}$$

where IP_2 is the input intercept point of the optical link and is given by Eq. (7). For the loss-compensated optical link, $G_2 \equiv G_{pr} G_{op} = 1$, Eq. (37) becomes

$$IP'_{sys} = \frac{IP_{sys}}{1 + G_1 IP_{sys}/IP_2} \tag{38}$$

The reduction of the system's intercept point in dB is, therefore,

$$\Delta(IP) = 10 \log \left(1 + \frac{G_1 IP_{sys}}{IP_2}\right) \tag{39}$$

Similarly, the 1-dB compression, P'_{sys}, of the system after the insertion of the optical link is

$$P'_{sys} = \frac{IP_{sys}^{1dB}}{1 + G_1 P_{sys}^{1dB}/P_2^{1dB}} \tag{40}$$

and the reduction of the system's 1-dB compression in dB is

$$\Delta\left(P^{1dB}\right) = 10 \log \left(1 + \frac{G_1 P_{sys}^{1dB}}{P_2^{1dB}}\right) \tag{41}$$

Finally, the spur-free dynamic range and compression dynamic range of the system after the insertion of the optical link can be expressed as

$$SFD'_{sys} = \left(\frac{IP'_{sys}}{F'_{sys} kT_1}\right)^{2/3}$$

$$= \frac{SFD_{sys}}{\left[(1 + \Delta F_2/F_{sys})(1 + G_1 IP_{sys}/IP_2)\right]^{2/3}} \tag{42}$$

$$D'_{sys} = \frac{P'_{sys}}{F'_{sys} kT_1} = \frac{D_{sys}}{(1 + \Delta F_2/F_{sys})\left(1 + G_1 P_{sys}^{1dB}/P_2^{1dB}\right)} \tag{43}$$

where Eqs. (30), (38), and (40) are used. The degradation of the spur-free dynamic range (in dB-Hz$^{2/3}$) and the compression dynamic range (in dB-Hz) of the system are, therefore,

$$\Delta\left(SFD_{sys}\right) \approx \left(\frac{2}{3}\right)\Delta\left(IP\right) \approx 6.7 \log \left(1 + \frac{G_1 IP_{sys}}{IP_2}\right) \tag{44}$$

$$\Delta D_{sys} \approx \Delta\left(P^{1dB}\right) = 10 \log \left(1 + \frac{G_1 P_{sys}^{1dB}}{P_2^{1dB}}\right) \tag{45}$$

where $\Delta F_{op}/F_{sys} << 1$ is assumed.

When the preamplifier is properly chosen so that Eq. (5) is satisfied, then from Eqs. (7) and (8) we have $IP_2 = 4I_{ph}^2 R_L$ and $P_2^{1dB} = 2I_{ph}^2 R_L/5$. Consequently, the dynamic range of the system is solely limited by the photo-electric power. The higher the photo-electric power, the smaller the degradation. On the other hand, if the requirement of Eq. (5) cannot be satisfied, then both the preamplifier and the photo-electric power limit the system's dynamic range.

C. Photocurrent Requirement

If the preamplifier is properly chosen so that Eq. (5) is satisfied, then from Eq. (44), in order for the degradation $\Delta\left(SFD_{sys}\right)$ of the spur-free dynamic range to be less than 1 dB, the photocurrent must be

$$I_{ph} \geq \sqrt{\frac{5G_1 IP_{sys}}{8R_L}} \tag{46}$$

Similarly, from Eq. (45), in order for the degradation, ΔD_{sys}, of the compression dynamic range to be less than 3 dB, the photocurrent must be

$$I_{ph} \geq \sqrt{\frac{5G_1 P_{sys}^{1dB}}{2R_L}} \tag{47}$$

From Eqs. (46) and (47), one can see that the higher the input 1-dB compression of the system, the higher the optical power (photocurrent) of the optical link is required to be to preserve the dynamic range of the system.

IV. Summary

We analyzed the influence of an externally modulated fiber-optic link on a microwave communications system and determined quantitatively the degradation of the noise figure and the dynamic range caused by inserting the link in the system. We found that if the preamplifier is properly chosen, the photo-electric power in the photodetector is the only parameter of the link that affects the dynamic range of the system. The higher the photo-electric power, the less dynamic range degradation of the system. For a system of a given dynamic range, we deduced the minimum photocurrent (optical power) requirement for the system.

We also determined quantitatively the effect of different noise terms of the optical link on the noise figure of the system. We paid special attention to the laser's relaxation

oscillation noise and showed how it is multiplied up in frequency by the modulation signal to degrade the signal. The maximum amplitude of the relaxation oscillation peak is inversely proportional to the achievable dynamic range of the system.

Finally, we studied the requirements of the preamplifier required for the optical link. Both the required gain and the third-order intercept of the preamplifier are proportional to V_π^2/R_m, a quantity solely determined by the characteristics of the modulator. In practice, modulators with small V_π and large R_m should be used so that the preamplifier can meet the requirements of Eqs. (4) and (5).

When designing a photonic link for analog communications systems, the following procedure is recommended. First, a laser with low relaxation oscillation noise should be chosen according to Eq. (17). This ensures that the multiplied noise peaks are well below the noise floor of the system. Second, the minimum photocurrent I_{ph}^{min} (or minimum optical power) of the photonic link should be determined using Eq. (46) or Eq. (47). This is to make certain that the dynamic range of the system is preserved. Third, a modulator with low enough V_π should be chosen, and the RF loss of the photonic link should be evaluated by substituting V_π and I_{ph}^{min} in Eq. (1). Fourth, a preamplifier should be selected with a gain large enough to compensate for the link's RF loss. To ensure that the preamplifier does not limit the dynamic range of the system, its intercept point and 1-dB compression should satisfy Eqs. (7) and (8), respectively. Finally, the noise factor degradation of the system caused by the insertion of the photonic link should be evaluated using Eqs. (33) and (34).

Acknowledgments

We thank G. Lutes and R. Logan for sharing their technical expertise.

References

[1] G. E. Betts, L. M. Johnson, C. H. Cox III, and S. D. Lowney, "High-Performance Optical Analog Link Using External Modulator," *IEEE Photonics Technology Letters*, vol. 1, no. 11, pp. 404–406, November 1989.

[2] R. T. Logan and G. F. Lutes, "High Stability Microwave Fiber Optic Systems: Demonstration and Applications," *Proceedings of the 46th Annual Symposium on Frequency Control*, Hershey, Pennsylvania, pp. 310–316, May 27–29, 1992.

[3] W. E. Stephens and T. R. Joseph, "System Characteristics of Direct Modulated and Externally Modulated RF Fiber-Optic Links," *Journal of Lightwave Technology*, vol. LT-5, no. 3, pp. 380–387, March 1987.

[4] W. L. Glomb, "Fiber Optic Links for Antenna Remoting," *SPIE Proceedings*, vol. 1703, pp. 523–527, 1992.

[5] W. L. Glomb, J. D. Farina, and S. W. Merritt, "Minimum Noise Figure Microwave Transmission System," *SPIE Proceedings*, vol. 1703, pp. 528–534, 1992.

[6] R. T. Logan, G. F. Lutes, and L. Maleki, "Microwave Analog Fiber Optic Link for Use in the NASA/JPL Deep Space Network," *The Telecommunications and Data Acquisition Progress Report 42-100, vol. October–December 1989*, Jet Propulsion Laboratory, Pasadena, California, pp. 21–33, February 15, 1990.

[7] B. H. Kolner and D. W. Dolfi, "Intermodulation Distortion and Compression in an Integrated Electrooptic Modulator," *Applied Optics*, vol. 26, no. 17, pp. 3676–3680, September 1987.

[8] D. E. Norton, "The Cascading of High Dynamic Range Amplifiers," *Microwave Journal*, vol. 16, no. 6, pp. 57–71, June 1973.

[9] T. J. Kane, "Intensity Noise in Diode-Pumped Single-Frequency Nd:YAG Lasers and Its Control by Electronic Feedback," *IEEE Photonics Technology Letters*, vol. 2, no. 4, pp. 244–245, April 1990.

[10] D. M. Dagenais, K. P. Koo, and D. Dandridge, "Demonstration of Low-Frequency Intensity Noise Reduction for Fiber Sensors Powered by Diode-Pumped Nd-YAG Lasers," *IEEE Photonics Technology Letters*, vol. 4, no. 5, pp. 518–520, May 1992.

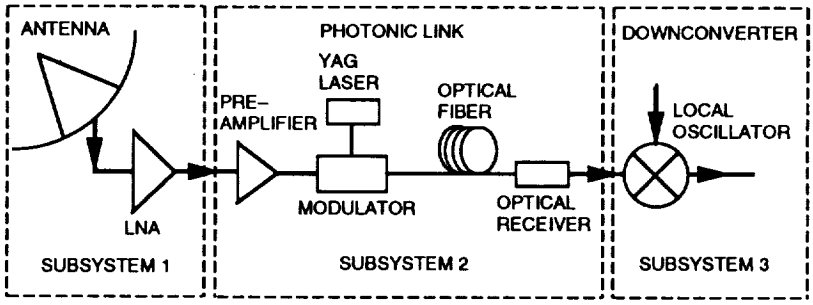


Fig. 1. An illustration of a photonic link in an analog system and how to group the system into three subsystems.

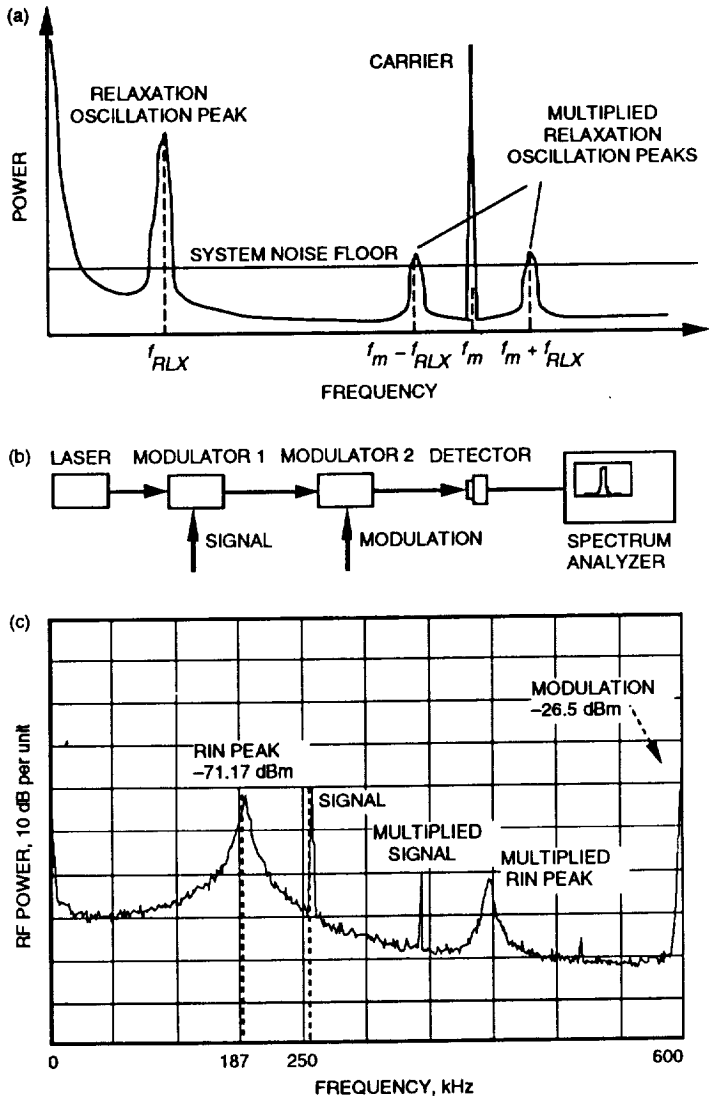


Fig. 2. The influence of multiplied relaxation oscillation peaks on the system: (a) the laser relaxation oscillation peak is multiplied up by the modulation signal. The multiplied peaks may be mistaken for the signal when they are above the system noise floor; (b) diagram of an experiment to verify that the strength of the multiplied noise peak obeys the same rule as a multiplied signal; and (c) results of the experiment.

Appendix A

Derivation of Gain, 1-dB Compression, and Intercept Point of an Optical Link

The optical transmission function of a Mach–Zehnder modulator biased at 50 percent of its transmission peak is

$$T(t) = \frac{1}{2}\left[1 - \sin\frac{\pi V(t)}{V_\pi}\right] \tag{A-1}$$

where V_π is the half-wave voltage of the modulator and $V(t)$ is the voltage of the driving signal. For $|\pi V(t)/V_\pi| < 1$, we can expand Eq. (A-1) into a Taylor series:

$$T(t) \approx \frac{1}{2}\left\{1 - \frac{\pi V(t)}{V_\pi} + \frac{1}{6}\left[\frac{\pi V(t)}{V_\pi}\right]^3\right\} \tag{A-2}$$

The error of the approximation is less than 0.83 percent.

A. Gain of the Optical Link

For a single-tone driving signal $V(t) = V_o \sin \omega t$, Eq. (A-2) becomes

$$T(t) = \frac{1}{2}\left[1 - m\left(1 - \frac{m^2}{8}\right)\sin \omega t - \frac{m^3}{24}\sin 3\omega t\right] \tag{A-3}$$

where m is the modulation depth and is defined as

$$m \equiv \frac{\pi V_o}{V_\pi} \tag{A-4}$$

The total photocurrent current $I(t)$ in the load resistor of the photoreceiver is

$$I(t) = \alpha\eta W_{in} T(t) = I_{ph} T(t) \tag{A-5}$$

where $I_{ph} = \alpha\eta W_{in}$ is the total average photocurrent in the load resistor, W_{in} is the input optical power to the modulator, η is the responsivity of the photoreceiver, and α is the total optical loss of the optical link, including modulator insertion loss, fiber attenuation loss, and optical coupling loss.

In the frequency domain, the optical link's output power $P_{op}(\omega)$ of the fundamental frequency component is

$$P_{op}(\omega) = I^2(\omega)\frac{R_L}{2} = \frac{1}{2} I_{ph}^2 R_L m^2 \left(1 - \frac{m^2}{8}\right)^2 \tag{A-6}$$

On the other hand, the total input RF power $P_m(\omega)$ to the modulator is

$$P_m(\omega) = \frac{V_o^2}{2R_m} = \frac{m^2 V_\pi^2}{2\pi^2 R_m} \tag{A-7}$$

where R_m is the input impedance of the modulator. Eq. (A-4) is used in deriving Eq. (A-7).

Substituting Eq. (A-7) in Eq. (A-6), we obtain

$$P_{op} = G_{op}\left(1 - \frac{G_{op} P_m}{4 I_{ph}^2 R_L}\right)^2 P_m \tag{A-8}$$

where G_{op} is the small signal gain of the optical link and is defined as

$$G_{op} \equiv \frac{P_{op}}{P_m} = \pi^2 \frac{I_{ph}^2 R_L}{V_\pi^2 / R_m} \tag{A-9}$$

B. The 1-dB Compression of the Optical Link

From Eq. (A-8), 1-dB compression occurs when $(1 - G_{op} P_m / 4 I_{ph}^2 R_L)^2 = 0.8$. Therefore, the input 1-dB compression of the optical link is

$$P_m^{1\text{dB}} \approx \frac{2 I_{ph}^2 R_L}{5 G_{op}} \tag{A-10}$$

C. Third-Order Intercept of the Optical Link

For a two-tone driving signal of equal amplitude $V(t) = V_o(\sin \omega_1 t + \sin \omega_2 t)$, Eq. (A-2) becomes

$$T(t) \approx \frac{1}{2}\left\{\left[1 - m\left(1 - \frac{3m^2}{8}\right)(\sin \omega_1 t + \sin \omega_2 t)\right]\right.$$

$$\left. + \frac{m^3}{8}\left[\sin\left(\omega_1 - \Delta\omega\right) + \sin\left(\omega_2 + \Delta\omega\right)\right]\right\} \quad \text{(A-11)}$$

where $\Delta\omega = \omega_2 - \omega_1$. In Eq. (A-11), the second term is the intermodulation product and other higher harmonic terms were neglected. Similar to the derivation of Eq. (A-6), the RF power of the output intermodulation product is

$$P_{IM}\left(\omega_1 - \Delta\omega\right) = P_{IM}\left(\omega_2 + \Delta\omega\right) = \frac{I_{ph}^2 R_L m^6}{128} \quad \text{(A-12)}$$

Substitution of Eqs. (A-7) and (A-9) in Eq. (A-12) yields the output intermodulation products

$$P_{IM} = \frac{\left(G_{op} P_m\right)^3}{\left(4 I_{ph}^2 R_L\right)^2} \quad \text{(A-13)}$$

At the third-order intercept point, $P_{IM} = G_{op} P_m$. Substituting in Eq. (A-13), we obtain

$$IP_m = P_m = \frac{4 I_{ph}^2 R_L}{G_{op}} \quad \text{(A-14)}$$

Appendix B

Multiplied RIN Noise

The total optical power incident on the photoreceiver is

$$\Delta P(t) = [P_o + \Delta P(t)]\,(1 + m \sin \omega_m t) \tag{B-1}$$

where $\Delta P(t)$ is the optical power fluctuation, ω_m is the modulation frequency, and m is the modulation depth defined in Eq. (A-4). Because the optical fluctuation of a laser peaks at relaxation oscillation frequency ω_R, as an approximation, $\Delta P(t)$ can be written as

$$\Delta P(t) = \Delta P_o(t) \sin \omega_R t \tag{B-2}$$

Substituting Eq. (B-2) in Eq. (B-1), we obtain

$$P(t) = P_o\,(1 + m \sin \omega_m t) + \Delta P_o(t) \sin \omega_R t$$

$$- \frac{m \Delta P_o(t)}{2} [\cos\,(\omega_m - \omega_R) - \cos\,(\omega_m + \omega_R)\,t] \tag{B-3}$$

The photocurrent at $\omega_m - \omega_R$ and $\omega_m + \omega_R$ is

$$I_{mRIN} = \frac{\eta \Delta P_o(t) m}{2} \tag{B-4}$$

The corresponding rms noise power at $\omega_m - \omega_R$ and $\omega_m + \omega_R$ is

$$P_{mRIN} = \frac{I_{mRIN}^2 R_L}{2} = \frac{1}{2}\left(\frac{\eta m}{2}\right)^2 R_L < \Delta P_0(t)^2 > \tag{B-5}$$

where <> denotes for time average.

Because $RIN = < \Delta P_0(t)^2 > / P_o^2$, the last equation becomes

$$P_{mRIN} = \frac{1}{2}\left(\frac{I_{ph} m}{2}\right)^2 R_L RIN \tag{B-6}$$

Substitution of Eqs. (A-7) and (A-9) in Eq. (B-5) yields

$$P_{mRIN} = \frac{1}{4} G_{op} P_m RIN \tag{B-7}$$

Field Demonstration of X-Band Photonic Antenna Remoting in the Deep Space Network

X. S. Yao, G. Lutes, R. T. Logan, Jr., and L. Maleki
Communications Systems Research Section

We designed a photonic link for antenna remoting based on our integrated system analysis. With this 12-km link, we successfully demonstrated photonic antenna-remoting capability at X-band (8.4 GHz) at one of NASA's Deep Space Stations while tracking the Magellan spacecraft.

I. Introduction

Photonic antenna remoting has become increasingly important in high-stability microwave and millimeter-wave communications systems. In an antenna remoting system, a photonic link is generally inserted between the low-noise amplifier (LNA), at the antenna front-end, and the downconverter, as shown in Fig. 1. One of the major advantages of such a link is that the downconverter and related equipment may be moved out of the antenna area to a remote command and signal processing center to substantially reduce the amount of equipment located at the antenna area. In systems with multiple and widely separated antennas, this new configuration will lower hardware and operating costs and increase system performance, flexibility, and reliability. In addition, this approach offers the capability to coherently array widely separated antennas at RF frequencies to increase receiving sensitivity.

For systems with high-frequency and high-dynamic ranges, externally modulated photonic links generally are preferred over the directly modulated links [1,2]. The performance of such links previously has been analyzed [3-6] and demonstrated in the laboratory environment [1,2,7]. However, most of the previous analyses were performed by considering the photonics link as an isolated entity rather than as a segment of the antenna-receiver system. In addition, in none of the previous studies was account taken of the influence of the laser's relaxation oscillation noise on the system performance, an effect that is critical in a number of antenna remoting applications.

In the present work, we present results of a study [8] to quantitatively determine the influence of a photonic link on the entire system and to explicitly account for the influence of the laser's relaxation oscillation noise on the system performance. From the analysis, we deduce the requirement for each component of the link, according to system specifications. Finally, we report on the successful demonstration of a photonic link capable of directly transmitting X-band (8.4-GHz) microwave signals from an antenna site to a remote signal processing facility located 12 km away in NASA's Venus Deep Space Station (DSS) 13. This demonstration is critical for implementing photonic antenna remoting throughout NASA's Deep Space Network [2,6].

The externally modulated photonic link essentially consists of a diode-pumped YAG laser, a high-speed $LiNbO_3$ modulator, an RF preamplifier, and a high-speed optical receiver. For ultimate performance in the antenna remoting applications, the insertion of the link between the low-

noise amplifier and the receiver downconverter should be transparent to the rest of the system, and thus leave the system's noise temperature, dynamic range, and gain profile unaltered. This requirement forms the basis of our analysis of the link performance.

II. The Preamplifier Requirements

The RF insertion loss G_{op} (or gain) of a photonic link without a preamplifier is given by [5,8]

$$G_{op} = \pi^2 \frac{I_{ph}^2 R_L}{V_\pi^2 / R_m} \tag{1}$$

where I_{ph} is the photocurrent across the load resistor R_L of the receiver, R_m is the input impedance of the modulator, and V_π is the half-wave voltage of the modulator. In Eq. (1), the numerator is the electrical power generated by the photocurrent in the receiver, and the denominator is the input electrical power to the modulator with a modulation voltage of V_π. We, therefore, arrive at the requirement for a preamplifier with a gain of $G_{pr} = 1/G_{op}$ in order to compensate for any loss resulting from the insertion of the photonic link.

The input third-order intercept of the modulator IP_m can be expressed as [5,8,9]

$$IP_m = 10P_m^{1\text{dB}} = \frac{4}{\pi^2} \frac{V_\pi^2}{R_m} \tag{2}$$

where $P_m^{1\text{dB}}$ is the input 1-dB compression of the modulator. To ensure that the preamplifier does not limit the dynamic range of the link, the output third-order intercept IP_{pr} of the preamplifier must be much larger than IP_m, thus leading to the condition that

$$IP_{pr} >> \frac{4}{\pi^2} \frac{V_\pi^2}{R_m} \tag{3}$$

III. Impact of the Laser's Relaxation Oscillation Peak

Any diode-pumped YAG laser has a relaxation oscillation noise peak around a few hundred kHz [10]. However, this low frequency relative intensity noise (RIN) peak can be multiplied up to the modulation frequency by the modulator, as shown in Fig. 2. In systems where the frequency of the signal of interest is less than a few hundred kHz from the carrier frequency, these multiplied noise peaks interfere with the signal and may cause serious difficulties. To avoid this, the amplitude of the multiplied RIN should be kept sufficiently below the noise floor of the system. We found [8] that to keep the multiplied noise peaks 10 dB below the noise floor of the system,

$$RIN(f_{RLX}) \leq \frac{0.4}{D_{sys}} \tag{4}$$

is required. In Eq. (4), $R(f_{RLX})$ is the RIN at the relaxation oscillation frequency f_{RLX} and D_{sys} is the compression dynamic range (Hz) of the system.

IV. Dynamic Range Degradation

The output intercept IP_{op} of the optical link is simply the product of the RF gain G_{op} and input intercept IP_m of the link: $IP_{op} = G_{op} IP_m$. Using Eqs. (1) and (2), we obtain

$$IP_{op} = 10P_{op}^{1\text{dB}} = 4I_{op}^2 R_L \tag{5}$$

where $P_{op}^{1\text{dB}}$ is the output 1-dB compression of the optical link. It is important to note that the output intercept of the link is determined solely by the photoelectric power in the photoreceiver and is, therefore, independent of the modulator characteristics. It is also evident that the higher the photocurrent, the larger IP_{op} will be.

From our analysis, the degradation $\Delta(SFD_{sys})$ of the spur-free dynamic range of the system caused by the insertion of the optical link can be expressed as [8]

$$\Delta(SFD_{sys}) \approx 6.7 \log\left(1 + \frac{G_{LNA} IP_{sys}}{IP'_{op}}\right) \tag{6}$$

where G_{LNA} is the gain of the low-noise amplifier of the antenna, IP_{sys} is the input third-order intercept of the system, and IP'_{op} is both the input and the output (because of the unity gain) third-order intercept of the loss-compensated link. One may express IP'_{op} as

$$\frac{1}{IP'_{op}} = \frac{1}{G_{op} IP_{pr}} + \frac{1}{IP_{op}} \tag{7}$$

If Eq. (3) is satisfied, $IP'_{op} \approx IP_{op} = 4I_{ph}^2 R_L$, and from Eq. (7), the dynamic range of the system is solely limited by the photoelectric power. On the other hand, if

$IP_{op} >> G_{op}IP_{pr}$, then the preamplifier limits the system's dynamic range.

V. Noise Factor Degradation

The noise factor increase ΔF_{op} of the system, caused by the optical link, can be written as [8]

$$\Delta F_{op} \approx \frac{T_{op}F_{pr} + 1160I_{ph} + 116I_{ph}^2}{T_{LNA}G_{LNA}} + \frac{1}{4}SNR_{sys}RIN(f - f_m) \quad (8)$$

where T_{op} and T_{LNA} are the temperatures of the link and the low-noise amplifier, respectively; F_{pr} is the noise factor of the preamplifier; I_{ph} is expressed in mA; and SNR_{sys} is the signal-to-noise ratio (in a 1-Hz bandwidth) at the input of the low-noise amplifier. In Eq. (8), the first term is the noise factor contribution from the preamplifier; the second term is from the shot noise of the optical link (assuming $R_L = 50\ \Omega$); the third term is from the baseband RIN of the laser (assuming a baseband RIN of -165 dB/Hz); and the last term is due to the multiplied relaxation oscillation noise. This last term is frequency dependent and peaks at $f = f_m \pm f_{RLX}$.

At low modulation levels, the contribution to the noise factor from the multiplied RIN is small. However, at high modulation levels, the contribution from the multiplied RIN peaks becomes more important around $f = f_m \pm f_{RLX}$. Because the largest signal-to-noise ratio of the system is set by the dynamic range D_{sys}, the maximum ΔF_{op} is obtained by replacing SNR_{sys} with D_{sys} in Eq. (8).

VI. Link Design

As mentioned above, the ultimate aim of the study was to design a photonic link for the Venus station of the NASA Deep Space Network at Goldstone, California. Thus, the parameters of the system in the Venus station were used to determine the appropriate parameters of the photonic link. In the Venus station, the gain of the antenna's low-noise amplifier was $G_{LNA} = 36$ dB, and the input noise temperature was $T_{LNA} = 36.8$ K. The temperature at the input of the optical link was $T_{op} = 290$ K. The input third-order intercept point and 1-dB compression of the system before inserting the optical link were $IP_{sys} = -47$ dBm and $P_{sys}^{1dB} = -53$ dBm, respectively, and were limited by the system's X-band downconverter. The corresponding compression dynamic range of the system was $D_{sys} = 131$ dB-Hz, and the bandwidth and the total noise temperature of the system were 500 MHz and 40.4 K, respectively.

According to Eq. (4), a laser with noise $RIN(f_{RLX}) \leq -135$ dB/Hz is required in order to obtain a D_{sys} of 131 dB-Hz. The $RIN(f_{RLX})$ of a diode-pumped YAG laser without noise reduction circuitry is typically as high as -100 dB/Hz and cannot meet this stringent requirement. We selected a diode-pumped YAG ring laser with a built-in noise reduction circuit [10] that has a noise peak $RIN(f_{FLX})$ just low enough (-135 dB/Hz) to meet the requirement of Eq. (4). This is the lowest noise laser that is currently commercially available. For systems with a higher dynamic range, additional external noise reduction schemes [11] must be deployed. The output power of the laser used in our link was 50 mW.

We used a Ti:LiNbO$_3$ Mach–Zehnder modulator with a frequency response of 18 GHz. The stability of this modulator is excellent, but the optical insertion loss is high (10 dB). The receiver we chose has a frequency response from 0.1 to 12 GHz and a load resistance of 50 Ω. With 12 km of single-mode fiber between the modulator and receiver, the photocurrent in the load resistor of the optical receiver was $I_{ph} = 0.25$ mA. The frequency response of the link without the preamplifier was very flat from 130 MHz to 10 GHz.

Due to the high V_π (about 30 V) of the modulator used in the link and the low photocurrent of the optical receiver, the RF insertion loss G_{op} was high at about -60 dB. The preamplifier used to compensate for the loss consisted of two cascaded amplifiers and had a gain of 60 dB, a noise factor of $F_{pr} = 2$, and an output intercept $IP_{pr} = 36$ dBm. Because of the high V_π, IP_{pr} could not meet the requirement of Eq. (3), and the dynamic range of the system was limited by both the preamplifier and the photoelectric power. From Eqs. (8) and (9), the expected degradation $\Delta(SFD_{sys})$ of the system's spur-free dynamic range is calculated to be 9.6 dB. This degradation can easily be reduced to less than 1 dB, as will be shown below.

With the system parameters given above, the maximum noise factor contribution from the multiplied RIN peaks is calculated to be 0.1 at $f = f_m \pm f_{RLX}$, and the noise temperature degradation caused by the first three terms in Eq. (8) is 0.75 percent. Such a small degradation is not expected to be measurable in our experiment.

VII. System Evaluation

As a first step, we evaluated the performance of the photonic link in the laboratory. With the preamplifier

described above, the frequency response of the link was very flat (±1 dB) in the frequency range of interest (from 8 to 9 GHz). The phase noise [12] of the laboratory-based photonic link with 10 m of fiber was measured to be less than −110 dBc-Hz at 1 Hz from the carrier frequency of 8.4 GHz, as shown in Fig. 3. This measurement value was limited by the noise floor of the measurement system. The phase noise at this level is consistent with the previously reported results [2] and is 56 dB below the local oscillator phase-noise specification of the Deep Space Network.

As a next step, we evaluated our optical link at S-band (2.295 GHz) in a test facility (CTA-21) that simulated the operating conditions of a real antenna receiving system in the Deep Space Network. A telemetry signal of 7.2 kbit/sec was first sent to the receiving system and was attenuated to be slightly above the noise floor of the system. The noise figure of the system was calculated from the received bit-error rate. Using this method, we measured the system noise figures with and without the optical link inserted between the low-noise amplifier and the downconverter. We found that the insertion of the 12-km fiber-optic link added no measurable degradation to the receiving system.

Finally, we evaluated the 12-km photonic link at the Venus Deep Space Station (DSS 13) while the Magellan spacecraft was being tracked. As shown in Fig. 1, the optical link was inserted between the low-noise amplifier and the X-band downconverter, both of which were located at the pedestal room of the antenna. When the fiber-optic link was inserted, no changes in the level and the quality of the received spacecraft signal at 8.426 GHz were observed. We also measured the noise temperature of the system before and after the insertion of the optical link and found no observable difference. This is consistent with the result of Eq. (8).

Because of time and equipment constraints, the dynamic range degradation of the system, expected to be 9.6 dB, was not measured. However, this degradation can be greatly reduced by simply replacing the modulator with a state-of-the-art unit that has a V_π less than 10 V and an optical insertion loss less than 4 dB. With the 6-dB improvement in the optical insertion loss, the photocurrent I_{ph} would be increased to 1 mA. With such a modulator in the link, the required gain of the preamplifier would be reduced to 36 dB, and the minimum required output third-order intercept point would be reduced to 29 dBm. Commercial amplifiers are currently available to meet these requirements and, thus, will not limit the system's dynamic range. From Eq. (7), it can be seen that the total degradation in the spur-free dynamic range of the system would be 1 dB. We are currently exploring alternative options [13] for improving the photonic link for systems requiring higher dynamic range.

VIII. Summary and Conclusions

In summary, we have presented analytical results for predicting the noise temperature and the dynamic range degradation of the antenna receiving system caused by the insertion of a photonic link. We have specified design parameters for each component of the link according to the system requirements. In this study, special attention was paid to the effect of the relaxation oscillation noise peak of the laser, resulting in the important conclusion that the maximum allowed noise amplitude is inversely proportional to the dynamic range of the system. We also found that a high level of photoelectric power is critical for a system with a high dynamic range.

Based on the analytical results, we designed and fabricated a photonic link. With this link, we successfully demonstrated photonic antenna remoting capability at X-band in an operating antenna receiving system in NASA's Deep Space Network and concluded that the link added no observable degradation to the 40.4-K noise temperature of the system. We also determined that a modulator with an optical insertion loss of less than 4 dB and a half-wave voltage of less than 10 V may be used in the future to reduce the degradation of the system's dynamic range to less than 1 dB.

Acknowledgments

We thank P. Clark of Lightwave Electronics Corporation for lending us the diode-pumped YAG laser and C. Mitchell, C. Goodson, and DSS 13 crew members for technical assistance.

References

[1] G. E. Betts, L. M. Johnson, C. H. Cox III, and S. D. Lowney, "High-Performance Optical Analog Link Using External Modulator," *IEEE Photonics Technology Letters*, vol. 1, no. 11, pp. 404–406, November 1989.

[2] R. T. Logan and G. F. Lutes, "High Stability Microwave Fiber Optic Systems: Demonstration and Applications," *Proceedings of the 46th Annual Symposium on Frequency Control*, Hershey, Pennsylvania, pp. 310–316, May 27–29, 1992.

[3] W. E. Stephens and T. R. Joseph, "System Characteristics of Direct Modulated and Externally Modulated RF Fiber-Optic Links," *Journal of Lightwave Technology*, vol. LT-5, no. 3, pp. 380–387, March 1987.

[4] W. L. Glomb, "Fiber Optic Links for Antenna Remoting," *SPIE Proceedings*, vol. 1703, pp. 523–527, 1992.

[5] W. L. Glomb, J. D. Farina, and S. W. Merritt, "Minimum Noise Figure Microwave Transmission System," *SPIE Proceedings*, vol. 1703, pp. 528–534, 1992.

[6] R. T. Logan, G. F. Lutes, and L. Maleki, "Microwave Analog Fiber Optic Link for Use in the NASA/JPL Deep Space Network," *The Telecommunications and Data Acquisition Progress Report 42-100, vol. September-December 1989*, Jet Propulsion Laboratory, Pasadena, California, pp. 21–33, February 15, 1990.

[7] G. E. Betts, C. H. Cox III, and K. G. Ray, "20 GHz Optical Analog Link Using an External Modulator," *IEEE Photonics Technology Letters*, vol. 2, no. 12, pp. 923–925, December 1990.

[8] X. S. Yao and L. Maleki, "Influence of an Externally Modulated Photonic Link on a Microwave Communications System," *The Telecommunications and Data Acquisition Progress Report 42-117, vol. January-March 1994*, Jet Propulsion Laboratory, Pasadena, California, pp. 16–28, May 15, 1994.

[9] B. H. Kolner and D. W. Dolfi, "Intermodulation Distortion and Compression in an Integrated Electrooptic Modulator," *Applied Optics*, vol. 26, no. 17, pp. 3676–3680, September 1987.

[10] T. J. Kane, "Intensity Noise in Diode-Pumped Single-Frequency Nd:YAG Lasers and Its Control by Electronic Feedback," *IEEE Photonics Technology Letters*, vol. 2, no. 4, pp. 244–245, April 1990.

[11] D. M. Dagenais, K. P. Koo, and D. Dandridge, "Demonstration of Low-Frequency Intensity Noise Reduction for Fiber Sensors Powered by Diode-Pumped Nd-YAG Lasers," *IEEE Photonics Technology Letters*, vol. 4, no. 5, pp. 518–520, May 1992.

[12] R. T. Logan, L. Primas, G. F. Lutes, and L. Maleki, "Design of a Fiber-Optic Transmitter for Microwave Analog Transmission With High Phase Stability," *The Telecommunications and Data Acquisition Progress Report 42-102, vol. April-June 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 27–33, August 15, 1990.

[13] C. H. Cox III, G. Getts, and L. M. Johnson, "An Analytic and Experimental Comparison of Direct and External Modulation in Analog Fiber-Optic Links," *IEEE Transactions on Microwave Technology and Techniques*, vol. 38, no. 5, pp. 501–508, May 1990.

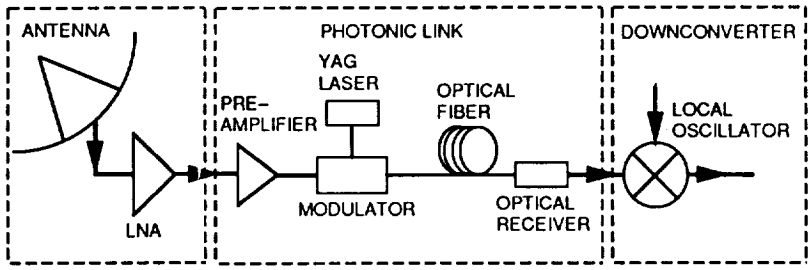


Fig. 1. An illustration of photonic remoting. The downconverter and local oscillator are ordinarily directly connected to the LNA at the antenna. The photonic link permits them to be moved out of the antenna to a remote signal processing center.

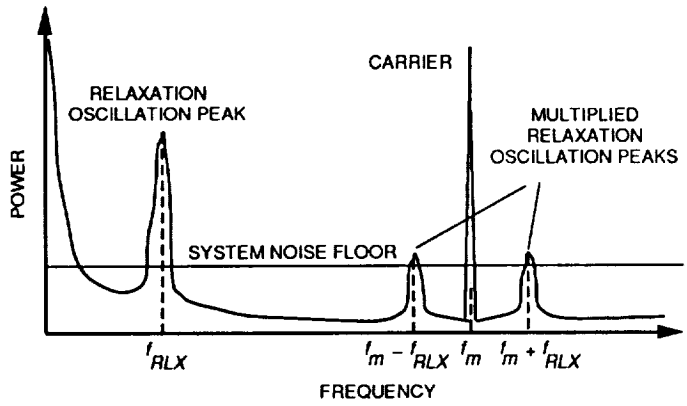


Fig. 2. The laser relaxation peak is multiplied up by the modulation signal. The multiplied peaks may be mistaken for the signal when they are above the system floor noise.

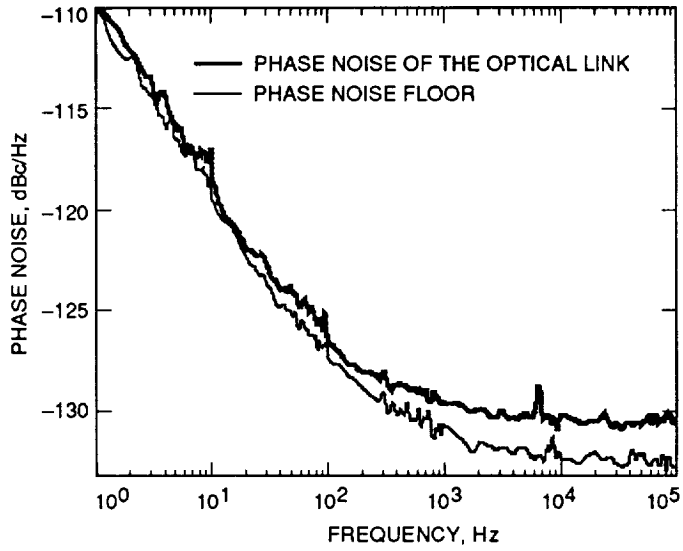


Fig. 3. Close-in phase noise measurement of the photonic link. The noise floor is set by the X-band microwave amplifier used in the measurement. A fiber length of about 10 m was used between the modulator and receiver.

The Electrical Conductivities of Candidate Beam-Waveguide Antenna Shroud Materials

T. Y. Otoshi and M. M. Franco
Ground Antennas and Facilities Engineering Section

The shroud on the beam-waveguide (BWG) antenna at DSS 13 is made from highly magnetic American Society for Testing and Materials (ASTM) A36 steel. Measurements at 8.42 GHz showed that this material (with paint) has a very poor electrical conductivity that is 600 times worse than aluminum. In cases where the BWG mirrors might be slightly misaligned, unintentional illumination and poor electrical conductivity of the shroud walls can cause system noise temperature to be increased significantly. This potential increase of noise temperature contribution can be reduced through the use of better conductivity materials for the shroud walls. An alternative is to attempt to improve the conductivity of the currently used ASTM A36 steel by means of some type of plating, surface treatment, or high-conductivity paints. This article presents the results of a study made to find improved materials for future shrouds and mirror supports.

I. Introduction

The technique used to measure the resistivity of flat stock materials at microwave frequencies was the cavity technique described in [1,2]. Resistivity data on previously measured samples of various metals and plated surfaces have been previously presented in [1–4]. Figure 1 shows the X-band cavity that was used for these measurements. The measurement technique involves placing a flat plate sample of the material to be tested on top of an open cylindrical cavity operating in the TE_{011} mode. Resistivity is determined from measured loaded- and unloaded-Q at a nominal frequency of 8.420 GHz. Slight deviations from perfect flatness of the test sample will cause this center frequency of the test to deviate slightly from nominal. Electrical conductivity is calculated from the measured resistivity using an equation given in [2]. Although electrical conductivities of metals theoretically are frequency independent, in practice when surface roughness becomes a significant fraction of skin depth, the electrical conductivities (that are determined from measured resistivities) could be somewhat frequency dependent.

II. Test Sample Description

Some of the samples tested are shown in Figs. 2 and 3. The ASTM A36 steel samples shown in Fig. 2 were cut from a section of the former bypass shroud on the BWG antenna. It was previously reported [2] that bare-metal-surface ASTM A36 samples had average conductivity values of about 0.01×10^7 mhos/m, while samples painted with thermal diffusive white paint had worse conductivities of about 0.0036×10^7 mhos/m.

For this article, a study was made to find ways to improve the electrical conductivity of ASTM A36 steel by treating the surface with (1) a zinc-plating process and (2) spray painting with cold galvanized paint. For the zinc-plating process, a black-colored dye was used. The use of black dye color was arbitrary, and clear zinc plating could have been specified instead.

Another steel material tested was type 1018 steel (see Fig. 3). This material was used to simulate flat surfaces of steel mirror-support structures in the BWG antenna in some of the shroud noise temperature tests performed at the Microwave Test Facility (MTF) at Goldstone, California. This material was used for these MTF tests because it was known to have dc conductivity and magnetic properties similar to ASTM A36 steel. It was also used because it was readily available in flat sheet stocks at Goldstone. This material was tested with and without white diffusive thermal paint.

Aluminum material is sometimes treated with a type of surface treatment to prevent oxidation. Two types of aluminum samples with surface treatments were fabricated and tested. These samples of aluminum were treated with (1) irriditing and (2) anodizing processes (Fig. 3). In this article, the term "irridite" will be used to describe the surface treatment of aluminum samples or parts by a chemical dipping process. Not generally well known is the fact that irridite, yellow chemical film, and alodine are trade terms referring to identical surface-treatment processes. Confusion sometimes occurs between the terms alodine and anodize, which are not equivalent processes. The former involves a chemical dipping process, while the latter refers to an electrochemical-oxidizing surface-treatment process.

III. Test Results

Table 1 shows a summary of the test results of the described samples. As may be seen in Table 1, the type 1018 steel conductivities for unpainted and painted samples were measured to be 0.0226×10^7 and 0.0081×10^7 mhos/m, respectively, at 8.420 GHz. These results may be compared to 0.01×10^7 and 0.004×10^7 mhos/m for the ASTM A36 structural steel unpainted and painted samples, respectively.

The zinc-plating process on ASTM A36 steel improved the conductivity from 0.01×10^7 to about 0.44×10^7 mhos/m, but it is still about 5 times worse than type 6061 aluminum. As may be seen in the table, for two of the samples, the galvanized-paint process made the conductivity much worse than that for bare ASTM A36 steel and even worse than ASTM A36 steel painted with thermal diffusive white paint. It is suspected that galvanized spray paint is not purely metallic and has lossy epoxy compounds to make the surface more like a lossy dielectric surface. It might seem that if the paints were more highly conductive, then better results would be obtained. However, it was shown in a previous article [2] that a very high grade of silver paint only improved the conductivity from 0.01×10^7 mhos/m for bare metal ASTM A36 steel to 0.022×10^7 mhos/m when silver painted.

Table 1 also shows the results of aluminum samples with surfaces treated with (1) irriditing and (2) anodizing processes. Irriditing caused no noticeable change in the conductivity properties of bare aluminum, while anodizing only degraded the conductivity from 2.31×10^7 to 1.96×10^7 mhos/m, which is still acceptable. From previous tests [2], it was found that primer and thermal diffusive white paint also did not significantly degrade the resultant electrical conductivity of aluminum, but did significantly degrade the conductivity of ASTM A36 steel.

Relative permeability values (relative to air) of the test samples are shown in Table 1. It can be seen that the steel materials are highly magnetic with relative permeabilities being in the 9000–10,000 range. For comparison, aluminum and copper have relative permeabilities of unity, while most types of stainless steel have relative permeabilities of less than 1.10 [2].

It was pointed out by Vane [5] that conductive metals having high permeability tend to have poor *effective* conductivities due to the fact that high permeabilities cause the skin depth to become very small. The effects of surface roughness and surface layers (of paints and oxides) are accentuated because, when skin depth is small, most of the RF currents will tend to flow along the irregular surfaces and even partially inside the treated layers (oxides and paints). Hence, it is not surprising that highly magnetic materials such as ASTM A36 and type 1108 steel with painted layers and poor surface finishes tend to have very poor conductivities. For materials with relative permeabilities close to unity, the skin depth is larger, and RF currents will tend to flow through more volume of the conductor rather than just at the surface. An equation showing the relationship between skin depth and relative permeability was given in [2].

IV. Conclusions

Test results showed that zinc plating the surface of ASTM A36 steel improved the electrical conductivity of

bare metal from 0.01×10^7 to 0.44×10^7 mhos/m. For comparison, the conductivity was 0.0036×10^7 mhos/m for a sample of this steel that was primed and painted with thermal diffusive white paint. Even with this improvement, the zinc-plated samples were still about 5 times worse than aluminum. Galvanized spray paint is not recommended because galvanized-painted ASTM A36 samples resulted in conductivities about the same or much worse than samples with thermal diffusive white paint. It can be stated that for highly magnetic steels with rough surfaces, the plating process will help to improve the conductivity significantly, but painting the surfaces with cold conductive paints will not.

Either an irriditing or an anodizing process should be considered for preventing oxidation of aluminum shrouds or mirrors. If anodized aluminum material is chosen as the material for future BWG shrouds, then it is recommended that, for better optical lighting purposes, a clear dye anodizing process should be specified.

Acknowledgment

Pablo Narvaez of the JPL Magnetics Laboratory measured the relative permeabilities of the steel samples.

References

[1] R. C. Clauss and P. D. Potter, *Improved RF Calibration Techniques—A Practical Technique for Accurate Determination of Microwave Surface Resistivity*, Technical Report 32-1526, vol. XII, Jet Propulsion Laboratory, Pasadena, California, pp. 54–67, December 15, 1972.

[2] T. Y. Otoshi, M. M. Franco, and H. F. Reilly, Jr.,"The Electrical Conductivities of the DSS-13 Beam-Waveguide Antenna Shroud Material and Other Antenna Reflector Surface Materials," *The Telecommunications and Data Acquisition Progress Report 42-108, vol. October–December 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 154–163, February 15, 1992.

[3] E. H. Thom and T. Otoshi, "Surface Resistivity Measurements of Candidate Subreflector Surfaces," *The Telecommunications and Data Acquisition Progress Report 42-65, vol. July–August 1981*, Jet Propulsion Laboratory, Pasadena, California, pp. 142–150, October 15, 1981.

[4] H. F. Reilly, J. J. Bautista, and D. A. Bathker, "Microwave Resistivity of Several Materials at Ambient Temperature," *The Telecommunications and Data Acquisition Progress Report 42-80, vol. October–December 1984*, Jet Propulsion Laboratory, Pasadena, California, pp. 8–11, February 15, 1985.

[5] A. B. Vane, *Measurements of Effective Conductivity of Metallic Surfaces at 3000 Megacycles and Correlation with Surface Conditions and DC Conductivity*, Stanford Microwave Laboratory Report No. 4:III-9, Palo Alto, California, pp. 72–78, 1949.

Table 1. Summary of test results of the cavity samples.

Description	Relative permeability[a]	Number of samples tested	Average surface roughness, μm	Test frequency, GHz	Average surface resistivity, ohms/square	Effective conductivity, mhos/m	Comments
Type 1018 steel	>9000	2	1.47 (58 μin.)	8.42221	0.38350	0.02261 x 10^7	
Type 1018 steel, same as above except painted with primer and Triangle #6 thermal diffusive white paint	>9000	1	0.64 (25 μin.)	8.40200	0.63861	0.00813 x 10^7	Paint had a big effect.
BWG antenna shroud ASTM A36 bare steel	9985	1	>6.35 >(250 μin.)	8.42049	0.57737	0.00997 x 10^7	This compares well with the 0.01 x 10^7 value in [2].
BWG antenna shroud ASTM A36 bare steel, with rust spots	9985	2	12.7 (500 μin.)	8.42124	0.63401	0.00830 x 10^7	
BWG antenna shroud ASTM A36 bare steel, rust spots partially removed	9985	1	12.7 (500 μin.)	8.42176	0.63288	0.00830 x 10^7	Rust removal had no effect.
BWG antenna shroud ASTM A36 steel, primer only	9985	1	1.78–2.62 (70-103 μin.)	8.41173	0.63859	0.00814 x 10^7	Compare this with 0.01 x 10^7 for bare metal.
BWG antenna shroud ASTM A36 steel, primer and Triangle no. 6 thermal diffusive white paint	9985	1	2.36 (93 μin.)	8.43990	0.91065	0.00400 x 10^7	Compare this with 0.0036 x 10^7 reported in [2].
BWG antenna shroud ASTM A36 steel, zinc plating (with black dye)	9985	3	6.35 (250 μin.)	8.41959	0.08650	0.4442 x 10^7	Significant improvement over bare metal.
BWG antenna shroud ASTM A36 steel, cold galvanized spray paint	9985	2	2.79–4.27 (110–168 μin.)	8.41544	3.36860	0.000293 x 10^7	Very bad conductivity might be due to surface roughness and paint.
BWG antenna shroud ASTM A36 steel, cold galvanized spray paint	9985	3	3.18 (125 μin.)	8.41399	0.85831	0.00451 x 10^7	About the same as with thermal diffusive white paint.

[a] The relative permeability values are relative to air.

Table 1 (cont'd)

Description	Relative permeability[a]	Number of samples tested	Average surface roughness, μm	Test frequency, GHz	Average surface resistivity, ohms/square	Effective conductivity, mhos/m	Comments
Copper	1.000	3	0.71 (28 μin.)	8.42033	0.02666	4.6770 x 10^7	
Brass	1.000	1	0.53 (21 μin.)	8.42443	0.05020	1.3197 x 10^7	
Aluminum 6061-T6	1.000	2	0.33 (13 μin.)	8.42422	0.03792	2.3129 x 10^7	
Aluminum 6061-T6, irridite	1.000	1	0.41 (16 μin.)	8.42485	0.03790	2.3159 x 10^7	
Aluminum 6061-T6, black anodized type II[b]	1.000	1	0.46 (18 μin.)	8.42458	0.04122	1.9574 x 10^7	
Aluminum 6061-T6, black anodized type III[c]	1.000	1	0.46 (18 μin.)	8.42358	0.04121	1.9585 x 10^7	

[b] Type II refers to an anodizing process that treats the surface only.
[c] Type III refers to an anodizing process that typically goes about 0.013 mm (0.5 mil) into the metal.

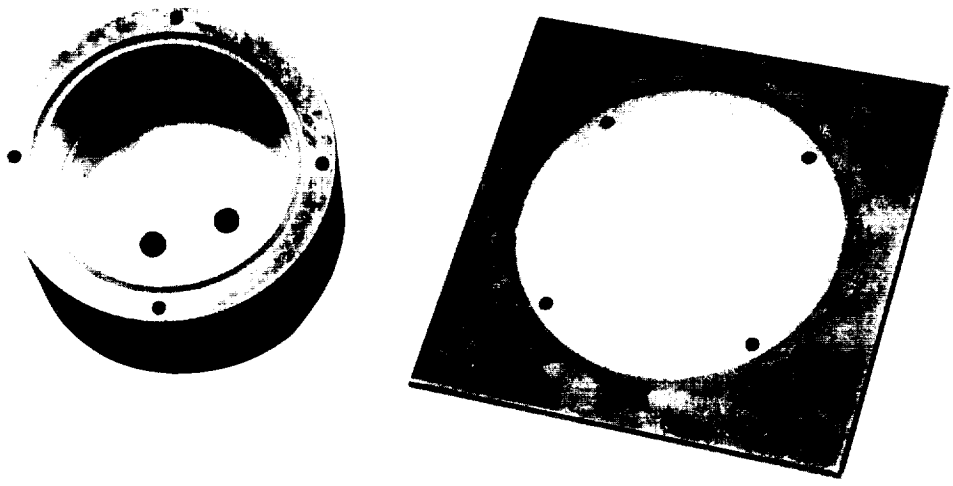


Fig. 1. The X-band cavity.

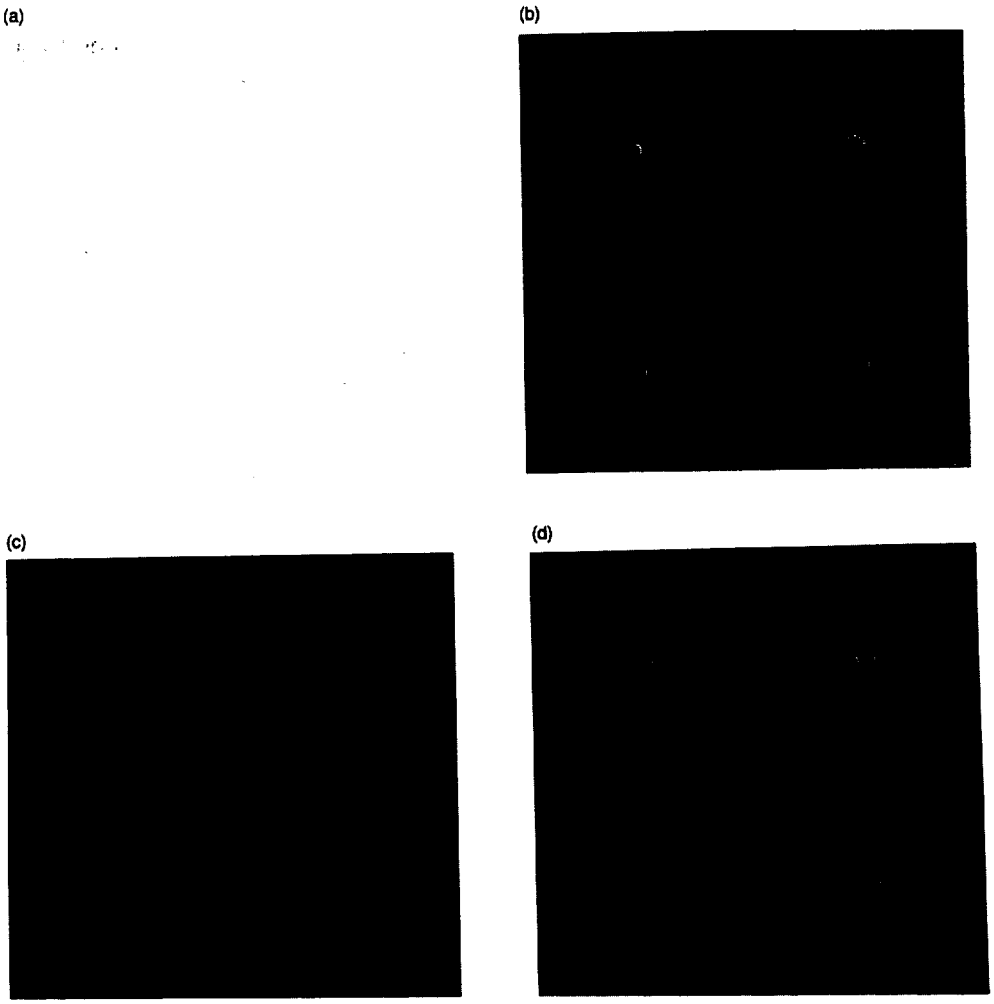


Fig. 2. Samples of the tested ASTM A36 shroud material (10.16 cm x 10.16 cm): (a) painted with primer and Triangle #6 thermal diffusive white paint; (b) bare metal; (c) zinc-plating (with black dye) surface treatment; and (d) galvanized spray-painted surface.

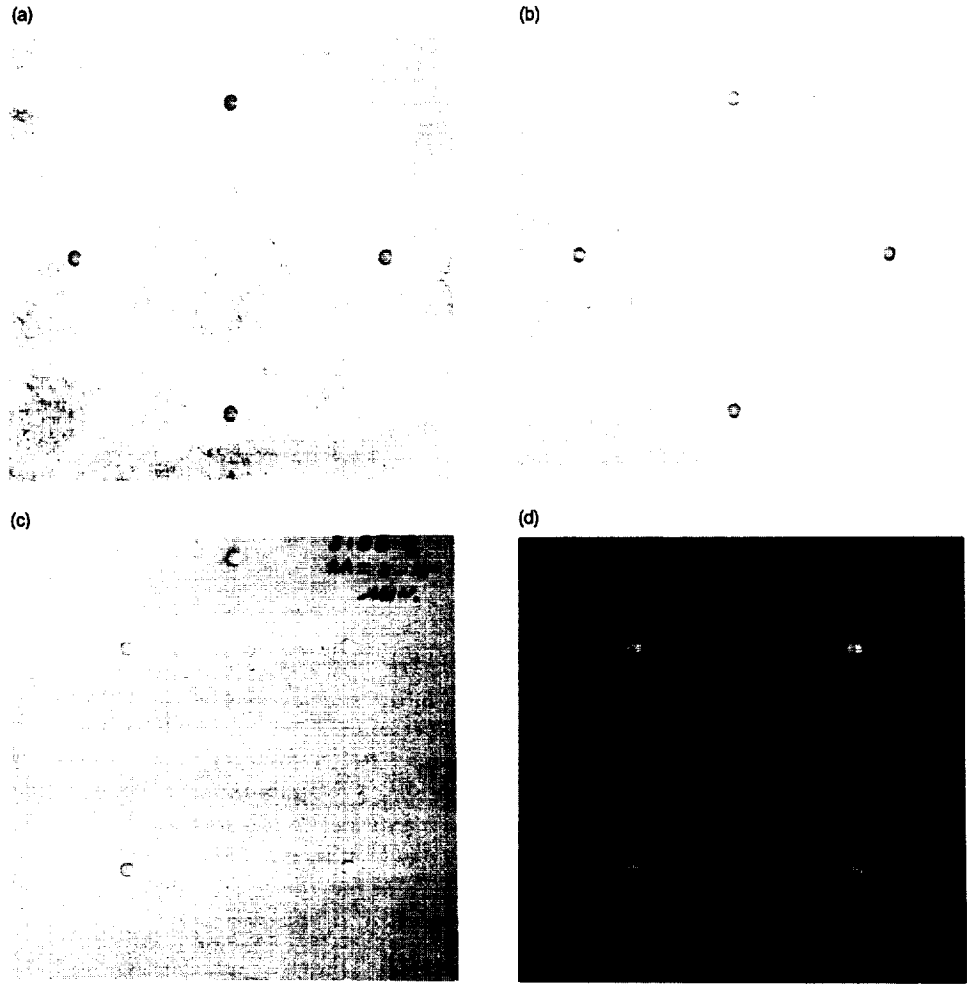


Fig. 3. Other sample candidate shroud materials tested (10.16 cm x 10.16 cm): (a) type 1018 steel bare metal; (b) aluminum 6061; (c) aluminum 6061 with irridite surface treatment; and (d) aluminum 6061 with black anodized surface treatment.

TDA Progress Report 42-117

May 15, 1994

Thin-Ribbon Tapered Coupler for Dielectric Waveguides

C. Yeh[1]

T. Y. Otoshi

Ground Antennas and Facilities Engineering Section

F. I. Shimabukuro

Aerospace Corporation

Los Angeles, California

A recent discovery shows that a high-dielectric constant, low-loss, solid material can be made into a ribbon-like waveguide structure to yield an attenuation constant of less than 0.02 dB/m for single-mode guidance of millimeter/submillimeter waves. One of the crucial components that must be invented in order to guarantee the low-loss utilization of this dielectric-waveguide guiding system is the excitation coupler. The traditional tapered-to-a-point coupler for a dielectric rod waveguide fails when the dielectric constant of the dielectric waveguide is large. This article presents a new way to design a low-loss coupler for a high- or low-dielectric constant dielectric waveguide for millimeter or submillimeter waves.

I. Introduction

A recent discovery shows that a high-dielectric constant, low-loss, solid material, such as TiO ($\epsilon_1/\epsilon_0 = 100$, $\tan\delta = 0.00025$) or Rexolite ($\epsilon_1/\epsilon_0 = 2.55$, $\tan\delta = 0.001$), can be made into a ribbon-like waveguide structure to yield an attenuation constant of less than 0.02 dB/m for single-mode guidance of millimeter or submillimeter waves [1]. This discovery provides the impetus to perfect a practical low-loss guided transmission system for these short wavelengths. As a comparison of loss, the attenuation constant of a WR28 copper waveguide is 0.58 dB/m at 32 GHz. One of the crucial components that must be invented in order to guarantee the low-loss utilization of this dielectric-waveguide guiding system is the excitation coupler.

A conventional technique to minimize the coupling loss of an excitation coupler is to taper the coupling end of a dielectric waveguide to a very narrow, sharp apex [2]. However, this method fails when the relative dielectric constant of the dielectric waveguide is much greater than unity, the free-space value. Another technique is to shape the coupling end of the dielectric waveguide into a cusp-like form [3]. This cusp design, which is based on the direct ap-

[1] Independent consultant to the Ground Antennas and Facilities Engineering Section.

plication of geometrical optics to minimize reflection, is applicable to guided wave structures whose cross-sectional dimensions are many free-space wavelengths.

Neither of the two techniques described above is applicable to the case of a dielectric waveguide with (1) a large dielectric constant difference between the core and cladding material and (2) a guide dimension of less than the free-space wavelength, such as the ultralow-loss millimeter or submillimeter dielectric ribbon waveguide [1]. This article will describe a different approach for designing a low-loss coupler for this case.

II. Approaches

A. Impedance Matching Approach

It is possible to define a wave impedance for a given propagating mode along a dielectric waveguide of a given cross section [4]. The wave impedance may be viewed as the characteristic impedance of a transmission line (i.e., the dielectric waveguide). Thus, a tapered (tapered in the direction of propagation of the guided wave, the z-direction) dielectric waveguide is then viewed as an inhomogeneous transmission line with z-dependent characteristic impedance [5].

The problem of designing a transition between a dielectric waveguide and free space becomes one of impedance matching the characteristic impedance of a given dielectric waveguide mode to the characteristic impedance (377 ohms) of free space. It is known that, between two transmission lines with different characteristic impedances, the best matching can be achieved over a broadband of frequencies with a tapered section of transmission line [6,7]. For this reason, in practice, transitions between a dielectric waveguide and free space are generally of the tapered-transmission line type. The shape of the taper is based upon one that produces the minimum mismatch losses over the frequency band of interest.

In principle, this tapering approach to a narrowed, thin apex is workable for any dielectric waveguide with any dielectric constant. In practice, however, this approach is workable only for a dielectric waveguide with a relative dielectric constant near unity. This is because when the core dielectric constant is large, to achieve good matching with free space, the tapered section must be very, very long and the cross-sectional dimension of the taper must be very small, making the tapered section extremely difficult to handle and align. The stability of the modal field in adhering to the tapered guiding section also becomes questionable. In other words, it is difficult to excite a guided wave along this type of tapered section for a high-dielectric constant guide, even though good impedance matching is present. It appears that the surface area for this high-dielectric constant tapered section is too small to capture the incident wave and to transform it to a guided wave [1].

B. Ribbon Transition Approach

Instead of tapering a high-dielectric constant guide to a very small cross-sectional area, the guide should be flattened to a large surface area and very thin thickness [1], i.e., the transition region should be tapered to a thin, flat, but wide ribbon, as shown in Fig. 1. To provide further improvements in matching, the end of the transition region can be further extended with a comb-like structure, as shown in Fig. 2. Because of the large surface area of this structure, it can be easily supported mechanically without causing noticeable interference to the electromagnetic field. The large surface area also enables the guided surface wave to better attach to the guiding structure, thus improving the launching efficiency [1].

It can be seen from Fig. 2 that this ribbon transition region is very different from the conventional tapered-to-a-point transition region used by all earlier investigators to obtain a better match between the free-space region and the dielectric waveguide region.

III. Transitions

The analytical foundation for the transitions described in this section is given in the Appendix.

A. Metallic Rectangular Waveguide-to-Ribbon Dielectric Waveguide Transition

The ribbon transition is tailor-made for the ultralow-loss ribbon dielectric waveguide for millimeter or submillimeter wavelengths. Figure 3 shows the transition region between a metallic rectangular waveguide supporting the dominant TE_{10} mode and a dielectric ribbon waveguide supporting the dominant ${}_eHE_{11}$ mode [1]. A flared metallic horn is used to provide the wide width for the ribbon transition region.

B. Microstripline-to-Ribbon Dielectric Waveguide Transition

Another important, practical transition is between the microstripline [8] and the ribbon dielectric waveguide. Figure 4 is a sketch of such a transition. In order to minimize Fresnel-type reflection losses and to accommodate the wide width of the ribbon transition, an additional transition section is added as shown. In the added transition region, it

is necessary to taper the microstripline dielectric filling to a vanishingly thin wedge while the upper conducting strip is flared to a wide width. It is seen that the field concentrated under the narrow, upper conducting strip of the microstripline is spread out through the added transition region to cover the wider width of the ribbon transition.

C. Transition for a Round Dielectric Waveguide

If it is desirable to excite propagating fields on a round (circular) dielectric waveguide, the transition region, as shown in Fig. 5, can be designed. Here, the circular core is flared in one transverse direction and compressed in the other transverse direction into a flat ribbon and then tapered to a very narrow wedge. Again, the idea is to retain the largest possible surface area to capture the guided fields. Thus, an incident plane wave can easily be captured smoothly by the wide-but-thin wedge-shaped dielectric transition.

IV. Conclusions

A semiheuristic way to design a low-loss excitation coupler for a high-dielectric constant dielectric waveguide has been presented. Unlike the traditional tapered-to-a-point transition region, a tapered-to-a-thin-sheet transition region is proposed. It is found that the thin sheet gives better stability for the surface wave in the transition region, thus improving the launching efficiency for the wave onto a dielectric waveguide.

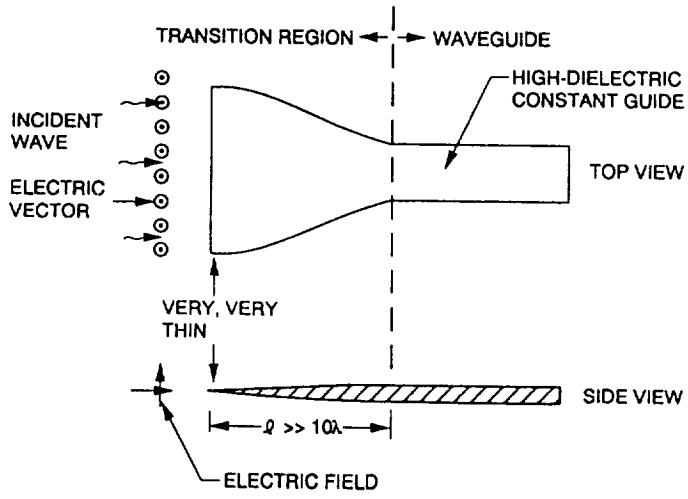


Fig. 1. Tapered section for a high-dielectric constant waveguide.

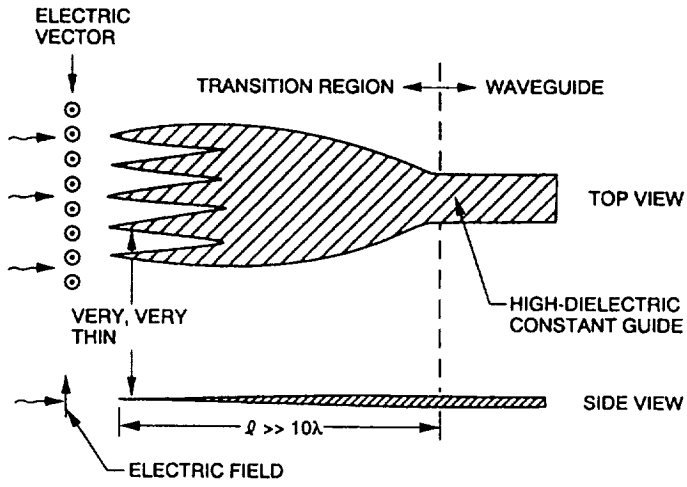


Fig. 2. Comb-like transition region.

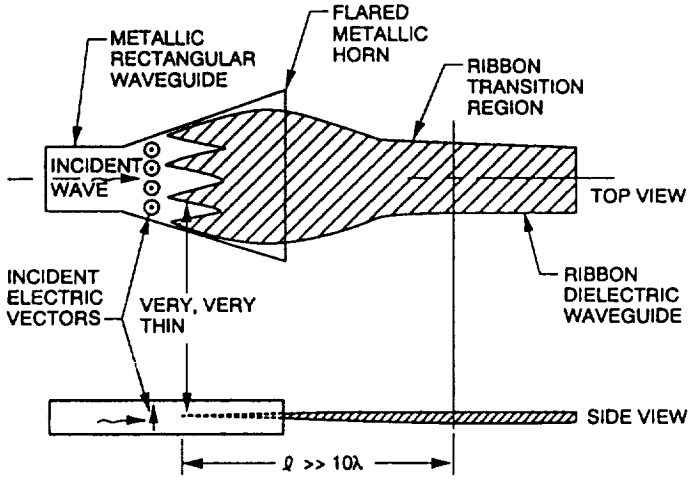


Fig. 3. Metallic rectangular waveguide-to-ribbon dielectric waveguide transition.

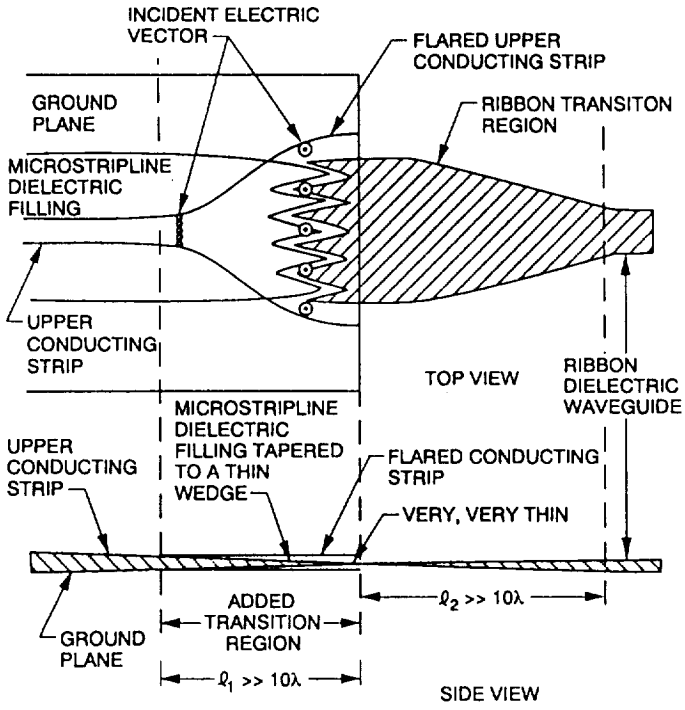


Fig. 4. Microstripline-to-ribbon dielectric waveguide transition.

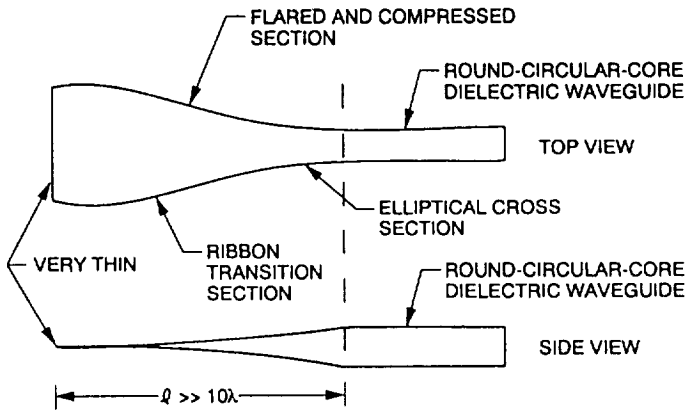


Fig. 5. Transition from a ribbon-wedge to a round-circular-core dielectric waveguide.

Appendix

Analytical Foundation

The following provides the theoretical foundation [9,10] for the heuristic approach to the design of low-loss, efficient transitions to dielectric waveguides. The canonical problem is the study of the propagation characteristics of a guided wave on a dielectric slab of thickness d, enclosed by two conducting plates as shown in Fig. A-1. Referring to Fig. A-1, it is of interest to learn how a guided wave is detached from the conducting planes as a (the separation of the two conducting plates) is increased. We wish to know how the guided field and its wave impedance of a parallel plate guide are affected by the presence of a very thin high-dielectric constant slab.

The guided mode of interest is the one whose electric field is polarized in the x-direction and whose electric field distribution is even with respect to positive and negative values of x. A propagation factor and a time-dependent factor of $e^{-j\beta z + j\omega t}$, where β is the propagation constant and ω is the frequency of the wave, is implied and suppressed for all field components. In region (1), the field components for the mode of interest are

$$E_z^{(1)} = A_1 \sin \sqrt{k_1^2 - \beta^2} \quad x \tag{A-1}$$

$$E_x^{(1)} = \frac{-j\beta A_1}{\sqrt{k_1^2 - \beta^2}} \cos \sqrt{k_1^2 - \beta^2} \quad x \tag{A-2}$$

$$H_y^{(1)} = \frac{-j\omega\epsilon_1 A_1}{\sqrt{k_1^2 - \beta^2}} \cos \sqrt{k_1^2 - \beta^2} \quad x \tag{A-3}$$

where A_1 is an arbitrary constant and $k_1^2 = \omega^2\mu_0\epsilon_1$. In region (0), the field components are

$$E_z^{(0)} = A_0 \sin \left[\sqrt{k_0^2 - \beta^2}\left(x - \frac{a}{2}\right)\right]$$

$$\text{for } \beta < k_0 \tag{A-4a}$$

$$= A_0 \sinh \left[\sqrt{\beta^2 - k_0^2}\left(x - \frac{a}{2}\right)\right]$$

$$\text{for } \beta > k_0 \tag{A-4b}$$

$$E_x^{(0)} = \frac{-j\beta A_0}{\sqrt{k_0^2 - \beta^2}} \cos \left[\sqrt{k_0^2 - \beta^2}\left(x - \frac{a}{2}\right)\right]$$

$$\text{for } \beta < k_0 \tag{A-5a}$$

$$= \frac{-\beta A_0}{\sqrt{k_0^2 - \beta^2}} \cosh \left[\sqrt{\beta^2 - k_0^2}\left(x - \frac{a}{2}\right)\right]$$

$$\text{for } \beta > k_0 \tag{A-5b}$$

$$H_y^{(0)} = \frac{-j\omega\epsilon_0 A_0}{\sqrt{k_0^2 - \beta^2}} \cos \left[\sqrt{k_0^2 - \beta^2}\left(x - \frac{a}{2}\right)\right]$$

$$\text{for } \beta < k_0 \tag{A-6a}$$

$$= \frac{-\omega\epsilon_0 A_0}{\sqrt{k_0^2 - \beta^2}} \cosh \left[\sqrt{\beta^2 - k_0^2}\left(x - \frac{a}{2}\right)\right]$$

$$\text{for } \beta > k_0 \tag{A-6b}$$

where A_0 is an arbitrary constant and $k_0^2 = \omega^2\mu_0\epsilon_0$. Matching the boundary conditions at $x = d/2$ and solving for the nontrivial solution yields the dispersion relation

$$\frac{\epsilon_0}{\epsilon_1}\frac{\sqrt{k_1^2 - \beta^2}}{\sqrt{k_0^2 - \beta^2}} \tan \left[\sqrt{k_1^2 - \beta^2}\frac{d}{2}\right] =$$

$$\tan \left[\sqrt{k_0^2 - \beta^2}\left(\frac{d}{2} - \frac{a}{2}\right)\right]$$

$$\text{for } \beta < k_0 \tag{A-7}$$

or

$$-\frac{\epsilon_0}{\epsilon_1}\frac{\sqrt{k_1^2 - \beta^2}}{\sqrt{\beta^2 - k_0^2}} \tan \left[\sqrt{k_1^2 - \beta^2}\frac{d}{2}\right] =$$

$$\tanh \left[\sqrt{\beta^2 - k_0^2}\left(\frac{d}{2} - \frac{a}{2}\right)\right] \quad \text{for } (k_1 > \beta > k_0) \tag{A-8}$$

It is seen that, if $d << a$ and if $\beta < k_0$, the fields inside the conducting plates are basically those of a parallel-plate conducting waveguide whose propagation constant is given by

$$\left(\frac{\beta a}{2}\right) \simeq \sqrt{\left(k_0 \frac{a}{2}\right)^2 - \pi^2} \qquad \text{(A-9)}$$

with a phase velocity, ω/β, faster than the speed of light in a vacuum. As $k_0 a$ becomes large and $\beta > k_0$, the fields are basically governed by the fields on a slab dielectric waveguide with a phase velocity slower than the speed of light in a vacuum. A sketch of the transformation (or launching) of the transverse electric field for the dominant mode from an enclosed waveguide to a dielectric ribbon guide through a wedge dielectric flared-horn transition region (which slows down the phase velocity of the wave) is shown in Fig. A-2. It is seen that the TM_1 mode in a conducting parallel-plate waveguide can be launched smoothly onto a flat ribbon waveguide as the dominant TM_0 mode if the transition region contains a gradually tapered section of the dielectric ribbon and a gradually flared conducting-plates region. Due to the thinness and wideness of the tapered dielectric ribbon, this type of transition remains effective even when the dielectric constant of the ribbon is very large.

Let us consider this transition from the impedance matching point of view. The wave impedance of the TM_1 mode in the parallel-plate waveguide is [11]

$$_pZ_{TM_1} = \frac{\beta^{(TM_1)}}{\omega\epsilon_0} \qquad \text{(A-10)}$$

where the subscript p means a parallel-plate guide and $\beta^{(TM_1)}$ represents the propagation constant of the TM_1 mode, expressed as

$$\beta^{(TM_1)} = \sqrt{\omega^2\mu\epsilon_0 - \left(\frac{2\pi}{a}\right)^2} \qquad \text{(A-11)}$$

Using the same definition for wave impedance as the conducting guide, for a dielectric slab guide, one has

$$_dZ_{TM_0}^{(0),(1)} = \frac{E_x^{(0),(1)}}{H_y^{(0),(1)}} = \frac{\beta^{(TM_0)}}{\omega\epsilon_{0,1}} \qquad \text{(A-12)}$$

where the subscript d means a dielectric slab (ribbon) waveguide and the subscript TM_0 means the dominant mode on the dielectric slab guide; E_x and H_y are, respectively, the transverse electric field and the transverse magnetic field of the TM_0 mode; and the superscript (0) or (1) refers to the region outside the dielectric slab (the cladding region) or the region inside the dielectric slab (the core region). Unlike the case for an empty conducting waveguide, the wave impedance for a dielectric slab guide, defined as the ratio of the transverse electric field and the transverse magnetic field, depends on the region in which it applies, as shown by Eq. (A-12). In fact,

$$\frac{_dZ_{TM_0}^{(0)}}{_dZ_{TM_0}^{(1)}} = \frac{\epsilon_1}{\epsilon_0} \qquad \text{(A-13)}$$

This perhaps highlights the reason why the usual tapering technique, i.e., conical tapering to a thin apex, does *not* work well for high-dielectric constant dielectric waveguides. The same can be said for the tapered ribbon guide, but, because the surface area for the surface wave to "cling to" does not decrease, the surface wave can still be attached to the ribbon structure, even when the thickness of the ribbon is very small and when the dielectric constant of the ribbon is high. Another way to look at this problem is that, since most of the guided energy of a thin-ribbon electric waveguide is outside the dielectric material, the wave impedance of the dielectric ribbon can be approximately represented by the "outside" wave impedance, i.e., $_dZ_{TM_0}^{(0)} = \beta^{(TM_0)}/\omega\epsilon_0$, which is very close to the wave impedance of the empty parallel-plate waveguide, i.e., $_pZ_{TM_1} = \beta^{(TM_1)}/\omega\epsilon_0$, implying good impedance matching or good launching of the TM_1 wave from the parallel-plate guide to the TM_0 mode of the thin-ribbon dielectric waveguide.

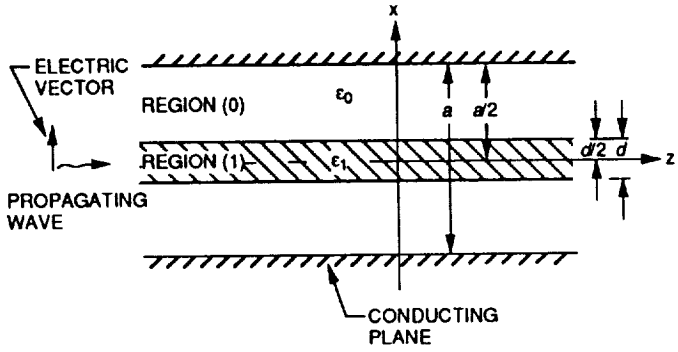


Fig. A-1. Geometry of the canonical problem.

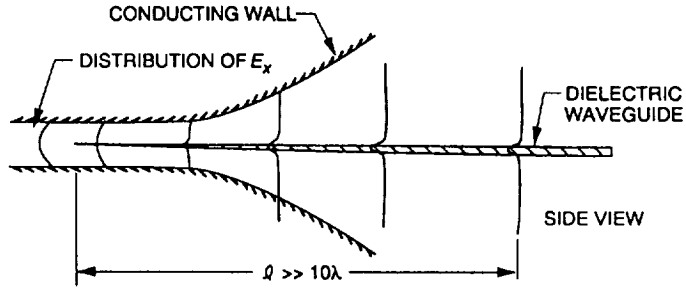


Fig. A-2. Transverse electric field transition from an enclosed guide to an open dielectric ribbon guide.

References

[1] C. Yeh, F. I. Shimabukuro, and J. Chu, "Dielectric Ribbon Waveguide: an Optimum Configuration for Ultra-Low-Loss Millimeter/Submillimeter Dielectric Waveguide," *IEEE Transactions on Microwave Theory Technology*, vol. 38, no. 6, pp. 691–702, June 1990.

[2] T. N. Trinh, J. A. G. Malberbe, and R. Mittra, "A Metal-to-Dielectric Waveguide Transition with Application to Millimeter-Wave Integrated Circuits," *1980 IEEE-MTT-S International Microwave Symposium*, Washington, DC, pp. 205–207, May 1980.

[3] H. M. Presby and C. A. Edwards, "Near 100% Efficiency Fibre Microlenses," *Electronic Letters*, vol. 28, no. 6, pp. 582–584, March 12, 1992.

[4] C. Yeh, "Elliptical Dielectric Waveguides," Ph.D. thesis, California Institute of Technology, Pasadena, California, 1962, and *Journal of Applied Physics*, vol. 33, no. 11, pp. 3235–3243, November 1962.

[5] A. Ishimaru, *Electromagnetic Wave Propagation, Radiation, and Scattering*, Englewood Cliffs, New Jersey: Prentice Hall, 1991.

[6] R. E. Collin, *Foundations for Microwave Engineering*, New York: McGraw-Hill, pp. 237–254, 1966.

[7] M. Jamil Ahmed, "Impedance Transformation Equations for Exponential, Cosine-Squared, and Parabolic Tapered Transmission Lines," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-29, no. 1, pp. 67–68, January 1981.

[8] K. C. Gupta, R. G. Garg, and I. J. Bahl, *Microstrip Lines and Slotted Lines*, Dedham, Massachusetts: Artech House Books, 1982.

[9] R. E. Collin, *Field Theory of Guided Waves*, New York: McGraw-Hill, 1960.

[10] S. Ramo, J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*, 2nd ed., New York: Wiley, 1984.

[11] F. E. Borginis and C. H. Papas, "Electromagnetic Waveguides and Resonators," *Handbuch der Physik*, vol. XVI, Berlin, Germany: Springer Verlag, pp. 285–423, 1958.

TDA Progress Report 42-117

May 15, 1994

Design and Implementation of a Beam-Waveguide Mirror Control System for Vernier Pointing of the DSS-13 Antenna

L. S. Alvarez, M. Moore, W. Veruttipong, and E. Andres
Ground Antennas and Facilities Engineering Section

The design and implementation of an antenna beam-waveguide (BWG) mirror position control system at the DSS-13 34-m antenna is presented. While it has several potential applications, a positioner on the last flat-plate BWG mirror (M6) at DSS 13 is installed to demonstrate the conical scan (conscan) angle-tracking technique at the Ka-band (32-GHz) operating frequency. Radio frequency (RF) beam-scanning predicts for the M6 mirror, computed from a diffraction analysis, are presented. From these predicts, position control system requirements are then derived. The final mechanical positioner and servo system designs, as implemented at DSS 13, are illustrated with detailed design descriptions given in the appendices. Preliminary measurements of antenna Ka-band beam scan versus M6 mirror tilt made at DSS 13 in December 1993 are presented. After reduction, the initial measurements are shown to be in agreement with the RF predicts. Plans for preliminary conscan experimentation at DSS 13 are summarized.

I. Introduction

Since 1976, the NASA/JPL Deep Space Network (DSN) antennas have successfully employed the conical scan (conscan) angle-tracking technique to track spacecraft in deep space. This method nutates the entire tipping portion of the antenna, and therefore displaces the microwave beam about 0.1 dB from boresight in elevation and cross-elevation. Beam-pointing error estimates are then derived from the resultant receiver output signature. The conscan technique has persevered through the DSN upgrade from S-band (2.3-GHz) to the X-band (8.45-GHz) frequency. However, as the DSN moves toward Ka-band (32-GHz) communications, the 0.1-dB beamwidth scan radii now are reduced to about 0.0016 and 0.0009 deg for the 34- and 70-m antennas, respectively. Alternatives to the current DSN conscan method (both by mechanical and electronic means) are being investigated. They are intended to relieve the antenna axis servomechanisms of these stringent Ka-band scanning requirements and to increase the bandwidth of the beam-switching capability.

In this article, a straightforward approach is presented to Ka-band vernier pointing of the 34-m DSS-13 beam-waveguide (BWG) antenna. The scheme is to rotate one of the reflector plates in the beam-waveguide optics path in order to scan the beam. It is conceivable that an ability to scan a beam quickly may allow for a single antenna to simultaneously support multiple spacecraft at two dis-

tinct, but neighboring positions in the sky. However, at this time, the DSN's motivation for scanning a beam with a BWG mirror lies principally in potential improvements in antenna-pointing capabilities. At DSS 13, a position control system has been implemented on the last flat-plate mirror (M6) specifically for demonstrating conscan at the Ka-band frequency.

This article covers the DSS 13 beam-waveguide mirror conscan project from conceptual design through initial beam-scan measurements made in December 1993. Beam-scan versus mirror tilt angle predictions, computed from a diffraction analysis for both the Ka- and X-band frequencies, are presented. From the radio frequency (RF) predicts, mechanical and servo position control requirements are then derived. The final mechanical positioner and servo control system designs, as implemented at DSS 13, are highlighted. Specific details of the designs are given in the appendices. Antenna-pointing measurements versus mirror tilt angle were made to verify the Ka-band beam-scanning predicts. Although preliminary, the initial results are in agreement with those predicted from the RF analysis. Plans for initial conscan experimentation are summarized.

II. Beam-Scanning RF Predicts

It is well understood that tilting a mirror in a beam-waveguide (BWG) antenna, like DSS 13, results in a scanned microwave beam. Indeed, the beam shift and resulting efficiency degradation of such a misaligned system can be shown to be equivalent to that of a paraboloidal reflector with an offset feed. This latter topology was analyzed many years ago by Ruze [1].

The use of the BWG mirror to scan the beam is very straightforward and robust. The physical principles underlying the beam-scanning mechanism come from simple geometrical optics. A diffraction analysis has been done using two JPL computer programs: Physical Optics (PO) and the Geometrical Theory of Diffraction (GTD). A single mirror can be used in conjunction with any feed package at every operational frequency to scan the beam.

At DSS 13, we wanted to use this scan capability with a conscan algorithm to estimate the antenna beam-pointing errors during Ka-band tracking operations. Here we prescribed the desired beam-scan angle associated with the conscan radius, which is a function of the beamwidth of the antenna (nominally 0.1 dB). A mirror tilt angle is then found. The narrower the antenna beamwidth, the smaller will be the desired scan angle and mirror tilt angle. Even though a scanned beam will always result in a degradation of the aperture efficiency of the antenna, it has been computed that, for X- and Ka- bands, the efficiency is reduced not more than 0.1 percent for the nominal scan angles of 0.1 dB. The curves in Fig. 1 show the relationship between the Ka-band scan angle of the electromagnetic beam of the antenna and the tilt angle of the conscan mirror at the two candidate positions (M1, M6). As shown, the Ka-band 0.1-dB beamwidth conscan radius of 0.0016 deg is achieved by tilting the M1 and M6 mirrors by 0.0375 and 0.076 deg, respectively. The reduced sensitivity of scanning the beam with either BWG reflector plate is a major advantage over direct scanning with the main antenna reflector. That is, a drastically smaller and lighter reflector surface can accomplish the desired beam scanning with a less precise positioning mechanism.

In Fig. 2, the X-band beam scan predicts are presented only for the M6 mirror. The 0.1-dB beamwidth conscan radius for X-band is 0.006 deg and the necessary mirror tilt to scan the beam is 0.365 deg. Because of its simplicity, this beam-scanning design may find application to still lower frequencies. The relevant parameters of the mechanical motion will be the limit for determining the highest and lowest frequencies for which the mirror system will be used. That is, the total stroke length of the actuator (i.e., the maximum rotation of the axes) will determine the system's lowest usable RF frequency, while the actuator position accuracy will determine its highest usable RF frequency.

III. Mirror Position Control System

A. Mirror Selection

We chose M6, the last flat-plate mirror, for demonstration at DSS 13. While M6 compares favorably to M1 by having a larger magnification factor in its mirror-rotation-to-beam-scan relationship, the main motivation for choosing M6 is the accommodating nature of its electromechanical servo positioner implementation. That is, the M6 mirror is easily accessible in the antenna pedestal room and installing a positioner at its location has minimal impact on the day-to-day station operations. The main advantage of a rotating mirror at the M1 position is that this single mirror could scan the beam of any feed in the pedestal room, much in the same manner as scanning the entire tipping structure. However, modification of the M1 mirror support structure involves taking the antenna down for a number of weeks. Due to the high subscription rate of DSS 13, the M1 mirror implementation is currently deemed unacceptable.

B. Mechanical System Requirements

A limited two-axis positioner is required for the M6 mirror above the third research and development feed position in the DSS-13 antenna pedestal room. We wished to scan the beam both at the Ka- and X-band frequencies for experimentation. The predicted X-band scan radius for the mirror is 0.365 deg; thus, the mirror positioner must have at least ±0.365 deg of angular travel in each axis of motion for operation at X-band. Although the Ka-band scan radius is roughly one-fourth that for X-band, additional mirror travel greater than the nominal 0.365 deg is desirable for studying the feasibility of implementing a true Ka-band vernier pointing system. That is, the mirror will be used for correcting pointing errors as well as estimating them, effectively closing the RF pointing loop about the mirror. We chose a pointing-correction margin of 0.008 deg. This is five times the Ka-band conscan radius (approximately 0.0016 deg for the 34-m antenna). The predicted M6 mirror tilt angle needed to scan the beam 0.0016 deg is 0.076 deg. The total Ka-band beam scan required for experimentation is then 0.0096 deg (the conscan radius plus the pointing-correction margin); this equates to a (predicted) angular travel of 0.46 deg needed for the M6 mirror.

Lastly, when not in use as a beam scanner, the mirror must function as a standard M6 mirror in the beam-waveguide optics path. This requires that it be located and stowed at an absolutely repeatable position corresponding to the nominal optical alignment position.

C. Servo Position Control Requirements

General requirements for the M6 mirror position servo are derived from the Ka-band beam-scanning predicts. The pointing accuracy requirement is obtained from the predicted M6 tilt angle of 0.076 deg needed for conscan at Ka-band. The mirror should be able to scan the beam to at least one-tenth of this angle, or to a 0.0076-deg mirror line-of-sight position. This requirement is then broken down to a 0.005-deg upper bound for position errors in each orthogonal axis of the mirror positioner. The tracking rates required for conscan are typically slow. In the DSN, the period for the circular beam scan ranges from 32 to 128 sec. While mechanical considerations may limit scanning the main antenna reflector with periods less than 32 sec, high-speed motion of a lightweight mirror gimbal assembly should be relatively easy to achieve with appropriate actuator selection. At a minimum, the mirror positioner must be able to circular scan the beam with a period of 10 sec for experimentation.

D. Mechanical System Design Description

Initially, a simple two-axes limited-motion positioner driven by linear actuators was proposed. The proposed design incorporated a mirror that rotated about the axes behind the front surface of the mirror. The resulting offset produced an undesirable parallax error. Subsequently, the design was changed and the bearings were moved outboard of the mirror, allowing their axes to intersect with each other, the front surface of the mirror, and the RF axis. Although the revised approach results in a more complex design, the utility of the positioner for other (future) applications is enhanced by elimination of the parallax error.

The final DSS-13 configuration is illustrated in Figs. 3 and 4. The mechanical design includes three major weldments: the inner gimbal, outer gimbal, and support stand; other parts are the mirror, the stepper motor linear-driven actuators, and the stow links. With the selected actuators installed, the mechanical axes can travel ±2 deg in the inner axis and ± 3.5 deg about the outer axis. This exceeds the operational requirements for conscan experimentation at Ka- and X-band beam scanning by a factor of four. The maximum slew rates of the mirror positioner are 6.6 and 11.5 deg per sec for the inner and outer axes, respectively. This exceeds the requirement to circular scan the beam within a period of 10 sec.

Built into the mirror structure, alongside the actuator mounts, are attachments in which stow links can be inserted. These rigid members attach via zero-clearance expandable locking pins between the rotating elements of the mechanism. Use of the zero-clearance pins allows the positioner to return to the exact position each time it is locked, so it can be used as a fixed mirror station when required. Further details of the mechanical design and hardware are given in Appendix A.

E. Servo Position Control System Design Description

Figure 5 shows a block diagram of the two-axis mirror position control system implemented at DSS 13. The inner and outer gimbal axes of rotation are denoted as the X and Y axes, respectively. The drive electronics (drive amplifiers/motor controllers, limit switches, etc.) are mounted near the mirror structure on the pedestal room ceiling, then interfaced down to a rack-mounted local control panel. Slew control of each mirror axis is available at the panel, along with a display of linear positions from linear variable differential transformers (LVDTs) mounted directly on the axis actuators. The mirror can be locally driven to the nominal alignment position (zero LVDT reading) before insertion of the stow links.

Remote position control from the DSS-13 control room is established through an RS232 interface. A rack-mounted 486 PC in the remote control room serves as the mirror position command generator/controller and angle display. Mirror position control can be accomplished either with feedback (from the LVDTs) or in an open-loop fashion (by counting motor steps). As discussed in Appendix B, both modes of operation offer more mirror position control accuracy than needed to meet the requirement of 0.005 deg. Further details of the servo position control system design and hardware are given in Appendix B.

IV. Initial Beam-Scanning Measurements

Pointing measurements were made at DSS 13 in December 1993 to validate the Ka-band beam-scanning predicts. Both planetary and broadband radio noise sources were tracked with the Ka-band array feed system installed at the F3 focus below the M6 mirror positioner. With the mirror tilted to a prescribed angle, the antenna was boresighted [2] using total (noise) power measurements made on the center horn channel of the array feed. As described in Appendix C, estimates of beam scan are computed from the measured cross-elevation and elevation pointing errors. The results are presented in Fig. 6.

The predicted slope of the beam scan magnitude versus mirror tilt angle shown in Fig. 6 differs from that shown in Fig. 1 due to a different geometry of the array feed system. The original slope predict is 47.5 to 1, corresponding to the standard horn Ka-band feed geometry. This is scaled by the ratio of 38.5 to 33.5 to account for the shorter phase-center-to-M6 mirror distance of the array feed center horn. The new predict is 55 to 1, which compares well with the best-fit slope of 54 to 1 computed from the measurements. This result is preliminary since about 30 percent of outlier measurement points were encountered during the experiment, although they are not shown in Fig. 6. Effort is currently being directed toward gaining an understanding of the outliers in terms of the errors in the beam-scan validation process. These include pointing measurement uncertainties associated with the low signal-to-noise ratios of the available Ka-band noise sources that are tracked, and errors dealing with the fact that the antenna orientation is not constant during the boresighting process.

V. Summary and Future Applications

The design and implementation of an antenna beam-waveguide mirror position control system at the DSS-13 antenna has been presented. The primary motivation for implementing the positioner on the M6 flat-plate mirror of the beam-waveguide optics path is to demonstrate the conical scan angle-tracking method at the Ka-band operating frequency. Radio frequency beam-scan predicts versus mirror tilt angle for the M6 mirror were presented and discussed in the context of the conscan application. From both the X- and Ka-band predicts, mechanical and servo control requirements for the mirror positioner were derived. The final control system implemented at DSS 13 was highlighted with detailed mechanical and servo design descriptions given in the appendices. Initial beam-scan measurements made at the Ka-band frequency in December 1993 were presented. After reduction, the preliminary measurements were shown to be in agreement with the Ka-band predictions.

The next experiment involving the M6 mirror position control system will be the demonstration of conscan at Ka-band. The conical scan algorithm presented in [3] is currently being integrated into the PC mirror controller shown in Fig. 5. Angle generation for circular beam scanning will be computed in the control room and downloaded to each mirror's axis motor controller in the pedestal room. As shown in Fig. 5, the PC mirror controller will then be interfaced to the receiver subsystems in the DSS-13 control room. Planetary and broadband radio noise sources will be tracked to demonstrate conscan at Ka-band with system noise temperature measurements input from the station's total power radiometer subsystem. The initial conscan experimentation will utilize the array feed center horn and front end electronics. However, any (Ka-band or lower frequency) feed package may be installed underneath the M6 mirror positioner for conscan or beam-scan testing.

Further applications in antenna pointing for the M6 mirror are likely to focus on the ease of obtaining precision pointing of the small and stiff mirror gimbal assembly and the potentially high bandwidths achievable by small-scale axis servos. For example, it may be desirable to decouple the RF pointing error compensation from the antenna mechanical (azimuth and elevation) servo position loops. Here, beam-pointing error measurements obtained from either conscan or an electronic tracking error sensor (e.g., a monopulse tracking feed) installed beneath the M6 mirror can directly drive the mirror position controllers. Such a vernier pointing system will essentially close the RF beam tracking loop around the mirror itself. This will give a special performance advantage if the pointing error measurements can be accurately measured at a frequency greater than that of the antenna azimuth and elevation servo bandwidths (each currently set at 0.1 Hz at DSS 13).

Acknowledgments

The authors wish to thank Mike Thorburn for his many contributions to the project. In addition, many individuals assisted with this extensive first stage of the mirror beam-pointing control system design and implementation. Dan Bathker contributed to the initial vernier pointing system concept and application development. Alfonso Feria performed the structural analysis on the mirror positioner and Don Ohashi completed the drafting. Richard Thomas of PRC, Inc., fabricated, welded, assembled, and aligned the mirror positioner structure at Goldstone. Seyavosh Ghamari and Edwin Khachatourian of PRC led the servo drives cabling effort and assisted in the software implementation, respectively. Mike Britcliffe is acknowledged for helpful discussions on the M6 mirror alignment, and Caroline Racho assisted in the preliminary beam-scan measurements and analysis.

References

[1] J. Ruze, "Lateral-Feed Displacement in a Paraboloid," *IEEE Trans. Antennas and Propagation*, vol. AP-13, pp. 660–665, September 1965.

[2] L. S. Alvarez, "Analysis and Applications of a General Boresight Algorithm for the DSS 13 Beam-Waveguide Antenna," *The Telecommunications and Data Acquisition Progress Report 42-111, vol. July–September 1992*, Jet Propulsion Laboratory, Pasadena, California, pp. 48–61, November 15, 1992.

[3] L. S. Alvarez, "Analysis of Open-Loop Conical Scan Pointing Error and Variance Estimators," *The Telecommunications and Data Acquisition Progress Report 42-115, vol. July–September 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 81–90, November 15, 1993.

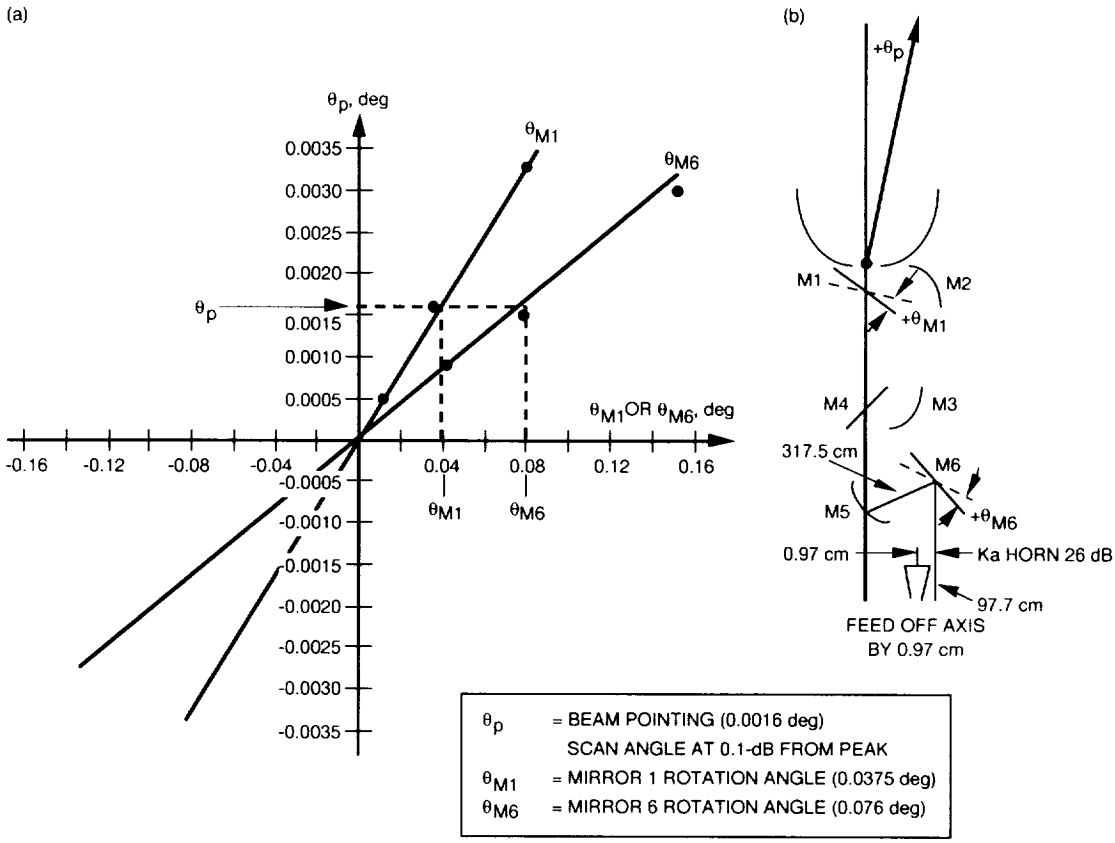


Fig. 1. RF predicts and corresponding feed geometry: (a) predicted Ka-band beam scan versus M1 and M6 mirror tilt and (b) geometry for a standard Ka-band feed at DSS 13.

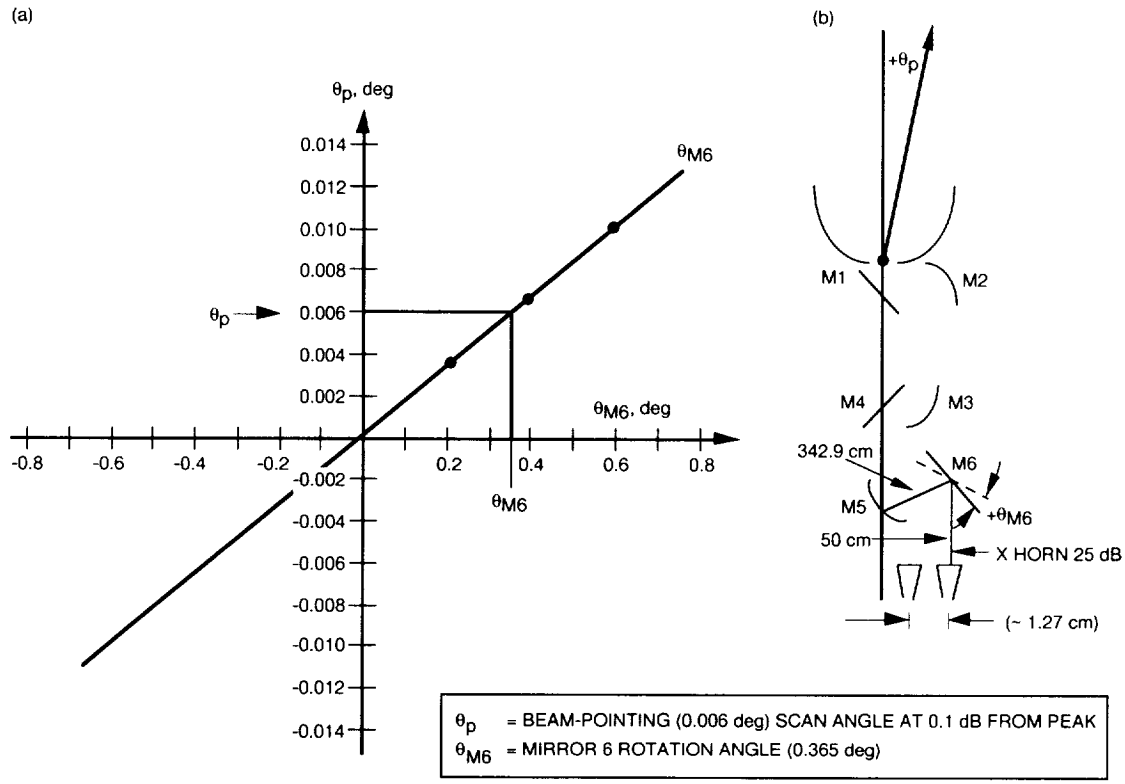


Fig. 2. RF predicts and corresponding feed geometry: (a) predicted X-band beam scan versus M6 mirror tilt and (b) geometry for a standard X-band feed at DSS 13.

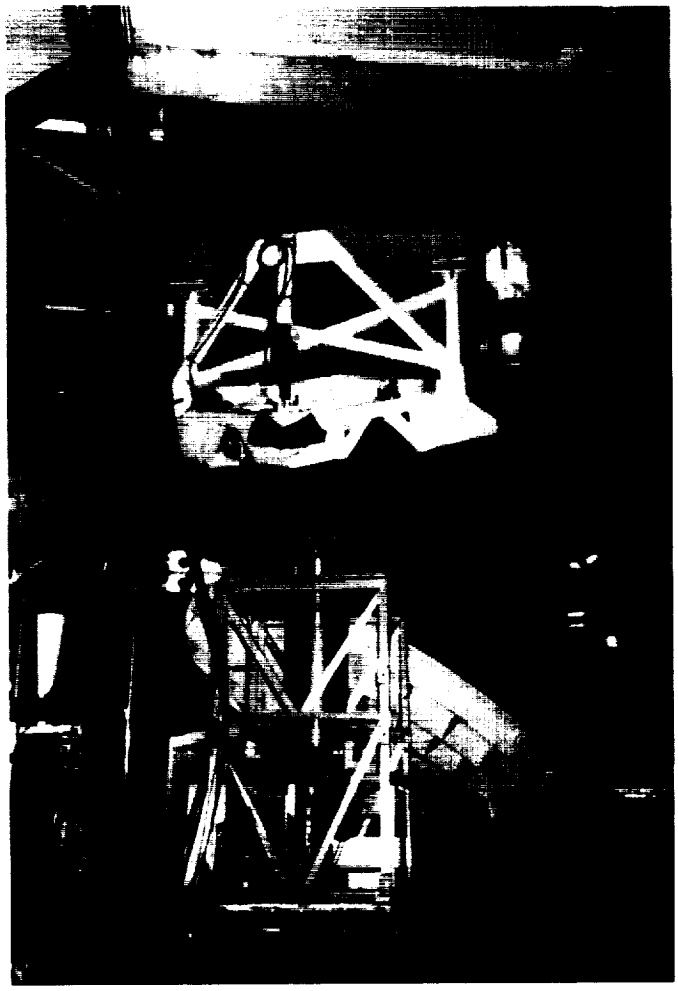


Fig. 3. The M6 mirror positioner at DSS 13, rear view.

Fig. 4. The M6 mirror positioner at DSS 13, side view.

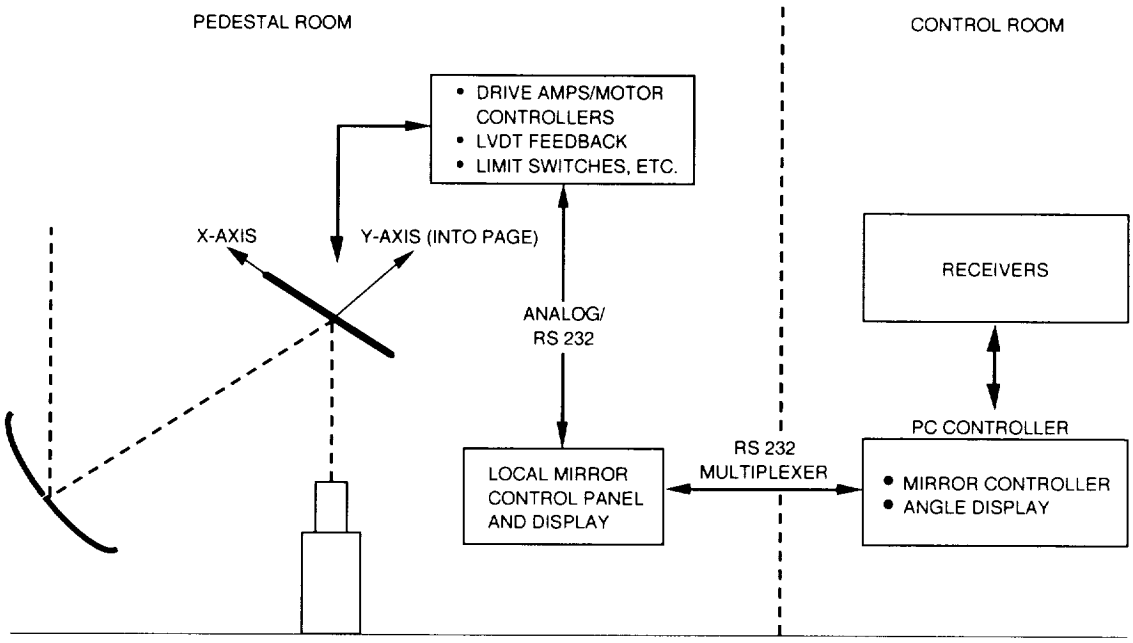


Fig. 5. Block diagram of the M6 mirror position control system at DSS 13.

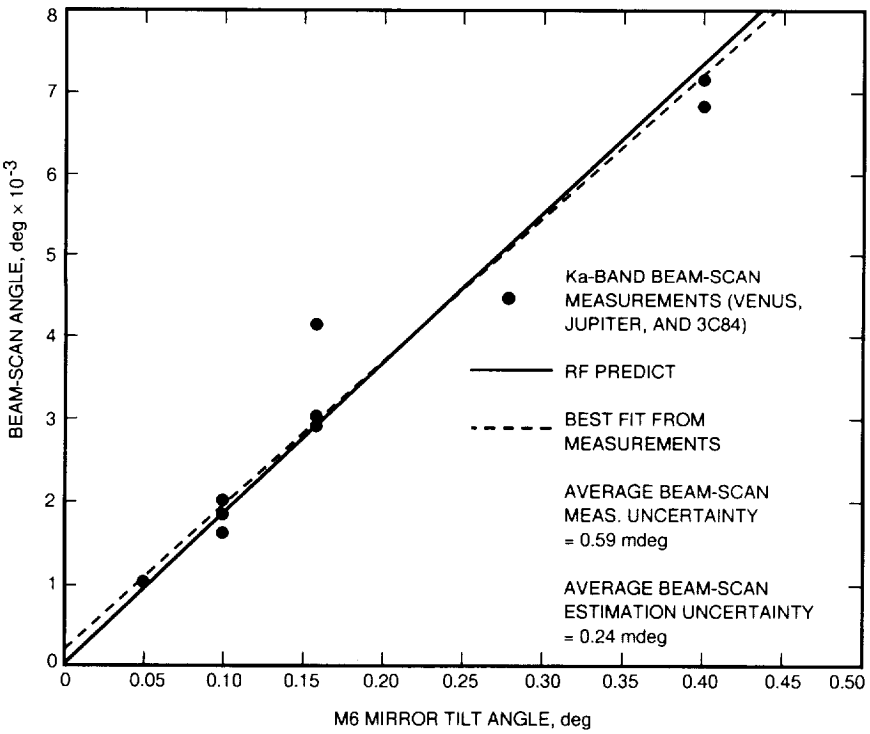


Fig. 6. Predicted and measured Ka-band beam scan versus M6 mirror tilt.

Appendix A

Mechanical Design Characteristics and Hardware Description

The overall mechanical design includes three major weldments: the inner gimbal, the outer gimbal, and support stand; other parts are the mirror, stow links, and drive actuators. These are described below.

I. Inner Gimbal

The inner gimbal is diamond shaped to provide a direct load path between the mirror and the ceiling. It is fabricated of 5- by 5- by 0.476-cm (2- by 2- by 3/16-in.) steel tubing. The inner gimbal supports the flat mirror in the center and rests within the perimeter of the outer gimbal. It attaches to the outer gimbal with bearings that are located at the top and bottom of the diamond-shaped weldments and provide the vertical axis of rotation. The bearings are shimmed for alignment with the positioner. There are also shimming provisions to adjust the mirror as required.

II. Outer Gimbal

The outer gimbal is also diamond shaped. It is fabricated of 7.6- by 7.6- by 0.635-cm (3- by 3- by 1/4-in.) steel tubing. The vertical axis bearings attach the inner gimbal to the outer gimbal at top and bottom. The horizontal axis bearings are attached at the other two corners of the outer gimbal weldment. They connect the outer gimbal to the support stand. There are shimming provisions at each of the four bearing sets that allow coplanar adjustment of the four bearings and four support yokes.

III. Support Stand

The support stand attaches to the ceiling via 7.6-cm (3-ft) plates. It also attaches to the horizontal axis bearings to the outer gimbal. There are shimming provisions to level the horizontal axis bearings.

IV. The M6 Mirror

A Blanchard ground aluminum casting was chosen over lighter more expensive alternatives. Neither weight nor inertia were considered since the positioner was designed with growth in mind and because it can position considerably greater weight than the mirror used. A dead load midspan deflection specification of 0.005 cm (0.002 in.) was established as a criterion in the modeling. A minimum 5-cm (2-in.) thickness of the aluminum plate was indicated to satisfy this requirement.

V. Stow Links

The stow links are rigid members that attach via zero-clearance locking pins between rotating elements of the mechanism. The stow links serve two purposes:

(1) They locate the mirror in a repeatable position, thus making it practical to use the equipment as a fixed mirror.

(2) In the event of a control system malfunction, the actuator loads can be transferred through the links without damage to the drives or the structure. The links are located immediately next to the actuators in order to minimize the load path.

Four locking pins are used, one at each end of each stow link. They are inserted into reamed holes and have zero clearance when installed and expanded.

VI. Drive Actuators

The actuator motors step 25,000 increments per revolution. Through the gear ratio and the ball nut, this results in 793,750 steps per centimeter (312,500 steps per inch) of linear motion. Since the actuators do not back drive, they require no brakes. A small slot is provided in the motor shaft to allow it to rotate.

The key feature of the actuators, designed and built by Industrial Drives Corporation, is that they have no more than 12.7-μm (0.0005-in.) backlash at the output. The actuators use tight gear meshes and two ball nuts loaded against each other to achieve this minimal backlash.

VII. Drive Characteristics

The main drive characteristics of the mirror positioner are summarized in Table A-1. A structural analysis of the positioner was also completed. It predicts that the first structural resonance is at 8.8 Hz.

Table A-1. Drive characteristics of the DSS-13 M6 mirror positioner.

Parameter	Inner gimbal	Outer gimbal
Torque	1179.5 J (870 ft-lbs)	677.9 J (500 ft-lbs)
Moment arm	50.5 cm (19.899 in.)	88.57 cm (34.879 in.)
Maximum velocity	6.6 deg/sec	11.5 deg/sec
Acceleration	1000 deg/sec/sec	7000 deg/sec/sec
Mechanical travel	±2 deg	±3.5 deg/sec

Appendix B

Servo Position Control Design Characteristics and Hardware Description

I. Drive Electronics

Actuation and low-level control of the linear drive motors is accomplished using the S5201 drive indexer motor controllers from Industrial Drives Corporation. Linear position commands issued in terms of motor steps are input through an RS232 interface. The microprocessor-based controllers also provide an interface for limit switches mounted on the actuators and for remote motor jogging actuated by push-button switches on the pedestal room floor control panel. Local and remote position feedback of each actuator is accomplished by Lucas-Schaevitz 500HR linear variable differential transformers (LVDTs). One of these standard product transformers is mounted directly to the body of each actuator. The total linear range of the sensor is ± 1.27 cm (± 0.5 in.) This is discussed in terms of angular travel below. In order to return the positioner to its nominal alignment location for use as a standard fixed M6 mirror, each transformer is nulled to zero when the stow links are installed and the mirror is in the fixed station position.

II. Direct Mirror Motion Control Drive Ratios

Linear control of the stepper motors is converted into angles through moment arms of 50.5 and 88.57 cm (19.899 and 34.879 in.) for the X and Y axes, respectively. As described earlier, each actuator provides 793,750 steps per centimeter (312,500 steps per inch) of travel. Using the small angle approximation $\tan(\theta) \approx \theta$, the number of steps-to-angle ratio needed for command generation is 108,532 and 190,235 steps per degree of motion of the X and Y axes, respectively. The position errors resulting from this angle approximation are negligible for the mirror tilt angles needed to conscan at the Ka- and X-band frequencies.

III. Position Feedback and Angle Accuracy

Position feedback can be obtained either by counting commanded steps to the stepper motors or from the LVDTs. The LVDTs are interfaced down from the ceiling into a two-channel Lucas-Schaevitz MP2000 amplifier/readout indicator. The readout indicator provides a linear motion display in the pedestal room. Each LVDT is zeroed at the nominal optical alignment position of the mirror, thus allowing local control of the mirror back to its stow position before manual insertion of the stow struts. Analog output corresponding to linear displacement for each axis is provided by the MP2000 in case a higher update rate of the position loops is desired. The DSS-13 mirror servo implementation obtains digital readouts from the MP2000 in the control room through an RS232 interface. The digital inputs for each axis are then converted to angles for terminal display and for comparison with the open-loop angle estimates computed from counting commanded motor steps.

Both the X- and Y- axis LVDTs are operated in a linear range of ± 1.27 cm (± 0.5 in.), corresponding to roughly ± 1.4 and ± 0.8 deg of travel for the X and Y axes. In this range, the chosen sensors are rated to be accurate to 2.54 to 0.254 μm (0.0001 to 0.00001 in.), which results in 0.0003 to 0.00003 deg of angular resolution for the (shorter moment arm) X axis. The number of motor steps needed to obtain the same range in resolution is from 32 down to 3. For extremely accurate position control applications, LVDT feedback is preferred since position errors caused by the motors skipping steps may become significant over time. A more significant error (when counting steps) may come from backlash at the actuator output. The maximum backlash for the actuators was stated to be 12.7 μm (0.0005 in.), or an equivalent angle error of 0.0014 deg for the X-axis. The position accuracy requirement for the beam-waveguide conscan mirror is only 0.005 deg for each axis of rotation. Thus, even under the worst-case position error scenario, pointing errors resulting from open-loop control of the stepper motors are still below the requirement.

Appendix C

Measurement Equations for the M6 Mirror

I. DSS-13 Pointing Derotation Algorithm

The geometry of the DSS-13 beam-waveguide antenna is shown in Fig. 1. During normal tracking operations, the mirror is held fixed in the pedestal room while the antenna moves independently relative to the mirror. Therefore, while tracking, fixed rotations of the mirror will not result in fixed beam-scan offsets relative to the antenna aperture. A derotation algorithm has been developed by Cramer[1] and validated at DSS 13 utilizing the array feed system. The algorithm predicts the relationship between a beam offset in the antenna aperture coordinate system and its associated phase center location in the feed position in the antenna pedestal room as a function of the antenna main reflector surface azimuth and elevation angles. The general derotation algorithm is

$$\Delta xel = -RB\cos(\phi) \tag{C-1}$$

$$\Delta el = RB\sin(\phi) \tag{C-2}$$

$$\phi = AZ - EL - \tau + \omega \tag{C-3}$$

where Δxel and Δel define a beam offset in the antenna aperture coordinate system, the radial distance R and angle ω define the phase center (in the pedestal room) associated with the offset beam given by Δxel and Δel, B is the beam deviation factor, AZ and EL are the antenna azimuth and elevation angles, and τ defines the angular position of the of the center of the focal plane in the pedestal room measured clockwise from true north.

II. Beam-Scan Measurement Equations

The Ka-band beam-scan predicts for small angles of rotation of the M6 mirror are given in Fig. 1. The predicts are validated by boresighting [2] the DSS-13 antenna on natural noise sources while the mirror is offset from the nominal optical alignment position. Let the beam-scan angle in Fig. 1 be denoted as Θ. It is related to the antenna aperture beam coordinates by the following expression:

$$\Theta^2 = \Delta xel^2 + \Delta el^2 \tag{C-4}$$

From the derotation algorithm above, this implies

$$\Theta = RB \tag{C-5}$$

During the tracking experiment, errors in cross-elevation and elevation are measured and then logged with the mirror tilt angle and the antenna position angles. Equation (C-4) is then used to estimate the magnitude of the beam scan. From Eqs. (C-1) and (C-2), estimates of the measured beam-scan angle can also be computed by

$$\Theta = -\frac{\Delta xel}{\cos(\phi)} \tag{C-6}$$

and

$$\Theta = \frac{\Delta el}{\sin(\phi)} \tag{C-7}$$

where ϕ is given by Eq. (C-3) and is dependent on antenna orientation.

[1] P. Cramer, "Tests of the Pointing Derotation Algorithm for DSS 13," JPL Interoffice Memorandum 3328-93-0037 (internal document), Jet Propulsion Laboratory, Pasadena, California, June 25, 1993.

N94-35549

TDA Progress Report 42-117

May 15, 1994

Use of the Sampling Theorem to Speed Up Near-Field Physical Optics Scattering Calculations

P. W. Cramer and W. A. Imbriale
Ground Antennas and Facilities Engineering Section

Physical optics scattering calculations performed on the DSN 34-m beam-waveguide antennas at Ka-band (32 GHz) require approximately 12 hr of central processing unit time on a Cray Y-MP2 computer. This is excessive in terms of resource utilization and turnaround time. Typically, the calculations involve five surfaces, and the calculations are done two surfaces at a time. The sampling theorem is used to reduce the number of current values that must be calculated over the second surface by performing a physical optics integration over the first surface. The additional number of current values required on the second surface by subsequent physical optics integrations is obtained by interpolation over the original current values. Time improvements on the order of a factor of 2 to 4 were obtained for typical scattering pairs.

I. Introduction

The technique discussed here was developed to reduce the amount of time required by physical optics to compute the antenna patterns of the 34-m beam-waveguide antennas at the DSN. Figure 1 illustrates a typical DSN beam-waveguide (BWG) antenna. As can be seen, the antenna has eight scattering surfaces. Beginning with the feed system in the pedestal room, there is a flat mirror, an ellipse, a second flat mirror, two parabolas, a third flat mirror, the subreflector and the main reflector. However, the flat mirrors are assumed not to affect the antenna pattern properties and are ignored in the analysis, leaving five scattering surfaces. With this number of mirrors, the analysis must be repeated four times before the final far fields are evaluated. For analysis up to X-band (8.45 GHz), the available computers could easily handle calculations of such size and complexity. However, with the shift to Ka-band (32 GHz) to support future deep-space missions, computational times are increased by a factor of about 16. Computational times of 12 hr on a Cray Y-MP2 single processor computer are typical.

This article presents a method to reduce the overall time by a factor of 4 or more for a typical pair of scattering surfaces and by a factor of 2 for the overall antenna system. The sampling theorem coupled with a near-field radial interpolation algorithm is used to speed up the physical optics calculations.

The sampling theorem has been used previously for the far-field analysis of reflecting surfaces [1]. A sampling-like technique [2] that allows the use of the Fast Fourier Transform (FFT) algorithm has been used to calculate the far fields of a parabolic reflector and the Fresnel zone fields of a planar aperture. Both methods [1] and [2] were extended to the near field [3,4], but were limited to calculating the fields on a spherical surface of constant radius.

Evaluating a BWG system requires multiple near-field calculations on arbitrary surfaces. To overcome the existing limitations, the sampling technique has been generalized for *arbitrary* field point calculations in the near field. Also, since evaluations on multiple surfaces are required, a technique is outlined for developing an equivalent source aperture that defines the geometry required to calculate the optimum sampling parameters.

II. Method

The basic method used to analyze the 34-m antennas consisted of performing a physical optics integration over the currents on the various surfaces of the antenna. The form of the physical optics program used here [5] is based upon a discrete approximation of the radiation integral. In this approach, the reflector surface is replaced by a triangular facet representation such that the surface resembles a geodesic dome. The physical optics currents are assumed to be constant in amplitude and phase over each of the facets so the radiation integral is reduced to a simple summation of the contributions of each facet.

To evaluate the complete antenna, an integration is performed over the currents on the first scattering surface to get the currents on the second surface. Using these new currents on the second surface, the process is repeated for the next surface, continuing the utilization of pairs of surfaces until the complete antenna has been analyzed. The final integration over the main reflector uses the Jacobi-Bessel form of physical optics [6], which is much faster for calculating far-field patterns from large reflectors, but unfortunately not amenable to the use of the sampling theorem. With the exception of the main reflector, if each surface is considered to be of comparable size, and if the required current resolution in any direction is $\mathcal{N}$, then $\mathcal{N}^2$ physical optics integrations are required over the first surface for each of the required $\mathcal{N}^2$ current points on the second scattering surface. This implies $\mathcal{N}^4$ calculations and is the real driver for the computational time.

Assuming first that the number of surface current values that must be calculated using physical optics on the second surface can be reduced significantly, and second that the $\mathcal{N}^2$ current values needed to perform the subsequent physical optics integration can be obtained by interpolating over the reduced surface current set, then the computational time will approach that of $\mathcal{N}^2$ operations on the first surface. The physical optics integral is composed of two basic parts, the current term and the kernel or exponential term. The current term is typically a slowly varying function of position, while the kernel varies rapidly as a function of position and observation point. The approach is to employ the sampling theorem to calculate the number of surface current values necessary to define the surface currents on the second surface, and then to use an interpolation algorithm to obtain the larger number of points required by the rapidly varying, but easily evaluated kernel.

A key problem is to define a field sampling function that could be used to determine the sampling frequency. Patterns produced by a uniform distribution have the narrowest beamwidths. Any deviations from a uniform distribution broadens the pattern shape or beamwidth. Therefore, the pattern derived from a uniform aperture distribution should have the highest frequency content and should provide a conservative estimator for the maximum sampling frequency. Figure 2 illustrates the far-field pattern distribution that results from a uniformly illuminated square aperture and is defined by the following function in the u direction:

$$E(u) = \frac{\sin(u)}{(u)}$$

where

$$u = \frac{2\pi X_m \sin\theta}{\lambda}$$

X_m = center to edge dimension of a square source aperture in the x direction

θ = angle to a field point on the sampling surface

λ = wavelength

If this distribution is evaluated on the sampling surface, then the distance from the surface center to its edge in (u, v) space is

$$u_m = \frac{2\pi X_m \sin\theta_m}{\lambda}$$

where θ_m = angle to the edge of the sampling surface. $E(v)$ in the orthogonal direction can be defined in a similar manner.

The distribution frequency is

$$B = \frac{u_m}{u_d}$$

where $u_d = 2\pi$ is the distribution period.

Since the sampling theorem requires sampling at twice the highest frequency and includes sampling points over the full width of the sampling surface, the number of sampling points required is

$$N = 2(2B) + 1$$

or

$$N = \frac{4X_m \sin \theta_m}{\lambda} + 1$$

The $\sin(u)/(u)$ function is based on a far-field derivation. Although it does not provide a rigorous basis for estimating the sampling frequency for sampling surfaces in the near field, it still gives a good estimate for typical source aperture fields. Also, the fields on the sampling surface are not a strictly band-limited function. To account for these limitations, an 18-percent oversampling was used, an approach that provides sufficient accuracy (as will be shown later).

Sinc functions are used to do the interpolations and to evaluate the fields at any point on a sampling surface; thus,

$$H_{(u)} = \sum_{n=-N/2}^{N/2} H_{(n)} \frac{\sin(u - n\pi)}{(u - n\pi)}$$

Since the $\sin(u)/(u)$ field function is defined on a spherical surface, the sampling must also be done on a spherical surface. Also, the spherical surface origin must be located at the center of the source aperture. In addition, the origin of the spherical surface must also be at the field function phase center so as to minimize the phase variations over the spherical surface. This implies that the field function phase center must also coincide with the source aperture origin. In general there are three problems: First, the reflector surfaces usually are not spherical; second, even if the reflector surface were spherical, its origin may not be located at the origin of the source aperture; third, the center of the source aperture may not be the phase center of the scattered fields. To accommodate nonspherical surfaces or spherical surfaces with offset phase centers, the sampling surface of interest is enclosed by two spherical surfaces, with the origins of the two surfaces at the phase center of the scattered fields. Figure 3 illustrates the geometry. In the figure are five surfaces: In addition to the physical source aperture (subreflector), and the ultimate sampling surface (main reflector), there are the two surfaces enclosing the ultimate sampling surface and an equivalent source aperture. If the scattered field phase center does not coincide with the center of the physical source aperture, an equivalent source aperture is constructed at the phase center.

To determine where the phase center should lie, a subset of the scattered field phase pattern is calculated on a spherical surface constructed midway between the two initial spherical reflectors, which have their origins at the center of the physical source aperture. A phase center location is computed that minimizes the phase pattern variations, in a least-squares sense, over the spherical surface. A discussion of this technique is beyond the scope of this article; however, W. Rusch and P. Potter [7] describe a two-dimensional technique that is the basis of the three-dimensional technique used here. The controlling equations are summarized here (see Fig. 4), where R is the radius of the spherical reflector. The objective is to minimize the following function:

$$\sigma = \Sigma w_i \left(kd \cos \gamma_i + c - \Phi_i - \overline{\Delta}\right)^2$$

with minimization conditions:

$$\frac{\delta \sigma}{\delta x_d} = 0$$

$$\frac{\delta \sigma}{\delta y_d} = 0$$

$$\frac{\delta \sigma}{\delta z_d} = 0$$

where

w_i = phase weight

k = propagation constant

d = (x_d, y_d, z_d) computed vector determining phase center location

γ_i = direction to field phase point

c = residual phase of pattern relative to computed phase center location

Φ_i = phase pattern on spherical surface relative to original pattern origin

$\overline{\Delta}$ = mean of $kd \cos \gamma_i + c - \Phi_i$

The phase center algorithm is based on a far-field approximation, i.e., the adjustment to the phase center location must be small compared to the size and radius of the spherical surface on which the phase is evaluated. This limitation is overcome by iterating the algorithm until the last estimate of the required adjustment to the phase center position is smaller than some specified value.

The equivalent aperture size is estimated to produce a field distribution on the sampling surface similar to that produced by the physical source aperture. Referring to Fig. 3, the equivalent aperture size is

$$X_m = \frac{S_f X_e \sin \theta_e}{\sin \theta_m}$$

where

S_f = oversampling parameter

X_m = center-to-edge dimension of the equivalent source aperture in the x direction

X_e = center-to-edge dimension of the physical source aperture (subreflector)

θ_m = angle to the edge of the sampling surface for the equivalent source aperture

θ_e = angle to the edge of the sampling surface for the physical source aperture

Since a square aperture is used, $Y_m = X_m$. This size is used in the calculation of the number of required sample points N.

After the determination of the sampling parameters and geometry and before any interpolations can be performed, the fields H_n must be computed on a square grid over the two spherical surfaces, the size of the grids being determined by the number of sampling points N. The overall interpolation is performed as follows. A radius from the center of the equivalent source aperture (phase center) is constructed to an interpolation point on the ultimate sampling surface of interest and then made to intercept the two spherical surfaces. Polar interpolated fields are computed on the two spherical surfaces at the intersection points of the radius. The polar interpolation at each of the spherical surfaces is expressed as follows:

$$H_{(u,v)} = \sum_{n_u=-N_u/2}^{N_u/2} \sum_{n_v=-N_v/2}^{N_v/2} H_{(n_u,n_v)} \times \frac{\sin(u - n_u\pi)}{(u - n_u\pi)} \frac{\sin(v - n_v\pi)}{(v - n_v\pi)}$$

where

$$u = \frac{2\pi x X_m}{\lambda r}$$

$$v = \frac{2\pi y Y_m}{\lambda r}$$

x = x coordinate of point on surface of interest

y = y coordinate of point on surface of interest

r = radial distance to point on surface of interest

$H_{(u,v)}$ is evaluated at r_1 and r_2 on the two surfaces and the results are defined as H_1 and H_2, respectively. Next, a radial interpolation is performed between these two points to obtain the interpolated field point on the sampling surface of interest. Since a near-field interpolation is required, terms of the order $1/r$ and $1/r^2$ are used. The appropriate equations are as follows:

$$H = \frac{e^{-jkr}}{4\pi r}\left(A_0 + \frac{A_1}{r}\right)$$

where

$$A_0 = \frac{4\pi}{(r_1 - r_2)}\left(r_1^2 H_1 e^{-jkr_1} - r_2^2 H_2 e^{-jkr_2}\right)$$

$$A_1 = \frac{4\pi r_1 r_2}{(r_2 - r_1)}\left(r_1 H_1 e^{-jkr_1} - r_2 H_2 e^{-jkr_2}\right)$$

This process is repeated until all the fields at all of the required points on the sampling surface have been calculated, then the associated currents that are required for the subsequent physical optics calculations can be calculated.

III. Program Design

This section briefly describes the changes required to an existing physical optics program to implement the sampling capability. The existing program is referred to as a POPO program in that it contains a cascaded design that allows the calculation of the scattered pattern from a two-reflector system. In effect, the currents on the first surface are either calculated from a feed pattern or imported from a previous scattering calculation. Then a physical optics integration is performed over the first surface to obtain the currents on a second surface. At this point, two options are available: Output the currents from the second surface to a file so that they might be imported into a second scattering calculation and/or perform a physical optics integration over the second surface to obtain the output scattered fields for the dual-reflector system. Figures 5(a) and 5(b) illustrate the program design. Figure 5(b) shows the lower level program routines for completeness and will not be discussed further.

Figure 5(a) shows three major blocks as follows: (1) the sampling surface (main reflector) routines, (2) the source aperture (subreflector) routines, and (3) the modules required to implement the sampling capability. The subreflector routines represent the first surface and are used to perform a physical optics integration on the first surface for points requested from the main reflector routines, which represent the second surface. The main reflector routines are then used to perform a physical optics integration over the second surface and to calculate the final scattered patterns. In Fig. 5(a), a dotted line labeled "original interconnection" shows the program in its unmodified form. To implement the sampling theorem mode, the only changes required to the original program were (1) break the link between routines INFLDM and SUBFLD and then insert the block of sampling routines, (2) change INFLDM to call sampling routine SMPFLD instead of SUBFLD, (3) change the main reflector routine DATAIN to read in the data parameters that the sampling routines require, and (4) change the main reflector routine DATAO1 to print out the data parameters associated with the sampling routines. Only simple changes to three routines were required. These routines are enclosed by dotted boxes.

Figure 6 shows a sample input data file for the POPO scattering program. The parameters in the box are the only changes required to the data file to support the sampling capability. The parameter SFACT allows changing the oversampling parameter from the default value of 18 percent to some other value. TFACT allows extending the sampling surface beyond its physical boundaries. XD,YD, and ZD specify the estimated location of the pattern phase center. SUBSW indicates whether the phase center location is defined in main reflector coordinates or subreflector coordinates. PHASW indicates whether the program is to compute an optimum phase center.

This discussion shows the simplicity of the approach and the ease with which it can be integrated into dual-surface physical optics scattering programs.

IV. Results

Figures 7 and 8 show the accuracy of the sampling approach. In the center of each figure, the geometry of the test case is illustrated to include a pair of parabolic mirrors such as used on a typical 34-m beam-waveguide antenna. The first parabola is the source aperture. The second parabola is the sampling surface on which a reduced set of fields are calculated and then interpolated to obtain the total set of fields and hence the currents required by physical optics to calculate the scattered field from the sampling surface. Figure 7 shows two curves describing the fields calculated on the sampling surface, one curve for the fields calculated in the normal manner and a second curve for the case where interpolation is used with a sparse set of sample points. There is a small amount of ripple between the two curves, less than 0.5 dB, but as seen in the Fig. 8, it has a negligible effect on the far fields calculated from the currents associated with the fields on the sampling surface. The curves shown in Fig. 8 are for the far fields calculated by performing a physical optics integration over the currents on the second parabola (sampling surface). One curve uses currents calculated using the sampling theorem and the other curve is based on the normal method where the currents are computed using physical optics integration for all the current points. As can be seen, the two curves are essentially identical over 40 dB. The differences are primarily in the side-lobe region. However, the side-lobe regions do not illuminate the subsequent scattering surfaces and therefore are of no interest in this particular application.

Figure 9 shows a combination of an elliptical source aperture and a parabolic sampling surface. Again, this is typical of a set of mirrors used in a beam-waveguide antenna. This case is different from the previous example in that the location of the phase center was known in the previous case, while in this case it was not. The source of the fields for the ellipse was from a feed with a transverse offset of approximately 8 wavelengths from the nominal focus, making it difficult to predict where the phase center would be for fields scattered from the ellipse. The program

was allowed to compute the location of the phase center starting with an estimated phase center that would be appropriate for a feed at the focus of the ellipse. The phase center had to be shifted by approximately 16 wavelengths. Except for one glitch, there is good agreement over 35 dB. The glitch represents a 1.5-dB error at a point 20 dB below the pattern peak. Some of the error could possibly be attributed to the difficulty of reproducing a pattern that contained some aberrations due to an offset feed.

An investigation was made to determine the effect of the size of the oversampling parameter on the accuracy of the computed scattered patterns. Figure 10 illustrates the effect of the oversampling parameter for a combination of an ellipse and a parabola. The upper 28 dB of the pattern was truncated to give more resolution to the area most affected by the oversampling parameter. As can be seen, a value of 1.18 is sufficient for a dynamic range of 37 dB (28 + 9). Except for the error at the 37-dB relative level, a value of 1.6 followed the main lobe down to at least the 58-dB relative level. As can be seen, larger oversampling values did not help in the side-lobe region. For most applications, performance below 30 dB is not required and the default value of 1.18 should be adequate. It is recommended that convergence tests be run for each general application or class to determine if the default value is adequate, or if a value of the oversampling parameter smaller than the default is desired. It must be pointed out that the computation time is strongly influenced by the size of the oversampling parameter.

The effect of increasing the sampling surface beyond the limits of the actual surface was investigated to see if this would improve the sampling accuracy, especially by better modeling the fields at the edge of the sampling surface. Figure 11 illustrates the case for an ellipse and parabola combination. A factor TFACT was defined and is a ratio of the angle subtended by a extended sampling surface to the angle subtended by the actual sampling surface. Thus TFACT = 1.0 refers to a sampling surface with no change. Except for some slight changes in the side-lobe region, it was found that increasing the sampling surface size had very little effect on pattern accuracy. This parameter did not show much potential and was not pursued further. Since increasing the sampling surface size strongly affects the computation time, a value of 1.0 should be used.

Figure 12 is a summary of the time improvement that was obtained for a calculation on a 34-m beam-waveguide antenna at Ka-band. Included in Fig. 12 is a schematic diagram of the 34-m BWG antenna with the three flat mirrors removed, which is the geometry analyzed and summarized in the figure. The appearance of the ellipse in front of the main reflector has no physical significance, but is what happens when the flat mirrors are removed. The results are shown by mirror pairs, the first mirror being the source mirror and the second mirror being the sampling mirror. The difference in time between the first two cases is easily accounted for. The sampling frequency is based on the size of the source aperture and the subtended angle produced by the sampling surface relative to the source aperture. In the second case the two mirrors are closer together than in the first case, increasing the subtended angle and in turn requiring a higher sampling frequency. See the equation for calculation of N. The improvements ranged between 1.69 and 4.39. The overall improvement up to and including the subreflector is a factor of 2.73. The sampling theorem was not applied to the main reflector calculation, so an improvement factor of 1.0 was assigned. Including the main reflector, a net improvement of 2.05 was obtained, reducing the computation time from 11.55 hr to 5.64 hr.

References

[1] O. M. Bucci, G. Franceschetti, and C. D'Elia, "Fast Analysis of Large Antennas: A New Computational Philosophy," *IEEE Trans. on Antennas and Propagation*, vol. AP-28, no. 3, pp. 306–310, May 1980.

[2] O. M. Bucci, G. Franceschetti, and R. Pierri, "Reflector Antennas Fields: An Exact Aperture-Like Approach," *IEEE Trans. on Antennas and Propagation*, vol. AP-29, no. 4, pp. 580–586, July 1981.

[3] G. D'Elia, G. Leone, R. Pierri, and D. Valentino, "Numerical Evaluation of the Near Field Using Sampling Expansions," *1982 APS Symposium*, vol. 1, pp. 241–244, June 1982.

[4] F. Benici, G. D'Elia, and R. Pierri, "Numerical Evaluation of Fresnel-Zone Fields by Sampling Like Technique," *1982 APS Symposium*, vol. 2, pp. 515–518, June 1982.

[5] W. A. Imbriale and R. Hodges, "Linear Phase Approximation in the Triangular Facet Near-Field Physical Optics Computer Program," *Applied Computational Electromagnetics Society Journal*, vol. 6, no. 2, pp. 74–85, Winter 1991.

[6] Y. Rahmat-Samii and V. Galindo-Israel, "Shaped Reflector Antenna Analysis Using the Jacobi-Bessel Series," *IEEE Trans. on Antennas and Propagation*, vol. AP-28, pp. 425–435, July 1980.

[7] W. V. T. Rusch and P. D. Potter, *Analysis of Reflector Antennas*, New York: Academic Press, pp. 145–153, 1970.

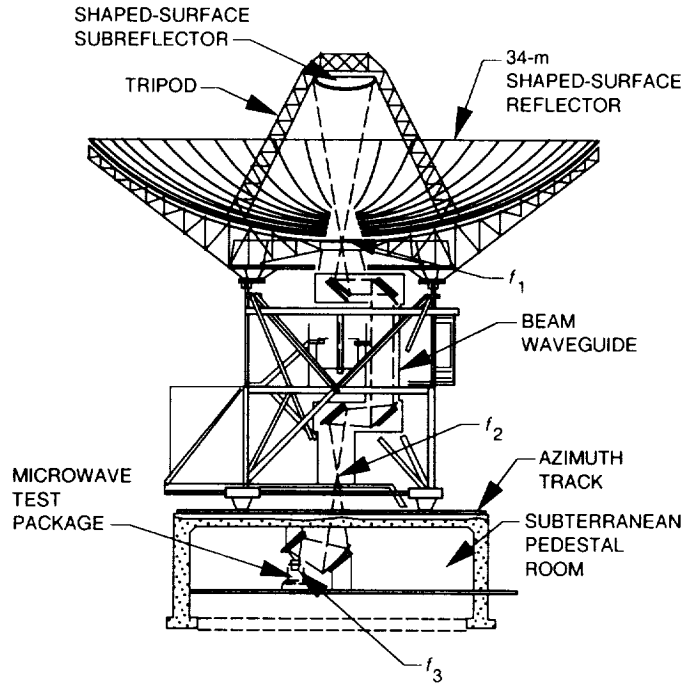


Fig. 1. The 34-meter BWG antenna.

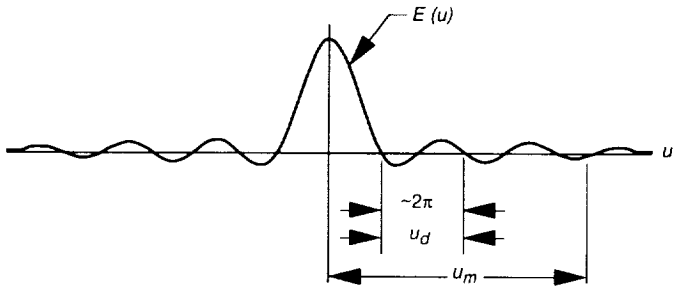


Fig. 2. Application of the sampling theorem.

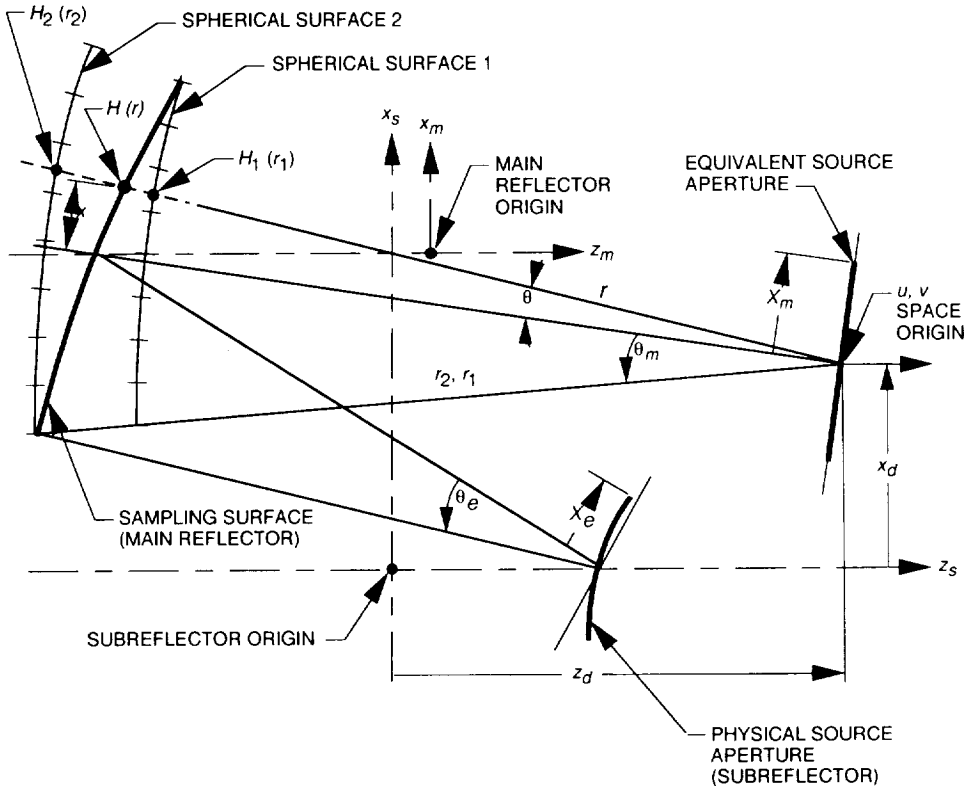


Fig. 3. Interpolation geometry.

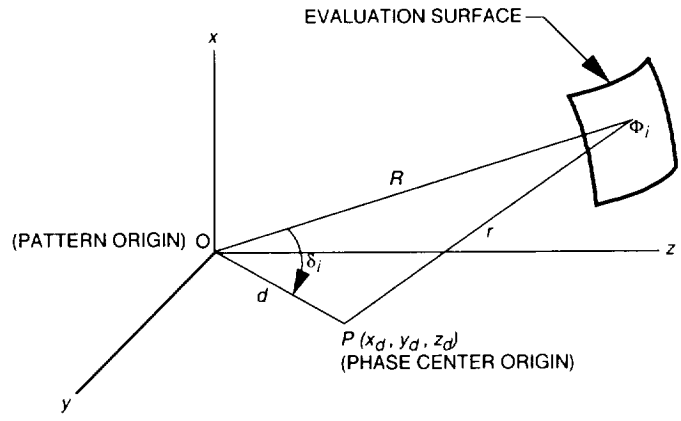


Fig. 4. The phase center algorithm.

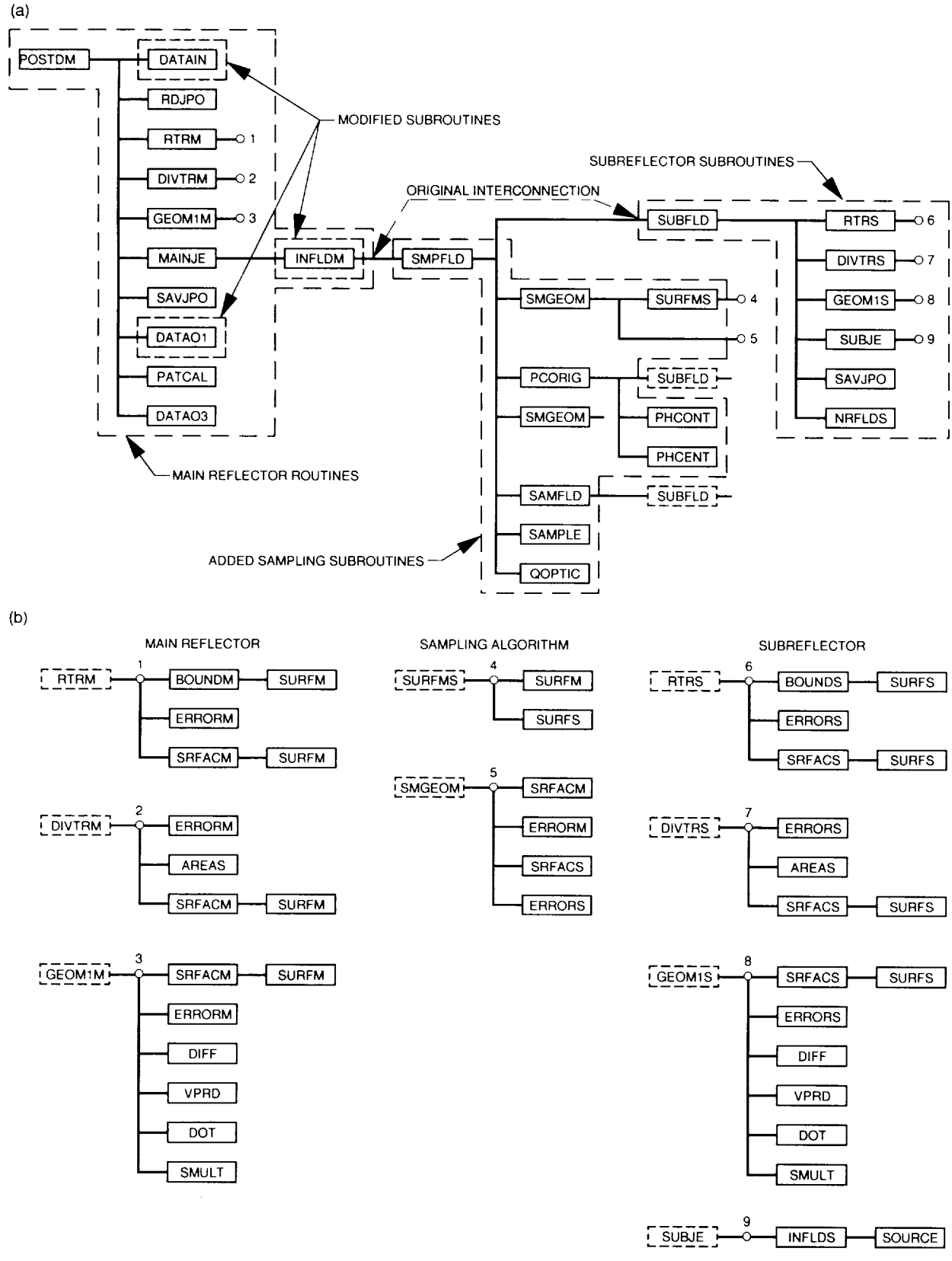


Fig. 5. Organization of the dual-reflector physical optics scattering program with the sampling algorithm: (a) main routines and (b) lower level routines.

SAMPLE DATA FILE

FREQ	UNITS						
32.0	METERS						
XS	YS	ZS					
-142.894190	0.0	-217.50					
ALPHAS	BETAS	GAMMAS					
90.0	120.0	90.0					
SCALEM							
1.							
MFILE							
dummy.dat							
NDXM	NDYM	XSYMM	YSYMM				
270	270	T	F				
IFLAGM	SIZEM						
-1	1.0						
READJE	SAVEJE						
f	t						
XF	YF	ZF					
0.00000	1.75000	0.3710					
ALPHAF	BETAF	GAMMAF					
0.0	0.0	0.0					
SCALES	SFACT	TFACT	XD	YD	ZD	SUBSW	PHASW
1.	0.0	0.0	0.0	0.0	9.0	F	T
SFILE							
dummy.dat							
NDXS	NDYS	XSYMS	YSYMS				
340	340	T	F				
IFLAGS	SIZES						
-1	1.0						
CIRPOL	POLX	LHCP					
T	F	T					
GAIN							
0.0							
HFILE							
f3-22-32.spw							
RR							
360.0							
PHII	PHIF	DPHI					
0.0	90.1	90.0					
THETAI	THETAF	DTHETA					
-16.0	16.1	0.5					
THTBPD	PHIBPD						
90.	0.						
PHSXC	PHSYC	PHSZC					
0.0	0.0	260.0					
PATNRM							
GAIN							
POLARZ							
SPHERICAL							
HFIELD	FRSNEL	ROTATE					
F	F	T					
PRTPAT	PLTPAT	PLTPHS	WRIT27				
F	F	F	T				
HFMAGN(1)	HFMAGN(2)	HFMAGN(3)					
1.	1.	1.					
HFPHN(1)	HFPHN(2)	HFPHN(3)					
0.	0.	0.					
FLM	DLM	OMEGAM	THDM				
130.0	130.0	90.0	10.4				
FLS	DLS	OMEGAS	THDS				
267.808	370.904	87.753	21.0				

Fig. 6. Sample input data sheet; added sampling data requirements shown in box.

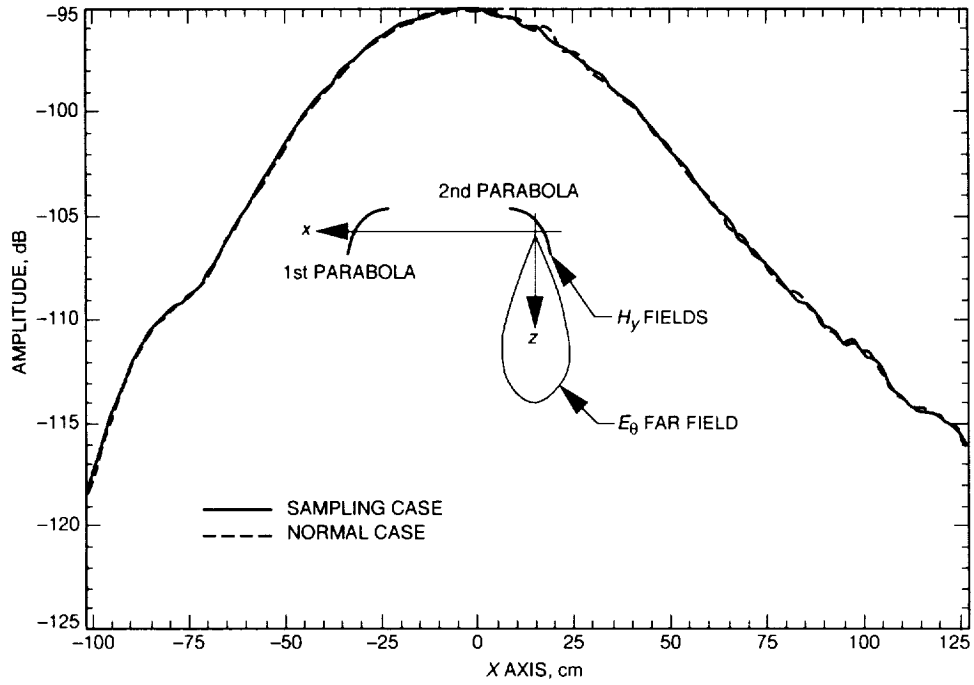


Fig. 7. The H_y field on the surface of the second parabola.

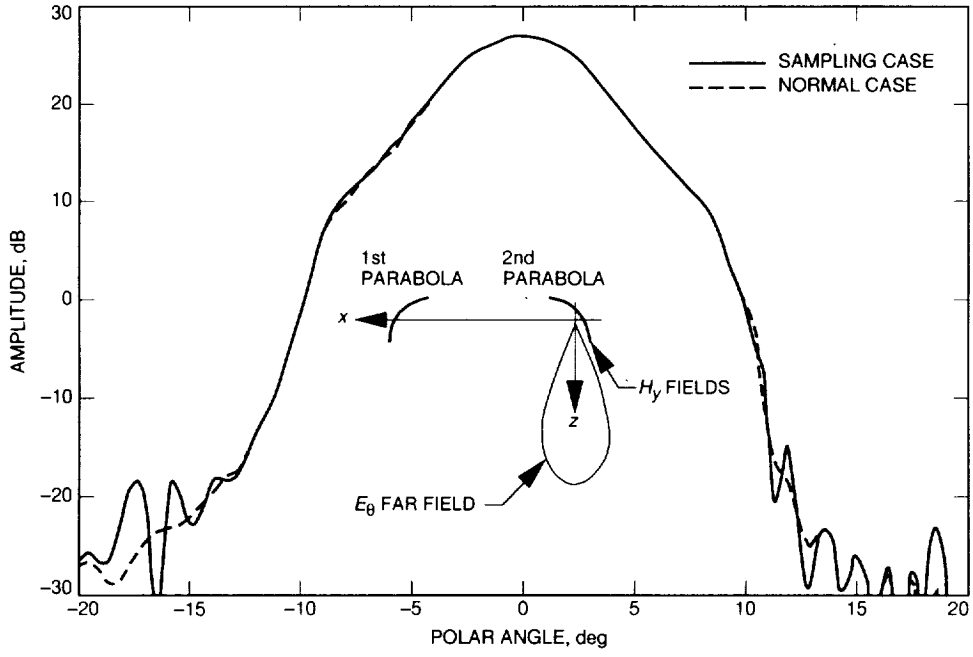


Fig. 8. The E_θ far-field component, parabola/parabola.

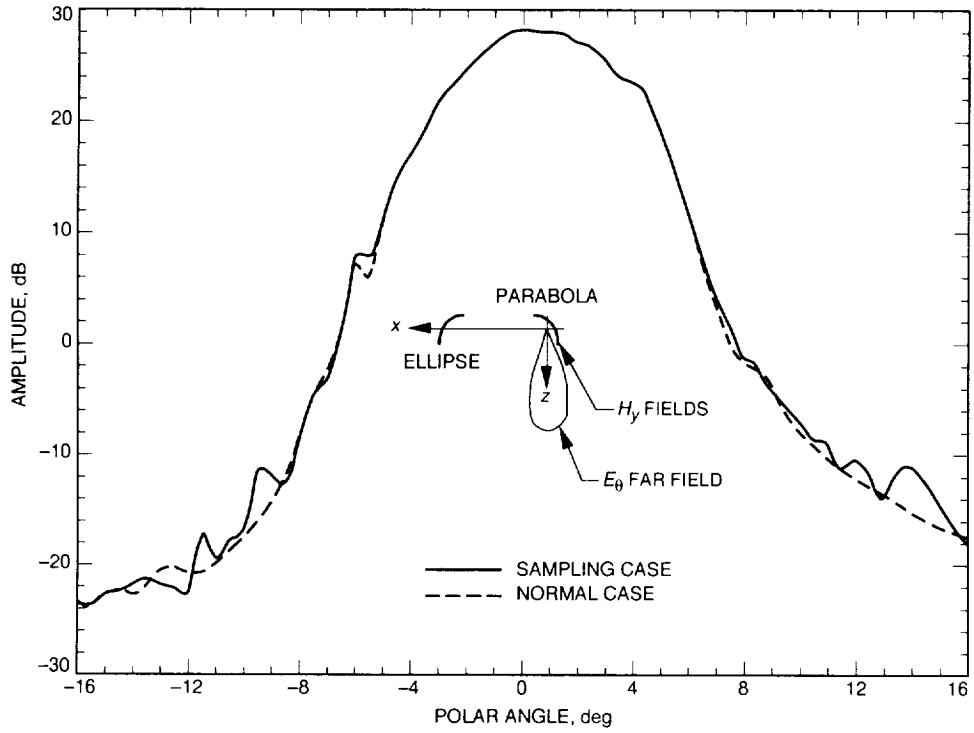


Fig. 9. The E_θ far-field component, ellipse/parabola, where the origin for sampling was determined by using the phase center algorithm.

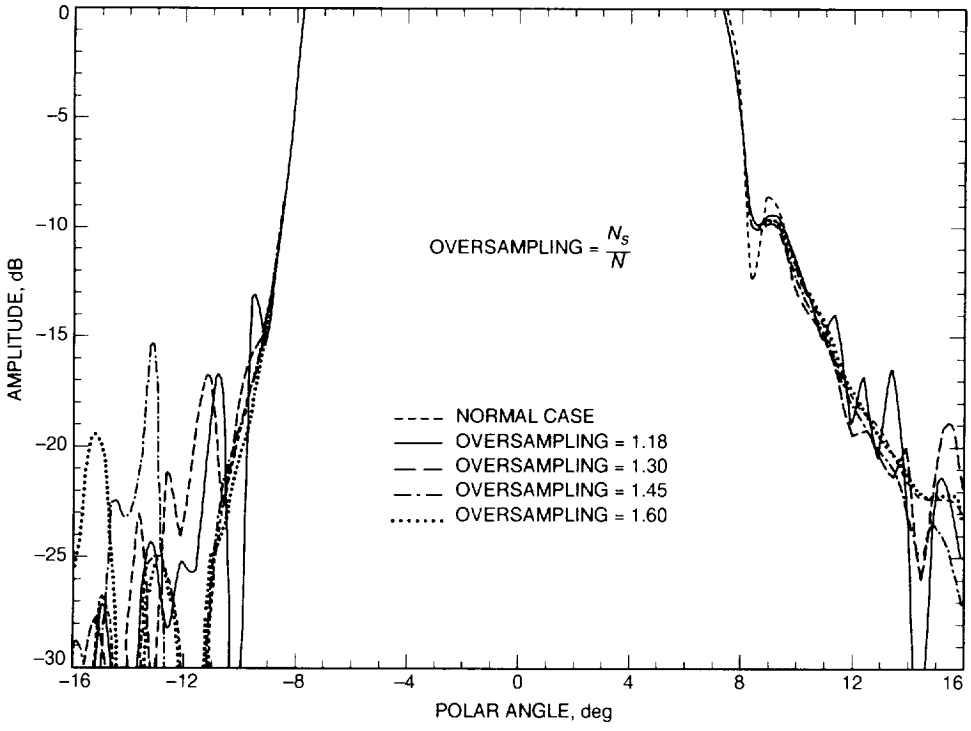


Fig. 10. The effect of sampling resolution, E_θ far-field component, ellipse/parabola.

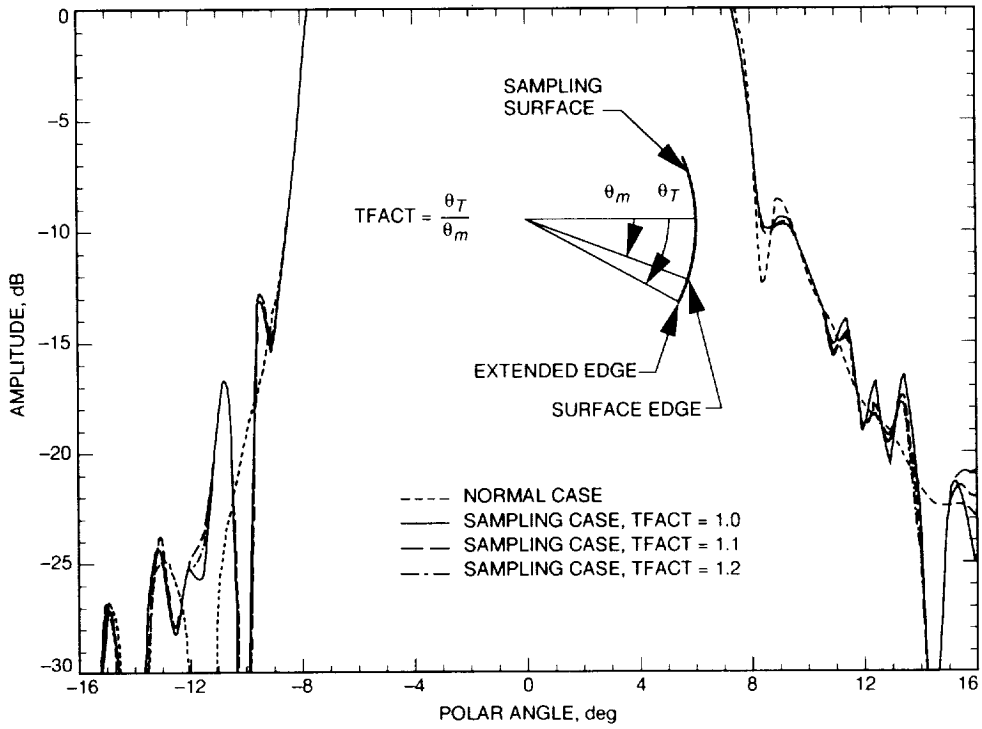
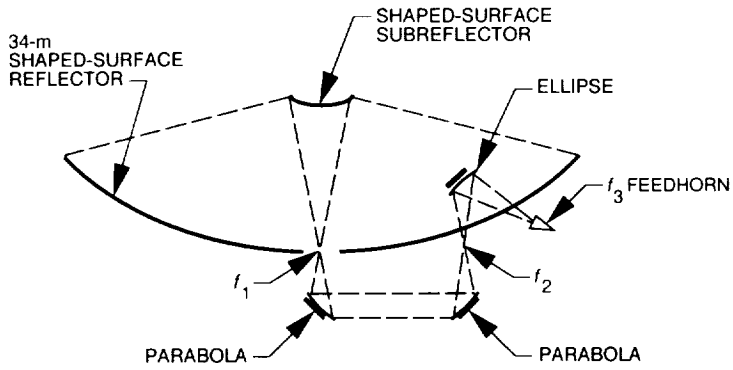


Fig. 11. The effect of increasing the size of the sampling surface, E_θ far-field component, ellipse/parabola.



GEOMETRY	INTEGRATION POINTS PER AXIS 1st/2nd SURFACE	POINTS PER WAVELENGTH	TIME SAMPLING, min	TIME NORMAL, min	TIME RATIO
ELLIPSE/PARABOLA	340/270	1/1	64.43	283.00	4.39
PARABOLA/PARABOLA	270/270	1/1	107.54	181.32	1.69
PARABOLA/SUBREFLECTOR	270/200	1/0.5	32.88	94.72	2.88
SUBTOTAL			204.85 (3.41 hr)	559.04 (9.32 hr)	2.73
SUBREFLECTOR/MAIN	200/24	0.5/••	133.82	133.82	1.00
TOTAL			338.67 (5.64 hr)	692.86 (11.55 hr)	2.05

Fig. 12. Schematic and analysis summary, 34-m BWG antenna, Ka-band.

DSS-13 S-/X-Band Microwave Feed System

F. Manshadi
Ground Antennas and Facilities Engineering Section

The configuration, detail design, and performance of the dual S-/X-band microwave feed system for the new DSN beam-waveguide antenna, Deep Space Station (DSS) 13, are reported. By using existing spare components, reducing fabrication cost of new components by simplifying their design, and using new fabrication techniques and material, this DSS-13 feed system was implemented successfully with a small budget and a very tight schedule. Measured noise temperature gains of the feed system are 17.5 K for S-band (2200–2300 MHz) and 24.0 K for X-band (8200–8600 MHz), which agree very closely with the predicted performance.

I. Introduction

DSS 13, located at the Deep Space Communications Complex in Goldstone, California, is NASA's first beam-waveguide tracking antenna. This antenna, which is primarily used for research and development tasks, is an elevation-over-azimuth type using wheel and track azimuth bearings.[1] DSS 13 has a parabolic main reflector and a hyperbolic subreflector that are 34 and 3.43 m in diameter, respectively. The subreflector is supported by a low-optical cross-section tripod. The main reflector and the subreflector are both shaped to provide near-uniform aperture illumination [1]. The reflector surfaces are designed and fabricated for providing high performance at 32 GHz.

In beam-waveguide antennas, the microwave front-end equipment is placed in a large room, called the pedestal room, under the main reflector structure. The microwave signal is guided from the pedestal room to the subreflector, or vice versa, through a system of flat or curved reflectors (see Fig. 1). There are several advantages to having the microwave feed equipment in the pedestal room: the microwave equipment will be stationary at all times; installation and maintenance are made easier; and there is the capability of using several feed systems without introducing additional loss due to the blockage.

At DSS 13, a large rotatable elliptical reflector in the basement (M5) and two flat and two parabolic reflectors

[1] M. Britcliffe, ed., *DSS-13 Beam-Waveguide Antenna Project: Phase 1 Final Report*, JPL D-8541 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 15, 1991.

in the beam-waveguide tube (M1 through M4) are used to guide the signal to the subreflector. Several different microwave feed systems are placed about the rotating ellipsoid. Each microwave feed can then be selected remotely by aligning the ellipsoid in the direction of that feed.

This article describes the S-/X-band feed system implemented for DSS 13. Section II covers the general theory of the operation of the feed system; Section III describes the detail design; and Section IV describes the performance. To reduce the cost of this task, attempts have been made to maximize the utilization of existing spare or decommissioned microwave components.

II. S-/X-Band Feed System Theory of Operation

The DSS-13 S-/X-band microwave feed system is capable of receiving S- and X-band signals simultaneously. Figure 2 shows the main components of this feed system: the X-band feed, the S-band feed, the S-/X-band dichroic reflector, and the X-band flat reflector. The S-band receiving frequency band is 2200 to 2300 MHz, and the X-band receiving frequency band is 8200 to 8600 MHz.

The S-band signal received from space is collected by the main- and sub-reflectors and is focused at F1 (see Fig. 1). Reflectors M1 through M4 guide the signal to the rotating ellipsoid focus F2. The signal is then scattered off the ellipsoid mirror, reflected by the dichroic reflector, and focused at the other focal point of the elliptical mirror. This signal is received by the S-band feedhorn in the S-band feed package.

The X-band signal is guided by the beam waveguide to the basement in the same manner as the S-band signal. However, after scattering off the ellipsoid, it passes through the dichroic mirror with very little loss, is reflected by the X-band flat reflector, and is focused at the other focal point of the ellipsoid. This signal is received by the X-band feedhorn in the X-band package.

III. Detail Design

A block diagram of the S-/X-band feed packages is shown in Fig. 3. The low-noise amplifier (LNA) used in this package was obtained from the decommissioned 26-m antenna at DSS 13. This is a dual-frequency LNA (i.e., it contains an X-band LNA and an S-band LNA in one cryogenic package). The S- and X-band feeds are packaged separately; however, they are physically connected since they share the same LNA package. The feedhorns are corrugated with the same corrugations and flare angle as the standard JPL feedhorns. The gain of the feedhorns is 19.1 dB for S-band and 25.0 dB for X-band. The polarizers provide the capability to select right-hand circular polarization (RCP) or left-hand circular polarization (LCP) reception. In the S-band package, the position of the polarizer can be changed easily by the use of the rotary joints; but in the X-band package, the position of the polarizer is fixed. To change polarization on the X-band, the polarizer has to be unscrewed and then rotated; however, spacers have been supplied to allow the addition of rotary joints at a future date. The couplers are used for injection of noise to check the linearity of the LNAs. The waveguide switches are used to connect the LNAs to the feedhorns or to the ambient loads for noise temperature and linearity measurements.

The S-/X-band dichroic reflector is a frequency-selective surface that passes the X-band signal but reflects the S-band signal. The S-/X-band dichroic plate used at DSS 13 is a 198.1-cm × 141.5-cm × 3.576-cm rectangular aluminum plate with an elliptical perforated area (see Fig. 4). The holes in the perforated area are based on an old dichroic plate design [2]. This design employs the Pyleguide holes originally used by Pyle [3]. However, to reduce the fabrication cost, the corner radius of the holes was increased from 0.013 to 0.318 cm, as shown in Fig. 4. An analysis of the propagation constant of the fields in the Pyleguide holes shows that the change in the propagation constant due to this modification is far less than the change due to the tolerances of the other critical dimensions of the holes.[2] This minor change reduced the fabrication cost of the S-/X-dichroic reflector by more than 60 percent.

The frames for the S- and X-band packages were fabricated using Bosch extruded aluminum struts. These struts are prefabricated, strong, lightweight, and flexible. Their anodized aluminum surface finish is scratch and corrosion resistant. Since all the elements of the frames are bolted together, it is very easy to modify these frames as needed in the future. The use of these materials resulted in a cost savings of more than 50 percent compared to conventional welded steel framing.

[2] J. C. Chen and P. H. Stanton, "Effect of Corner Radius on the Performance of an S-/X-Band Dichroic Plate With Pyleguide Aperture," JPL Interoffice Memorandum 3327-92-078 (internal document), Jet Propulsion Laboratory, Pasadena, California, November 24, 1992.

Figure 5 shows the S- and the X-band microwave feed assemblies and their overall dimensions.

IV. Feed System Performance

The preliminary predicted and measured noise temperatures of the S-/X-band LNAs, microwave feeds, and the overall DSS-13 beam-waveguide antenna are given in Table 1. The higher noise temperature measured for the X-band LNA is due to the age of that package and cannot be improved easily. The predictions are calculated from the theoretical or measured loss of the individual component of each system. The measurement data for the feeds were made at Goldstone before installation in the antenna pedestal room. The measurements for the overall antenna were made after the feed packages were installed and aligned in the pedestal room.

V. Conclusion

The DSS-13 feed system was implemented successfully with a small budget and a very tight schedule. This was accomplished by using existing spare components, reducing fabrication cost of the new components by simplifying their design, and using new fabrication techniques and material. The measured and predicted performance of the feed systems and the overall antenna agree closely.

Acknowledgments

This task owes its successful completion to the teamwork and dedication of the people who were involved. Many people from different organizations have made valuable contributions by offering their expertise and hard work. In particular, the author would like to acknowledge M. Esquivel for RF and optical design; J. Chen and P. Stanton for dichroic plate analysis; T. Otoshi, M. Franco, and S. Stewart for feed and antenna measurements; J. Fernandez and J. Loreman for repair and testing of the S-/X-band LNA; D. Ball, R. Bryant, and D. Ohashi for mechanical design; and M. Gatti for funding support.

References

[1] D. A. Bathker, W. Veruttipong, and P. W. Cramer, "Beam-Waveguide Antenna Performance Predictions with Comparison to Experimental Results," *IEEE Transactions on Microwave Theory and Technology*, vol. 40, no. 6, pp. 1274–1285, June 1992.

[2] P. D. Potter, *Improved Dichroic Reflector Design for the 64-Meter Antenna S- and X-Band Feed Systems*, JPL Technical Report 32-1526, vol. XIX, Jet Propulsion Laboratory, Pasadena, California, pp. 55–62, February 15, 1974.

[3] J. R. Pyle, "Cutoff Wavelength of Waveguides with Unusual Cross Sections," *IEEE Transactions on Microwave Theory and Technology*, vol. MTT-12, no. 5, pp. 556–557, September 1964.

Table 1. Noise temperature predictions and measurements.[a]

System	S-band predictions, K	S-band measurements, K	X-band predictions, K	X-band measurements, K
LNA	8.3	8.72	12.0	14.09
Feed	17.69	17.5	23.07	25.5
Antenna	37.26	38.0	32.90	35.5

[a] Values are not corrected for weather.

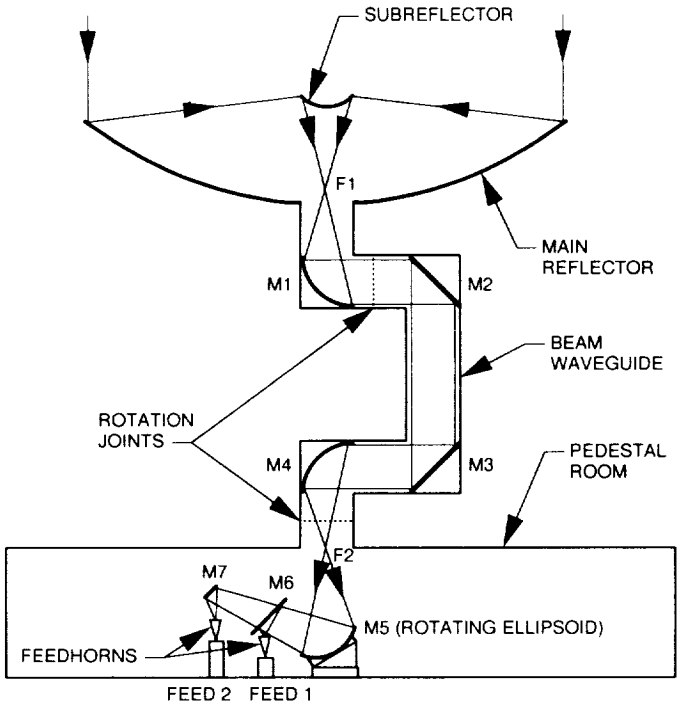


Fig. 1. The DSS-13 beam-waveguide antenna configuration.

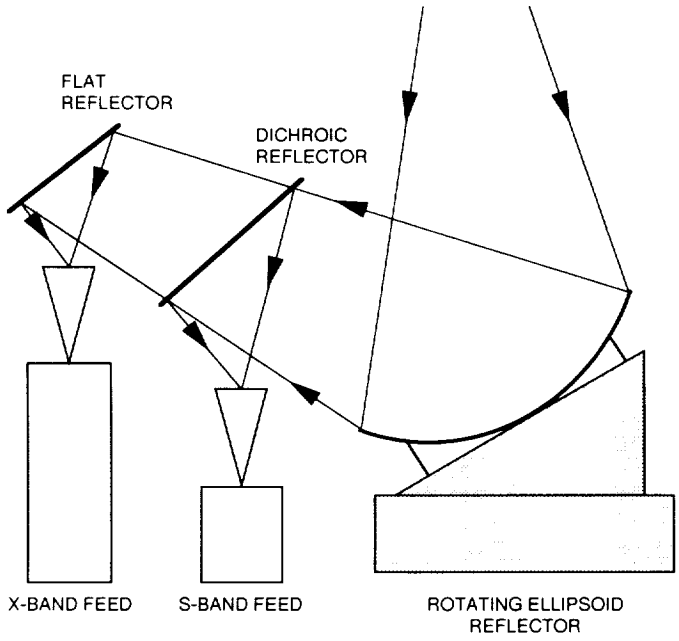


Fig. 2. The DSS-13 S-/X-band microwave feed system.

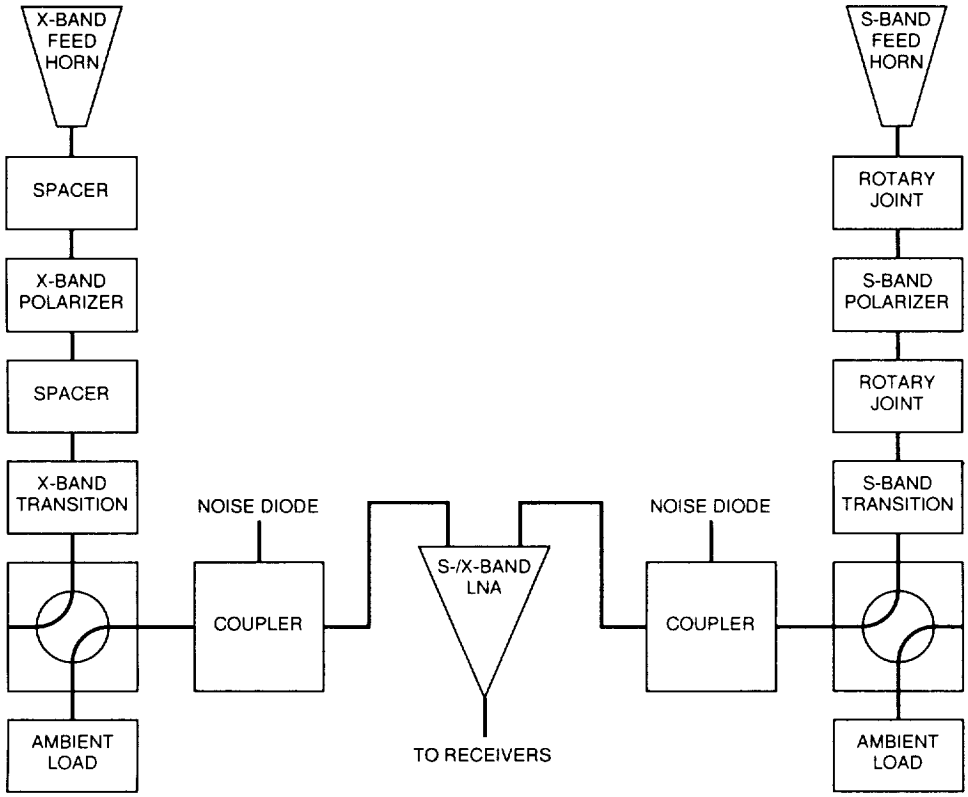


Fig. 3. The DSS-13 S-/X-band microwave feed block diagram.

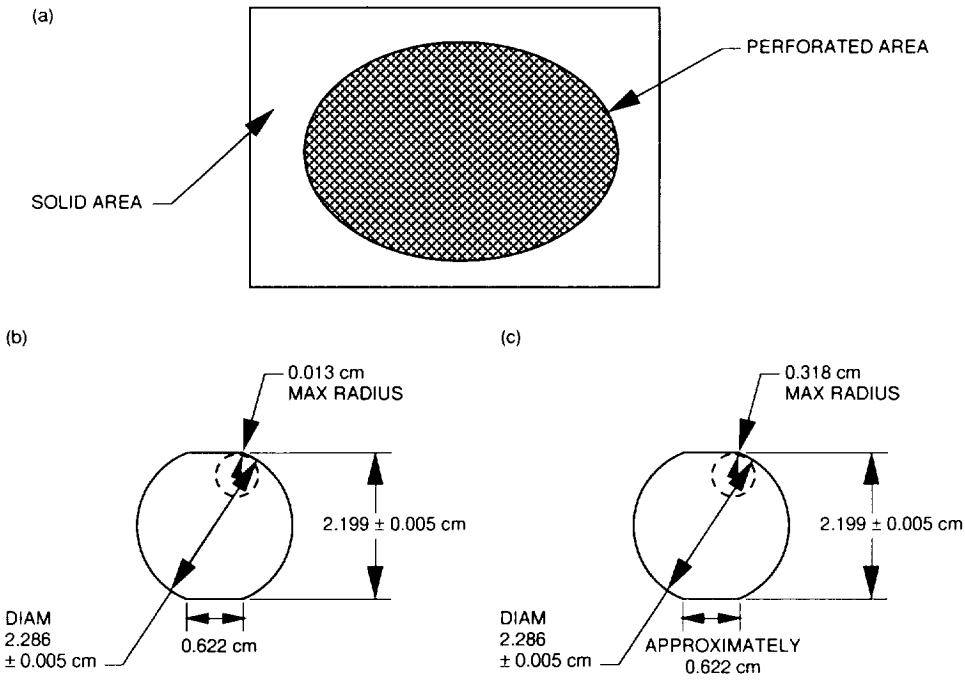


Fig. 4. The DSS-13 S-/X-band dichroic reflector: (a) aluminum plate with an elliptical perforated area; (b) original Pyleguide design; and (c) modified Pyleguide design.

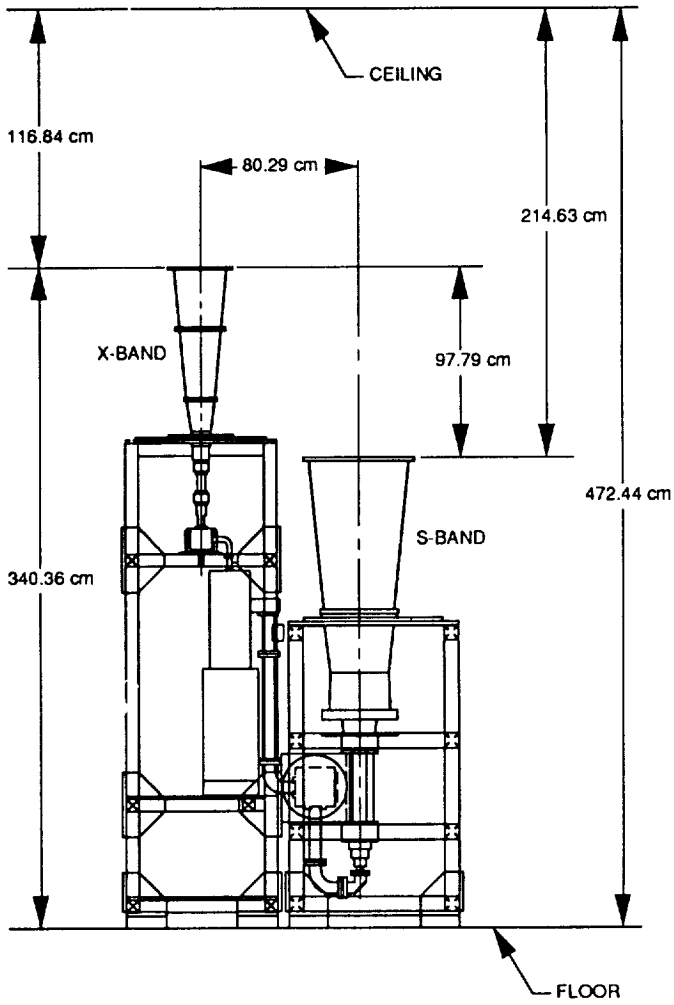


Fig. 5. The S- and X-band microwave feed assemblies.

A 1- to 10-GHz Downconverter for High-Resolution Microwave Survey

D. McWatters
Radio Frequency and Microwave Subsystems Section

A downconverter was designed, built, and tested for the High Resolution Microwave Survey project. The input frequency range is 1 to 10 GHz with instantaneous bandwidth of 350 MHz and dynamic range of 125 dB/Hz. Requirements were derived for the local oscillators and special design techniques were implemented to achieve the high degree of spectral purity required.

I. Introduction

The 1- to 10-GHz RF-to-IF downconverter (RID) was designed to perform the RF-to-IF frequency conversion portion of the High Resolution Microwave Survey (HRMS) RF system. The RF system is composed of a series of antenna feeds, hybrids, and low-noise amplifiers (LNAs) that cover the band of 1 to 10 GHz and deliver two polarization channels to the inputs of the RIDs. The RID is duplicated to process two channels simultaneously. The IF output is designed to be transmitted over a high dynamic range fiber-optic link (FOL) to the control room, where it is fed into an IF-to-baseband (IF/BB) frequency converter followed by a real-time 32-million-channel fast Fourier transform (FFT) analyzer [1] and sophisticated signal detection algorithms. This article describes the RID as well as the design process and rationale that led to its realization.

II. Requirements

The task of the HRMS project is to search for extremely weak narrow-band signals over a wide frequency range in the presence of very strong interfering signals. The challenge in this design was to trade off the competing requirements of wide bandwidth, high dynamic range, and high spectral purity. The requirements for the RID were derived from the RF system requirements and are listed in Table 1.[1]

III. Design Approach

To assist in achieving a design that meets these challenging requirements, two computer-based design tools were developed. The first tool is a Fortran program that is highly interactive with the user, performing mixer analysis to discover all the intermodulation products resulting from a particular frequency conversion scheme. The program inputs are the RF and local oscillator (LO) frequency ranges and increments, the IF filter shape, the mixer intermodula-

[1] G. A. Zimmerman, *Search for Extraterrestrial Intelligence Microwave Observing Project Sky Survey Element, Sky Survey Operational System, Functional Requirements Document*, Document No. 1720-4100 (internal document), Jet Propulsion Laboratory, Pasadena, California, June 10, 1992.

tion table, the required suppression, and the highest-order harmonic to be calculated. The program outputs a list of all the frequency spurs that result from this mix with suppression levels that exceed the requirement. Figure 1(a) is an example run showing that no intermodulation products violate 60-dB suppression for the first frequency conversion range. Figure 1(b) is the same frequency conversion with a more stringent suppression requirement. The program lists both the frequency components inside the IF band that are suppressed by the mixer as well as the components outside the IF band that are suppressed further by the bandpass filter.

The second tool is a custom spreadsheet that analyzes receiver performance component by component. The inputs to the spreadsheet are the receiver components in sequence with their associated gain, noise temperature, bandwidth, and compression point. The outputs of the spreadsheet are the contributions of each component to the overall receiver noise temperature, signal levels, total gain, noise power density, total noise power, and measures of dynamic range. This tool is helpful when performing system design and trade-off analysis. Figure 2 shows the end-to-end model developed for this RF system using this spreadsheet tool. The model includes every component of the system starting from the antenna all the way to the analog-to-digital converter. The group of components that constitutes the RID is shown shaded.

The RID frequency conversion scheme uses two frequency mixes. In the first mix, the entire 1- to 10-GHz block is upconverted using a high side LO (LO1). This LO is tuned over the range of 11.8 to 20.5 GHz with a maximum step size of 300 MHz. The IF is the lower sideband of the mix. The advantage of the high-side LO mix is that the LO harmonics generated by the mix, which create strong intermodulation products, are out of band. The upconversion is designed such that the IF band is outside the input RF band. Producing any signals in the receiver in the input band could lead to false-signal detection. The third constraint was to keep the frequencies as low as possible for reasons of cost and ease of implementation.

The second mix uses another high-side LO, thus restoring the inverted spectrum to its original orientation. The second LO (LO2) is fixed at a frequency of 11.2 GHz, which is simple to synthesize by phase-lock multiplying a 100-MHz pure reference signal by a factor of 112. The resulting output IF band is centered at 550 MHz with a bandwidth of 350 MHz. Since the IF bandwidth is narrower than an octave, it may be frequency converted as a block in the next stage of the receiver. Also, 550 MHz is a practical IF frequency for transmission over distance using a coaxial cable or a fiber-optic link.

Although wideband devices that cover the wide input band of 1- to 10 GHz exist, the band needed to be narrowed before the first mix to meet the high degree of spectral purity required. A switched-filter-bank approach was selected with eight contiguous bands that divide the 1- to 10-GHz block optimally for minimizing mixer intermodulation products, physical size, and complexity. Figure 3 is a block diagram of the RID.

IV. Local Oscillator Requirements

The requirements for the local oscillators of a downconverter are dependent on the specific frequency scheme and gain profile of the downconverter. For the RID to meet its phase noise, spectral purity, and dynamic range requirements, LO1 and LO2 must meet the requirements presented below. The requirement-derivation process is also described.

A. Local Oscillator Power Level

The LO power level requirement results from the required mixer performance: the higher the LO power, the higher the mixer dynamic range and intermodulation suppression. It was determined that the LO1 level is required to be +10 dBm while the LO2 level is required to be +20 dBm. LO2 power is higher than that of LO1 because the gain stages between the two mixers result in increased signal levels and thus increased difficulty in meeting the dynamic range requirements.

B. Spectral Purity

The spectral purity requirements result from analyzing all the different ways in which a spurious signal (spur) accompanying the LO could end up in the IF band and be falsely detected as a signal of interest. The required suppression of spurs relative to the LO is measured in dBc (dB relative to carrier). There are three frequency regions around the LO frequency that are derived separately [2]. For example, the requirements for LO2 are derived below. First, the frequency ranges of the second frequency conversion are listed:

(1) LO2 frequency is 11,200 MHz.

(2) Input RF range is 10,450 to 10,850 MHz.

(3) Output IF range is 350 to 750 MHz.

C-2

The first spectrally pure region extends over an IF bandwidth (400 MHz) in each direction around LO2. Any spur frequency within this range will modulate an incoming signal in the same way it modulates the LO signal. The required suppression for this range is derived by subtracting the minimum detectable signal from the maximum signal of interest:

$$-73 - (-172)\ \text{dBm} = 99\ \text{dB}$$

Therefore, for 10,800 to 11,600 MHz, the required suppression is −99 dBc.

The second region includes any spur frequency that, when mixed with LO2, will end up in the IF band. After leaking from the L port to the R port in the mixer, the spur is required to be 3 dB below the weakest incoming signal of interest. Therefore, for 10,450 to 10,850 MHz and 11,550 to 11,950 MHz, the required suppression is −102 dBc.

Finally, the third region consists of spur frequencies in the IF range that may leak from the L port to the I port in the mixer. After leaking from the L port to the I port, the spur is required to be 3 dB below the weakest signal of interest at the IF band. Therefore, for 350 to 750 MHz, the required suppression is −109 dBc.

C. Phase Noise

The receiver phase noise requirement (Table 1, no. 6) can be met only if the contributions from all the local oscillators do not add up to exceed it. Since no specific IF/BB design is discussed here, the LO contribution from the IF/BB downconverter will be omitted here.

LO1 and LO2 frequencies are derived from a common reference frequency using phase-locked multiplication techniques. The relative contribution from each LO depends on its frequency relative to some reference frequency. Within the phase-locked loop bandwidth, the phase noise of each LO is synchronously related to the reference phase noise by a factor of $20 \times \log N$, where N is the frequency multiplication factor [3]. The LO1 maximum frequency is 20,500 MHz, and the LO2 frequency is 11,200 MHz. Therefore, within the loop bandwidth, LO1 must exceed the receiver phase noise requirements by

$$20 \times \log((20,500 + 11,200)/20,500) = 3.8\ \text{dB}$$

LO2 must exceed the receiver phase noise requirements by

$$20 \times \log((20,500 + 11,200)/11,200) = 9.0\ \text{dB}$$

and a 100-MHz frequency reference has to exceed the receiver phase noise requirements by

$$20 \times \log(205 + 112) = 50\ \text{dB}$$

Outside the loop bandwidth, the phase noise of each LO is independently random. We may, for example, allocate equal contributions from each LO, such that LO1 and LO2 each has to exceed the receiver phase noise requirements by 3 dB.

V. Implementation

A. Circuit Description

We discussed the requirements and the design approach for the RID. The next step involves selecting circuit components and devices that implement the required performance called for by the design.

Figure 3 is a block diagram of the RID. A1 is a miniature wideband LNA, custom designed for high output power. A1 is followed by PA1, which is a programmable step attenuator covering the range of 0 to 11 dB in 1-dB steps. The attenuator can be used to compensate for a potentially long cable between the system LNA and the RID; this cable would cause insertion-loss variation with frequency. PA1 can also attain higher dynamic range (at a cost of increased system temperature) in the presence of very strong interfering signals. PA1 is computer monitored and controlled.

PA1 is followed by FB1, a computer-controlled, switched-filter bank implemented in microstrip with solid-state switches at the input and output. The solid-state switch approach has the advantages of higher reliability and smaller size as compared to the multiport coaxial electromechanical switches. These advantages are important in this system. The disadvantages of the solid-state switch as compared to the electromechanical switch are higher insertion loss, poorer impedance match, increased sensitivity to strong signals, and amplitude variation with temperature. All the disadvantages were adequately addressed. The switched-filter bank was designed to have adequate insertion loss, flatness, and immunity to strong signals. The impedance match was improved with attenuator pads, and amplitude drift can be minimized with temperature control.

FB1 is followed by a very wideband IF mixer (M1), which is followed by two identical bandpass filters (BPF1-A,B) with center frequencies (F_c) of 10,650 MHz and bandwidths (BW) of 400 MHz. The reason for filtering twice

is to achieve the high degree of rejection required to suppress leakage from LO1, which would mix with the LO2 frequency and result in spurs in the final IF band. The bandpass filter pair is followed by A2, an amplifier designed for a very flat frequency response. A2 is followed by a high power mixer (M2). The mixer is followed by a bandpass filter (BPF2) which establishes the IF output bandwidth. Additional gain is provided by A3, an amplifier designed for high dynamic range.

The LO1 and LO2 signals are each amplified in the RID using AL1 and AL2, respectively. This allows an input power level of 0 dBm, a requirement that can be easily achieved by most modern signal generators. LO2 is sampled through a 30-dB directional coupler (CPL2) with a diode detector (DET2). The detector is input to the monitor and control interface circuit (MCIC), which is a custom-designed printed circuit board. On the MCIC, the detector output is amplified and compared to a reference threshold, producing a digital indication of LO2 signal presence, which may be read by a computer. SW1 is an electromechanical coaxial switch that selects between the sampled outputs of LO1 and the IF output band. SW1 is computer controlled and monitored. The selected input is designed to be connected, directly outside the RID, to the power sensor of a computer-monitored power meter (PM). The reading of the LO1 power level can be used to adjust LO1 to the desired power level over its wide frequency band. The dc voltages used to power the active components in the RID are monitored in the MCIC and reported to the computer.

The directional coupler (CPLRS) at the output of the RID serves to inject a test signal for testing the fiber-optic link. Signal injection points for the RID are designed into the LNA package.

B. Packaging

Once the detailed circuit design is complete, the mechanical design follows. The main goal in the RID mechanical design was to achieve a low-noise environment for the circuit. The chassis designed for the RID circuit is shown in Figs. 4 and 5. Several techniques were used to reduce signal leakage and crosstalk, which would compromise the receiver spectral purity.

Coaxial cables are used to connect devices that are individually packaged with coaxial connectors. This method reduces signal leakage in comparison with a microstrip layout.

To isolate inputs and outputs of devices in the signal path, each critical signal-conditioning function is confined in a separate chamber. The chassis is machined from a single block of aluminum with chambers on the top and bottom sides. The signal is routed from chamber to chamber via coaxial feed-through adapters. The top and bottom covers are thick and flat, which facilitates sealing the box tightly by using closely spaced screws.

Isolators and pad attenuators are used to improve the impedance matching between filters, amplifiers, and mixer ports.

The dc voltages that enter the chassis are first filtered in their own chamber and then distributed to other chambers. To reduce signal leakage from one device to another via the dc voltage wires, feed-through capacitors are used to pass wires through the chassis walls. Each capacitor provides 60 dB of suppression to a 10-GHz signal. In addition, ferrite beads are used around each wire to further attenuate high-frequency leakage. Finally, the wire used has a special dielectric coating that provides additional high-frequency filtering.

VI. Test Results

The gain profile model (Fig. 2) was verified by testing. Testing included frequency response, noise temperature, gain stages, 1-dB compression, mixing intermodulation products, and *s*-parameters. The noise temperature and gain profiles of the RID generated by the model are shown in Figs. 6 and 7, respectively. One way we evaluated the effectiveness of the implemented design techniques was to tune LO1 to the frequency of 11,750 MHz. If this frequency leaks through the receiver, it may mix with the LO2 frequency (11,200 MHz) and result in a spur at 550 MHz, which is the center of the output IF band. The spur was measured at a level of −106 dBm, which is 17 dB below the noise floor of the RF system at the same point. This successful rejection of the spur is a result of the following measures: double filtering between M1 and M2; improved impedance matching of the mixers and filters using isolators; dc line filtering of AL1, AL2, and A2 using feed-through capacitors, ferrite beads, and lossy dielectric wire; shielded coaxial cables connecting the individually shielded signal processing components; and sealed compartments to separate the different frequency functions.

VII. Summary

A 1- to 10-GHz downconverter was designed, built, and tested. The requirements were listed, and the design process was described. Performance was verified in standalone bench testing using HP83712A synthesizers for LO1,

LO2, and RF input. The requirements described above in Section II were met. Only one channel was assembled and tested by the time the HRMS project was eliminated by Congress.

The RID can be useful for other RF-to-IF frequency-conversion applications featuring an input frequency range of 1 to 10 GHz with an instantaneous bandwidth of 350 MHz and a dynamic range of 125 dB/Hz. The RID was designed as a part of a complex computer-controlled RF system; however, it may be used as a stand-alone downconverter using commercial synthesizers for LO1 and LO2. A user interface box was constructed, facilitating local manual monitor and control functions.

References

[1] G. A. Zimmerman, M. F. Garyantes, M. J. Grimm, and B. Charny, "A 640-MHz 32-Megachannel Real-Time Polyphase-FFT Spectrum Analyzer," *The Telecommunications and Data Acquisition Progress Report 42-107, vol. July–September 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 132–140, November 15, 1991.

[2] W. F. Egan, *Frequency Synthesis by Phase Lock*, Chapter 4, Malabar, Florida: Robert E. Krieger Publishing Co., pp. 61–75, 1990.

[3] V. Manassewitsch, *Frequency Synthesizers Theory and Design*, New York: John Wiley & Sons, pp. 120–124, 1987.

Table 1. RID requirements.

No.	Parameter	Value
1	Input frequency range, GHz	1 to 10
2	Output instantaneous 0.5-dB bandwidth, MHz	320
3	Noise-temperature contribution of downconverter to front end LNA	5 percent max (1.2 K) for nominal attenuator settings
4	Two simultaneous channels with equal tuning frequency	
5	Computer controlled and monitored	
6	Receiver single-sideband phase noise	

Frequency offset from carrier, Hz	Single-sideband phase noise in 1-Hz BW, dBc
5	−49
10	−64
60	−97
100	−105
120	−105
180	−107
≥300	−112

7	Spectrally pure dynamic range for a single tone (of −73 dBm at antenna)	89 dBc at 10 GHz and 99 dBc at 1 GHz (dB from the tone down to mixing products with local oscillators and their spurs)
8	Nominal dynamic range (defined from noise floor to 1-dB compression point)	40 dB, for 320-MHz bandwidth; for a high dynamic range mode, 50-dB dynamic range is required and system noise temperature (T_{sys}) is allowed to degrade by 3 K
9	Minimum gain, dB	34
10	Programmable gain range, dB	10; step size: 1 dB
11	Gain deviation (instantaneous peak to peak over 320 MHz), dB	±2

(a)

```
program IMODP.EXE by Dalia McWatters, output
file: INTERMOD.FRQ
LISTING OF INTERMODULTION PRODUCTS SUPPRESSION
RESULTING FROM THE FOLLOWING FREQUENCY CONVERSION:

FRL =   1000.000 FRH =   1600.000
RF FREQ. INCREMENT =     50.000
FLL =  11650.000 FLH =  12250.000
LO FREQ. INCREMENT =     50.000
IF =  10650.000
IF FILTER SHAPE =   40.0 dB @   1.200  3dB bw from center freq.
BW =    400.000
REQUIRED REJECTION:  60.0dB
HIGHEST ORDER INTERMOD CALCULATED:    9
MIXER INTERMOD TABLE USED:   LRLL.TBL

F(M,N) = ABS(M*FR + N*FL)  if in BW
-----------------------------------------------
```

(b)

```
program IMODP.EXE by Dalia McWatters, output
file: INTERMOD.FRQ
LISTING OF INTERMODULTION PRODUCTS SUPPRESSION
RESULTING FROM THE FOLLOWING FREQUENCY CONVERSION:

FRL =   1000.000 FRH =   1600.000
RF FREQ. INCREMENT =    100.000
FLL =  11650.000 FLH =  12250.000
LO FREQ. INCREMENT =    100.000
IF =  10650.000
IF FILTER SHAPE =   40.0 dB @   1.200  3dB bw from center freq.
BW =    400.000
REQUIRED REJECTION:  99.0dB
HIGHEST ORDER INTERMOD CALCULATED:    9
MIXER INTERMOD TABLE USED:   LRLL.TBL

F(M,N) = ABS(M*FR + N*FL)  if in BW
-----------------------------------------------

FLO =  11650.000 FRF =   1000.000
Atten =  61.79dB
F(  0R , -1L ) =  11650.000

FLO =  11650.000 FRF =   1000.000
Atten =  61.79dB
F(  0R ,  1L ) =  11650.000

FLO =  11650.000 FRF =   1100.000
Atten =  61.79dB
F(  0R , -1L ) =  11650.000

FLO =  11650.000 FRF =   1100.000
Atten =  61.79dB
F(  0R ,  1L ) =  11650.000

FLO =  11650.000 FRF =   1200.000
Atten =  90.00dB
F( -9R ,  0L ) =  10800.000

FLO =  11650.000 FRF =   1200.000
Atten =  61.79dB
F(  0R , -1L ) =  11650.000

FLO =  11650.000 FRF =   1200.000
Atten =  61.79dB
F(  0R ,  1L ) =  11650.000

FLO =  11650.000 FRF =   1200.000
Atten =  90.00dB
F(  9R ,  0L ) =  10800.000

FLO =  11650.000 FRF =   1300.000
Atten =  97.14dB

    .
    .
    .
```

Fig. 1. Mixing products analysis program outputs: (a) 60-dB suppression required and (b) 99-dB suppression required.

DOWNCONVERTER GAIN AND NOISE TEMP PROFILE						Dalia McWatters		12/28/93				
HRMS RF/IF	with PROTOTYPE IF/BB					HRIDC.XLS						
BLOCK	Noise Temp	GAIN	Pol	BANDWIDTH	sigout	Tsys	TOTAL	PnOUT/Hz	Pn(out)	DR	HR	BLOCK
	Kelvin	dB	dBm	Hz	dBm	Kelvins	GAIN(dB)	dBm/Hz	dBm	(dB)	(dB)	
ANTENNA	14.4	0	7	2.50E+09	-62	14.4000	0.0	-187.0	-93.1			ANTENNA
PAD		0	7	2.50E+09	-62	14.4000	0.0	-187.0	-93.1			PAD
LNA	9.6	52	7	2.50E+09	-10	24.00	52.0	-132.8	-38.8	45.8	17.0	LNA
CABLE+PAD	(-5to-8)	-5	7	2.50E+09	-15	24.00	47.0	-137.8	-43.8			CABLE+PAD
A1	290	28	21	2.50E+09	13	24.01	75.0	-109.8	-15.8	36.8	8.0	A1
PA1	(-2 to-13)	-5	21	2.50E+09	8	24.01	70.0	-114.8	-20.8			PA1
amp		0	21	2.50E+09	8	24.01	70.0	-114.8	-20.8			amp
FB1+PADS		-16	21	5.00E+08	-8	24.01	54.0	-130.8	-43.8			FB1+PADS
M1	1164	-7	-3	5.00E+08	-15	24.02	47.0	-137.8	-50.8	47.8	12.0	M1
BPF1		-3	-3	5.00E+08	-18	24.02	44.0	-140.8	-53.8			BPF1
A2	440	24.5	24	5.00E+08	6.5	24.04	68.5	-116.3	-29.3	53.3	17.5	A2
PAD		-0.5	24	5.00E+08	6	24.04	68.0	-116.8	-29.8			PAD
M2	2014	-9	8	5.00E+08	-3	24.04	59.0	-125.8	-38.8	46.8	11.0	M2
BPF2+PADS		-12	8	5.00E+08	-15	24.04	47.0	-137.8	-50.8			BPF2+PADS
A3	360	40	30	5.00E+08	25	24.05	87.0	-97.8	-10.8	40.8	5.0	A3
CPLRS+PAD		-7	30	5.00E+08	18	24.05	80.0	-104.8	-17.8			CPLRS+PAD
FOL	6000000	-17	1	5.00E+08	1	24.11	63.0	-121.8	-34.8	35.8	0.0	FOL
pad		-15		5.00E+08	-14	24.1162	48.0	-136.8	-49.8			pad
IF/BB D/C in		0		1.10E+08	-14	24.1162	48.0	-136.8	-56.4			IF/BB D/C in
LPF+pads		-4		1.10E+08	-18	24.1231	44.0	-140.8	-60.4			LPF+pads
Mixer1	1200.0	-6	7	1.10E+08	-24	24.1709	38.0	-146.8	-66.4	73.4	31.0	Mixer1
BPF+circ		-2		1.10E+08	-26	24.1978	36.0	-148.8	-68.4			BPF+circ
AMP2	130.0	23	5	1.10E+08	-3	24.2304	59.0	-125.8	-45.4	50.4	8.0	AMP2
Mixer2+pad		-19	-3	1.10E+08	-22	24.2591	40.0	-144.8	-64.4	61.4	19.0	Mixer2+pad
AMP3	130.0	37	15	1.10E+08	15	24.2721	77.0	-107.8	-27.4	42.4	0.0	AMP3
SAW+pads		-42		4.00E+07	-27	24.3638	35.0	-149.8	-73.7			SAW+pads
AMP4	130.0	37	15	4.00E+07	10	24.4049	72.0	-112.7	-36.7	51.7	5.0	AMP4
MIXER3+pad		-14	12	4.00E+07	-4	24.4053	58.0	-126.7	-50.7	62.7	16.0	MIXER3+pad
AMP5	7000.0	20	29	4.00E+07	16	24.4164	78.0	-106.7	-30.7	59.7	13.0	AMP5
lpf+pads+PA2	(-26 to-57)	-32		4.00E+07	-16	24.4237	46.0	-138.7	-62.7			lpf+pads+PA2
AMP6	7000.0	20	29	4.00E+07	4	24.5995	66.0	-118.7	-42.7	71.7	25.0	AMP6
PAD		-2		4.00E+07	2	24.5996	64.0	-120.7	-44.7			PAD
AMP7	7000.0	21	29	4.00E+07	23	24.6024	85.0	-99.7	-23.7	52.7	6.0	AMP7
PAD		-15		4.00E+07	8	24.6024	70.0	-114.7	-38.7			PAD
AMP8	7000.0	21	29	4.00E+07	29	24.6031	91.0	-93.7	-17.7	46.7	0.0	AMP8
PAD		-3		4.00E+07	26	24.6031	88.0	-96.7	-20.7			PAD
IF/BB D/C out		0		4.00E+07	26	24.6031	88.0	-96.7	-20.7			IF/BB D/C out
WBSA PA		-28		4.00E+07	-2	24.6034	60.0	-124.7	-48.7			WBSA PA
CLC100	865.0	20	12	4.00E+07	18	24.6042	80.0	-104.7	-28.7	40.7	-6.0	CLC100
pad		-3		4.00E+07	15	24.6042	77.0	-107.7	-31.7			pad
ADC INPUT	14500000.0	0	-1	4.00E+07	15	24.8936	77.0	-107.7	-31.6	30.6	-16.0	ADC INPUT

Fig. 2. Receiver gain and noise profile spreadsheet.

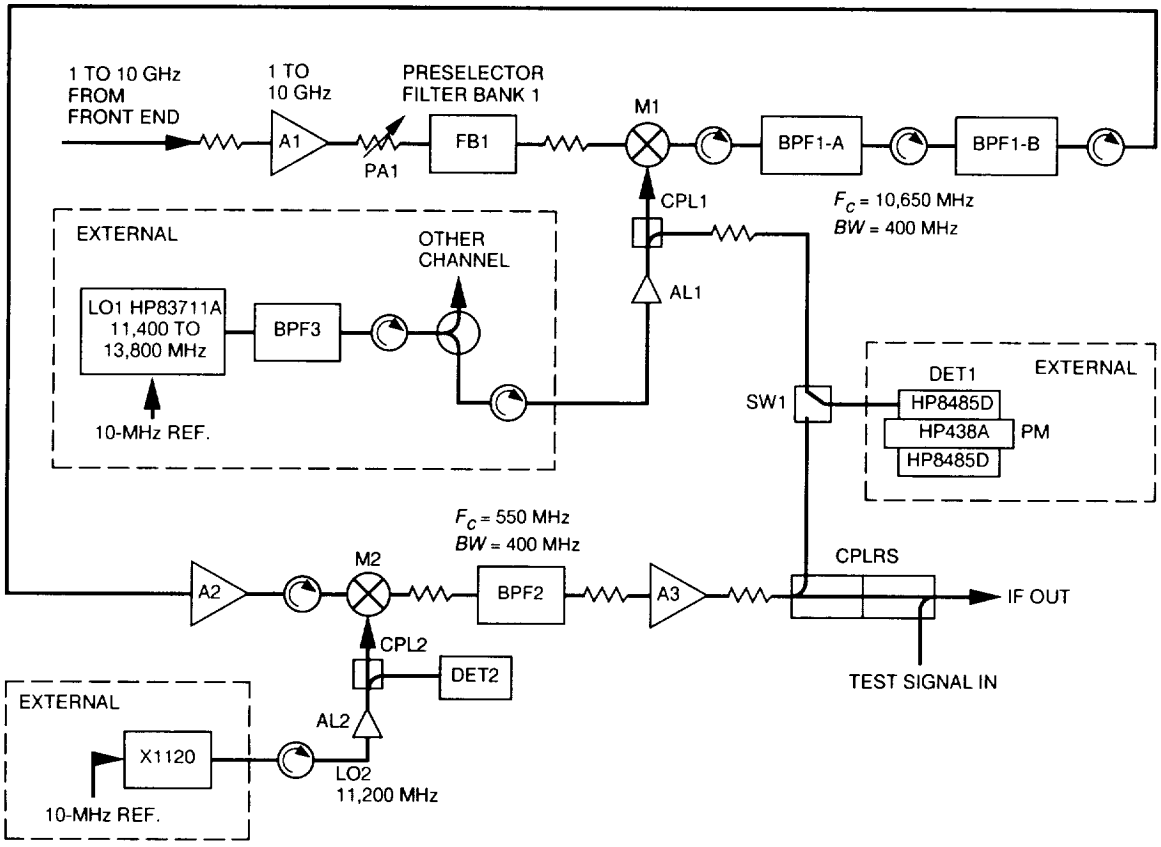


Fig. 3. RF/IF downconverter block diagram.

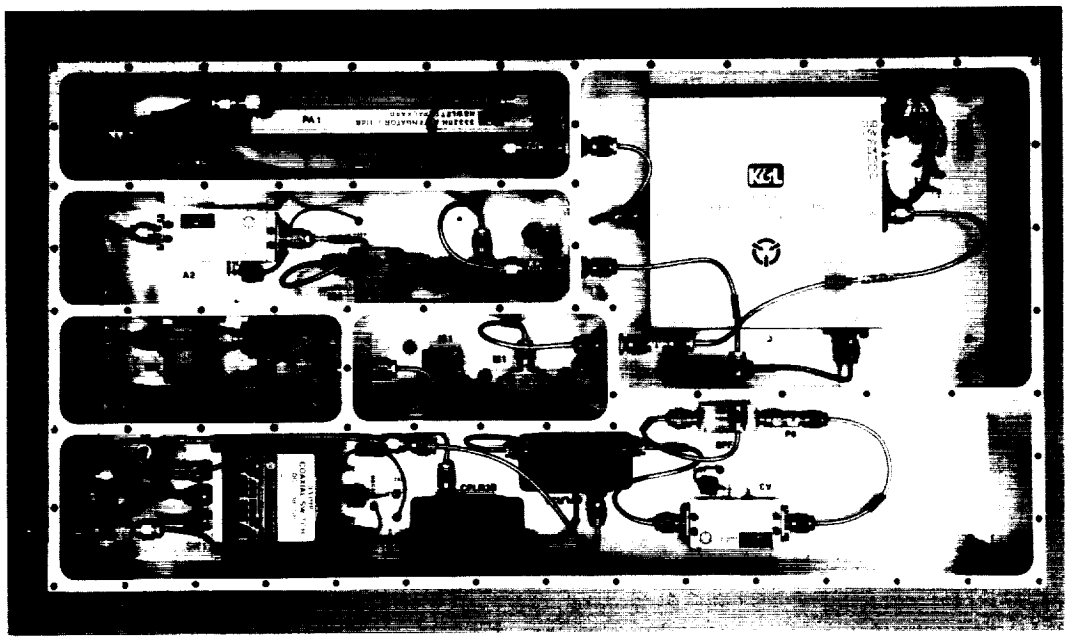


Fig. 4. RF/IF downconverter chassis fully assembled, top side.

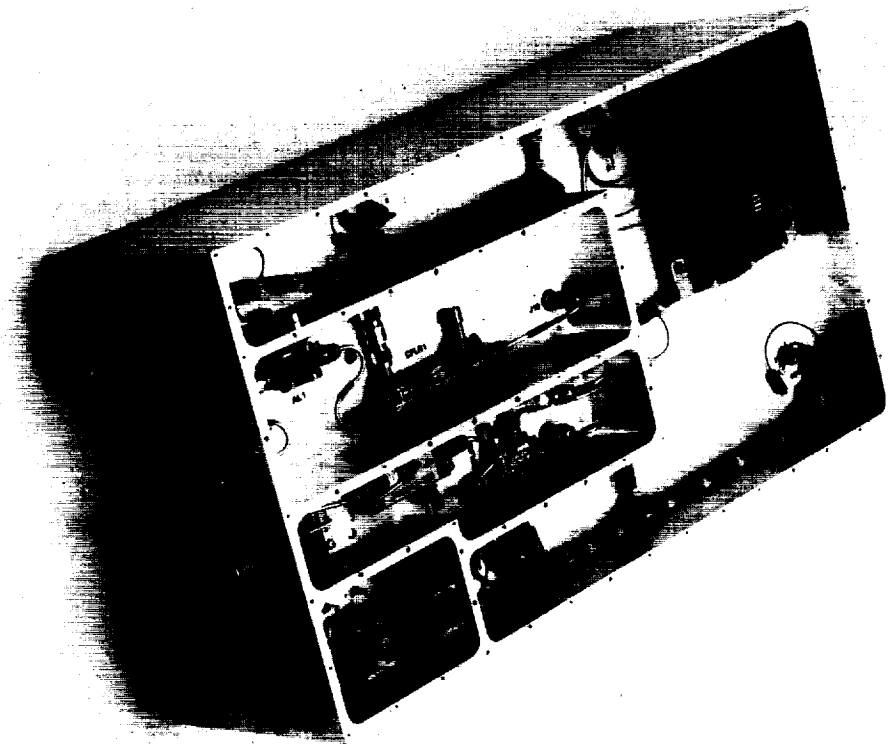


Fig. 5. RF/IF downconverter chassis fully assembled, bottom side.

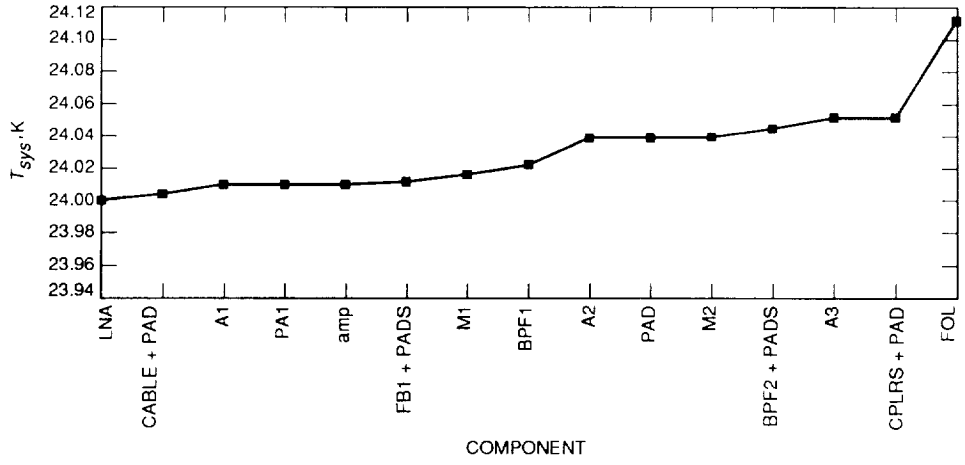


Fig. 6. RF/IF downconverter cumulative system temperature by component.

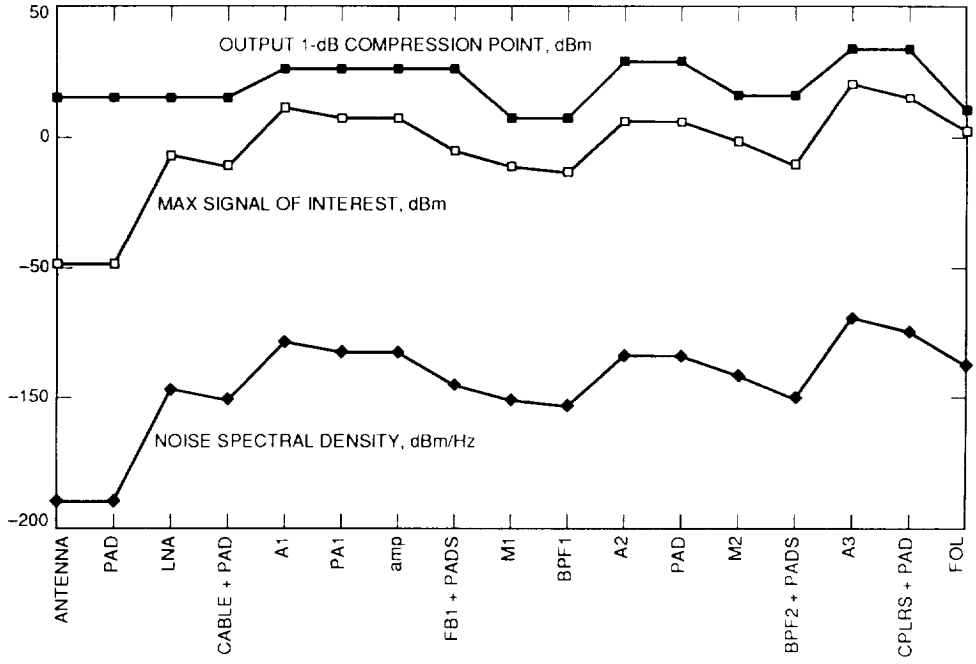


Fig. 7. RF/IF downconverter gain profile.

N94-35552

TDA Progress Report 42-117

May 15, 1994

A Closed-Loop Time-Alignment System for Baseband Combining

Y. Feria
Communications Systems Research Section

In baseband combining, the key element is the time alignment of the baseband signals. This article describes a closed-loop time-alignment system that estimates and adjusts the relative delay between two baseband signals received from two different antennas for the signals to be coherently combined. This system automatically determines which signal is advanced and delays it accordingly with a resolution of a sample period. The performance of the loop is analyzed, and the analysis is verified through simulation. The variance of the delay estimates and the signal-to-noise ratio degradation in the simulations agree with the theoretical calculations.

I. Introduction

Antenna array combining techniques have been used in the Deep Space Network to improve the signal-to-noise ratio (SNR) [1]. These techniques include full spectrum, baseband, and symbol stream combining. Each technique, however, has been applied only to a specific mission, and a rigorous comparison between the techniques is missing. To have a better understanding of how these techniques perform under the same or different conditions, all techniques are being studied and simulated in software.

This article concentrates on baseband combining where signals from two antennas have each had their carrier removed before going into the combiner. A diagram of baseband (BB) combining is shown in Fig. 1. The signals received from the ith antenna have the following form [1]:

$$\begin{aligned} r_i(t) &= \sqrt{2P_{C_i}}\sin\left(\omega_{c_i}(t-\tau_i)+\theta_{c_i}\right)+\sqrt{2P_{D_i}}d(t-\tau_i) \\ &\quad\times \operatorname{sgn}\left[\sin\left(\omega_{sc}(t-\tau_i)+\theta_{sc}\right)\right] \\ &\quad\times \cos\left(\omega_{c_i}(t-\tau_i)+\theta_{c_i}\right)+n_i(t) \end{aligned} \tag{1}$$

where $P_{C_i} = P_i\cos^2\Delta$ and $P_{D_i} = P_i\sin^2\Delta$, with P_i being the total signal power of the received signal from the ith antenna, and Δ is the modulation index. The parameters ω_{c_i} and θ_{c_i} denote the carrier angular frequency and phase of the received signal from the ith antenna, ω_{sc} and θ_{sc} denote the subcarrier angular frequency and phase, and τ_i is the relative delay between the reference signal and the ith signal. These received signals are analog-to-digital (A/D) converted and downconverted to an intermediate frequency (IF) and then sampled before going through the

carrier tracking in the model. After carrier removal, ignoring the filtering effects on the data and the squarewave subcarrier, the signals have the form

$$r_i'(k) = \sqrt{P_{D_i}} \cos \phi_{c_i} d(kT_s - \tau_i) \times \text{sgn}\left[\sin\left(\omega_{sc}(kT_s - \tau_i) + \theta_{sc}\right)\right] + n_i'(k) = s_i(k) + n_i'(k) \quad (2)$$

where T_s denotes the sampling period, and ϕ_{c_i} is the carrier phase error in the ith carrier-tracking loop. The signals $r_i'(k)$, $i = 1, 2, ...,$ are the to-be-combined baseband signals with different time-varying time delays. To coherently combine these signals, it is necessary to align them in time first.

To align two signals in time, first find the relative delay between the two signals. A standard procedure for finding a relative delay between two signals is to correlate them [2] since the cross-correlation is a (linear or nonlinear) function of the relative delay. The advanced signal will then be delayed by the amount computed from the cross-correlation. As mentioned before, the relative delay is not constant in time; therefore, to track the dynamics of the relative delay automatically, a closed loop is used. To close the loop, a number of components need to be added, e.g., a loop filter and an integrator.

This article presents a closed-loop time-alignment system, which takes two baseband signals, estimates the relative delay between the two signals using a correlator, and adjusts the time delay of the advanced signal through a loop filter and an integrator before combining the two baseband signals. The closed-loop time-alignment system, shown in Fig. 2, consists of a quadrature correlator, a loop filter, a unit delay, an integrator, a nearest integer device, and a decision-making device.

Each component in the loop has its function, as briefly described below and in detail in the subsections. The cross-correlator determines the relative delay between two signals. The loop filter tracks the dynamics of the relative delay and maintains the loop stability. The unit delay (z^{-1}) is required to close the loop. The integrator ($z/(z-1)$ transfer characteristic) combines the previously and currently computed relative delays. The nearest integer device is needed because the delay can be adjusted only by an integer number of samples. Finally, the decision-making device decides which signal needs to be delayed.

A similar time-alignment system called the real-time telemetry combiner (RTC) has been developed and analyzed [2,4,5]; however, this closed-loop time-alignment system overcomes the following RTC limitations:

(1) Since the length of the correlator in the RTC is fixed, the correlation feedback is not useful for low SNR and the loop has to be open.

(2) Because cross-correlation uses the product of the sign bits of the two signals, it causes the quantization-noise power to increase from −58.8 to −16.8 dB when compared with 8-bit quantization.

(3) Since the delay can be adjusted only in one branch, the to-be-adjusted branch must be delayed relative to the other branch. However, the signals' advance/delay relationship may change in time as the spacecraft moves and the Earth rotates.

(4) The RTC has an analog feedback to adjust delays, which is not suitable for software or completely digital hardware development.

The closed-loop time-alignment system presented here, on the other hand, has the following features:

(1) An adjustable correlation length to accommodate different SNRs.

(2) Cross-correlation, which is a product of the floating-point samples of the signals, reduces the quantization-noise power by 42 dB with 8-bit quantization.

(3) A decision-making device that automatically delays the advanced signal.

(4) Discrete-delay adjustments suitable for software or completely digital hardware development.

Although this time-alignment system can adjust delays only to the precision of an integer number of samples, accuracy is not a problem when the subcarrier is a perfect square wave and the data transitions are instantaneous. In practice, accuracy is not a problem if the signal is not filtered or the sample rate is much higher than the subcarrier frequency. In the case of a filtered signal sampled at a low rate, an added interpolation on the input waveforms can provide better resolution.

Due to all these differences, the RTC analysis does not apply to the closed-loop time-alignment system proposed

here; therefore, an analysis of the new system is needed. Because of the nonlinear device that takes the nearest integer in the loop, linear system theory, or analysis in the frequency domain or Z-domain, does not apply. The remaining choice is to use time-domain analysis for nonlinear systems even though it is not as straightforward as frequency-domain analysis for linear systems.

A system analysis will be given later, following a more detailed description of each of the components in the closed-loop time-alignment system. Simulation results and comparisons with the theoretical calculations will also be presented.

II. System Description

A block diagram of the closed-loop time-alignment system is shown in Fig. 2, and an equivalent diagram of the closed loop is shown in Fig. 3. In this section, each of the components and their functions are described in detail. The components include a cross-correlator, loop filter, unit delay, integrator, a device that takes the nearest integer, and a device that decides which signal to delay.

A. Cross-Correlator

To find the proportionality between cross-correlator output and relative delay, consider the noise-free case. The cross-correlator, shown in Fig. 4, computes the relative delay by taking the difference between the correlation of Signal 1, $s_1(k)$, with Signal 2 delayed by N_d samples, $s_2(k - N_d)$, and the correlation of Signal 2, $s_2(k)$, with Signal 1 delayed by N_d samples, $s_1(k - N_d)$; the result is divided by two. The correlator output becomes

$$c(m) = \frac{1}{2}\left[s_1(k)s_2(k - N_d) - s_1(k - N_d)s_2(k)\right] \quad (3)$$

where $s_1(k)$ and $s_2(k)$ are defined in Eq. (2).

This cross-correlator is similar to that in RTC [2,4,5], except the cross-correlation is floating point instead of sign bit, and discrete instead of continuous. By using correlations of 8-bit quantized samples, the quantization-noise power reduces from -16.8 dB of the single-bit correlation to -58.8 dB.

The analysis of this cross-correlator is similar to that in [2] except the cross-correlator is discrete instead of continuous. The proportionality of the cross-correlator output and the relative delay is given in Appendix A. Assuming that the signal power is normalized to 1, the correlator output in the linear region ($|m| < N_d$) is

$$\begin{aligned} c(m) &= \left[\frac{4f_{sc} - R_{sym}}{f_s}\right] m \\ &= am \end{aligned} \quad (4)$$

where f_{sc} is the subcarrier frequency, R_{sym} is the symbol rate, and f_s is the sampling rate. Dividing the correlator output by a, one obtains the relative delay m.

Appendix A shows that when the artificial delay, N_d, is one-quarter of the number of samples in a subcarrier period, N_{sc}, the linear region is the longest, which means that larger delays can be tracked. Hence, the correlator is called a quadrature correlator.

In a noisy environment, this output is further averaged over N correlator-output samples to reduce the noise effect.

B. Loop Filter

To close the loop, one needs to add a loop filter for dynamic tracking and loop stability. A higher order loop filter does better dynamic tracking. In this case, however, the relative delay is considered to have low dynamics; hence, only a first-order loop filter is used. That is, the filter transfer function is

$$F(z) = K$$

The choice of K influences the loop stability, loop SNR, and loop bandwidth. These issues will be discussed further in Section III.

C. Integrator, Unit Delay, and Decision Making

The integrator combines past and current delay computations; the unit delay is required to close the loop; and the decision-making device is basically a two-way switch that switches one way when input is positive and switches the other way when input is negative. This decision-making device allows the system to delay whichever signal is advanced.

III. System Analysis

To characterize the system quantitatively, a number of parameters need to be determined. Among them are the filter parameters for loop stability, the loop SNR, the loop bandwidth, and the SNR degradation due to misalignment. Except for loop stability, which depends on the filter parameters only, other parameters such as loop

SNR, loop bandwidth, and SNR degradation depend on the loop-output noise variance. This noise variance is analyzed following a brief analysis of loop stability.

A. Loop Stability

For stability analysis, the noise input in Fig. 3 is ignored because loop stability depends solely on the loop parameters and is input independent. Assume that the relative true delay is an integer multiple of the sample period. This assumption makes the nonlinear nearest integer device negligible in a noise-free environment. Note that when the delay is not an integer multiple of a sample period, the nearest integer device can be modeled as an additive quantization noise source of finite magnitude and thus can be ignored for the stability analysis.

Consider a first-order loop filter, $F(z) = K$. The open-loop transfer function is

$$G(z) = Kz^{-2}\frac{z}{z-1}$$

and the closed-loop transfer function is

$$H(z) = \frac{G(z)}{1+G(z)} = \frac{K}{z^2 - z + K} \tag{5}$$

A root-locus plot of this system with a first-order loop filter is shown in Fig. 5. From this plot, it can be observed that when $K \geq 1$, the root locus is outside the unit circle; thus, the loop is unstable. For $0.25 < K < 1$, the system is underdamped; for $K = 0.25$, it is critically damped; and for $0 < K < 0.25$, it is overdamped. Besides affecting system stability, the choice of K also affects the noise performance in the delay estimates, as will be discussed.

B. Noise Analysis

To analyze the noise, we first determine the amount of noise (mean and variance) at the output of the quadrature correlator in Fig. 4. This amount of noise is added at the output of the correlator, which acts as a unit delay in the closed loop, as shown in Fig. 6. The noise statistics at the correlator output are given next, followed by the noise statistics of the closed loop. Because of the nonlinear device of taking the nearest integer, the Z-domain linear system theory cannot be applied here, and time-domain analysis is used instead.

1. Noise Variance at Correlator Output. To determine the noise variance at the output of the correlator shown in Fig. 4, assume that the samples of the two inputs to the correlator consist of the sum of baseband signals, $s_1(k)$ and $s_2(k)$ as defined in Eq. (2), and independent samples of additive white Gaussian noise, $n_1'(k)$ and $n_2'(k)$, both with zero mean and variance σ_{in}^2. Denote the kth sample of the inputs of the quadrature correlator as

$$r_i'(k) = s_i(k) + n_i'(k) \tag{6}$$

where $i = 1, 2$.

The output of the correlator before "sum and dump" is

$$q(k) = \frac{1}{2}[r_1'(k)r_2'(k - N_d) - r_1'(k - N_d)r_2'(k)] \tag{7}$$

The sum-and-dump filter takes the average of its input over N samples, and its ouput is then divided by the slope a to produce an output sample

$$w_m = \frac{1}{aN}\sum_{k=mN+1}^{(m+1)N} q(k) \tag{8}$$

where a is defined in Eq. (4).

The problem here is to find the variance of w_m. With the assumption that $\{w_m\}$ is a stationary sequence and the noise samples at the inputs of the correlator are independent and identically distributed with zero mean, the variance of w_m is found to be

$$\sigma_w^2 = \frac{1}{(2a)^2 N}[2\sigma_{in_1}^2\sigma_{in_2}^2 + 4\sigma_{in_1}^2 P_2 + 4\sigma_{in_2}^2 P_1] \tag{9}$$

which is independent of m, and where $P_i, i = 1, 2$ is the ith signal power, and $\sigma_{in_i}^2$ is the ith noise variance at each correlator input. See Appendix B.

2. Noise Variance at Loop Output. An equivalent diagram of the closed loop is shown in Fig. 6, where noise is the input and the output corresponds to the location of the computed delay. Denote the kth noise samples at the input of the loop as w_k, those at the output of the integrator as x_k, and those after taking the nearest integer as y_k. The goal here is to find the variance at the output after the nearest integer device, that is, the variance of y.

The difference equation that describes the closed loop in Fig. 6 is

$$x_k - x_{k-1} = K(w_{k-1} - y_{k-2}) \tag{10}$$

Assuming that w and x are stationary Gaussian random processes with zero mean, then y also has zero mean, and the variances of w, x, and y, denoted by σ_w^2, σ_x^2, and σ_y^2, are found to have the following relationship (Appendix C):

$$\frac{2K}{1+K}\sigma_x^2 - K^2\sigma_y^2 = K^2\sigma_w^2 \tag{11}$$

The relationship between the variances of x and y is

$$\sigma_y^2 = \sum_k p_k k^2 \tag{12}$$

where

$$p_k = \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-x^2/(2\sigma_x^2)} dx \tag{13}$$

One can solve Eqs. (11), (12), and (13), for σ_y^2 for a given input noise variance σ_w^2. Note that in the absence of the nearest integer device, $\sigma_y^2 = \sigma_x^2$, and Eq. (11) becomes

$$\frac{2 - K^2 - K}{K^2 + K}\sigma_y^2 = \sigma_w^2 \tag{14}$$

which is exactly the same as Eq. (D-3) in Appendix D, in the linear system case.

Also note that for a large input noise variance ($\sigma_w^2 > 0.5$, observed from simulations), the noise generated by the nonlinear nearest integer device is approximated by a quantization noise with a uniform distribution and a variance of one-twelfth, or

$$\sigma_y^2 = \sigma_x^2 + \frac{1}{12}$$

C. Loop SNR and Loop Bandwidth

The loop SNR is usually defined as [1]

$$\rho = \frac{1}{\sigma_\tau^2} \tag{15}$$

where σ_τ^2 is the linear loop-output noise variance in units of squared radians.

However, the loop-output noise variance, σ_y^2, has units of squared number of samples. Converting the units to squared radians by equating a symbol duration to a cycle or 2π radians, one has

$$\sigma_\tau^2 = \left(\frac{2\pi R_{sym}}{f_s}\right)^2 \sigma_y^2 \tag{16}$$

where f_s is the sampling rate, and R_{sym} is the data rate. The loop SNR can be expressed as

$$\rho = \left(\frac{f_s}{2\pi R_{sym}}\right)^2 \frac{1}{\sigma_y^2} \tag{17}$$

The one-sided noise equivalent bandwidth for a linear system is defined as [4]

$$B_L = \frac{1}{2T_L} I_2 \tag{18}$$

where

$$I_2 = \frac{1}{|H(z)|^2} \frac{1}{2\pi j} \oint_{|z|=1} H(z)H(z^{-1}) \frac{1}{z} dz$$

$H(z)$ is the closed-loop transfer function, $T_L = NT_s$ is the correlation time, N is the number of samples in the correlation, and T_s is the original sample period. The loop input and output variances are related through I_2, that is, $var_{out} = I_2 var_{in}$. Note that var_{in} and var_{out} are not the same as σ_{in}^2 and σ_{out}^2 used previously.

In a nonlinear system, however, Z-domain analysis is no longer valid; thus, there is not an expression for $H(z)$, nor for I_2. Therefore, the loop bandwidth cannot be defined the same way as for linear systems. However, an expression for the relation between the input and output variances is available in Eq. (11); hence, the one-sided noise equivalent bandwidth is defined in this case as

$$B_L \equiv \frac{1}{2T_L} \frac{var_{out}}{var_{in}}$$

$$= \frac{1}{2T_L} \frac{\sigma_y^2}{\sigma_w^2} \tag{19}$$

where σ_y^2 and σ_w^2 have been defined in Sections III.B.1 and III.B.2.

D. SNR Degradation Due to the Loop

The magnitude reduction of the signal at the output of the sum-and-dump filter due to time misalignment in the baseband-combining scheme is (see Appendix E)

$$C_{\tau_i} = 1 + \left(\frac{1}{2} - 2\frac{f_{sc}}{R_{sym}}\right)\tau_i \tag{20}$$

where τ_i is the relative delay over a symbol period between the ith antenna and the reference antenna. Note that the expression in Eq. (20), derived in Appendix E, differs from the stated expression in Eq. 39 in [1] for the baseband-combining assembly containing the RTC. The first and second moments of the reduction function of the time-alignment loop are

$$\overline{C}_{\tau_i} = 1 + \left(\frac{1}{2} - 2\frac{f_{sc}}{R_{sym}}\right)\sqrt{\frac{2}{\pi}}\sigma_{\tau_i} \tag{21}$$

and

$$\overline{C_{\tau_i}^2} = 1 + 2\left(\frac{1}{2} - 2\frac{f_{sc}}{R_{sym}}\right)\sqrt{\frac{2}{\pi}}\sigma_{\tau_i} + \left(\frac{1}{2} - 2\frac{f_{sc}}{R_{sym}}\right)^2\sigma_{\tau_i}^2 \tag{22}$$

where $i = 1, 2$. Note that $\overline{C_{\tau_i}} \equiv 1$, and $\overline{C_{\tau_i}^2} \equiv 1$, which implies $\sigma_\tau = 0$.

Assuming perfect carrier and subcarrier tracking and symbol synchronization, the degradation due solely to the closed-loop time-alignment system is

$$D_{ta} = 10\log_{10}\left[\frac{\sum_{i=1}^{N_a}\gamma_i^2\overline{C_{\tau_i}^2} + \sum_{i,j,i\neq j}\gamma_i\gamma_j\overline{C}_{\tau_i}\overline{C}_{\tau_j}}{\Gamma^2}\right] \tag{23}$$

where $\gamma_i = (P_i/N_{o_i})/(P_1/N_{o_1})$, and $\Gamma = \sum_{i=1}^{N_a}\gamma_i$, with N_a being the number of antennas in the array, which is 2 in this case.

Having analyzed and defined all the necessary parameters, we now can check theoretical versus simulation results.

IV. Simulation Results

To verify the theoretical results, the P_t/N_o at the input of the cross correlator is fixed at 45 dB-Hz. The loop gain K is set at 0.25, and the sample rate at 500 kHz. The subcarrier frequency is 5 kHz, and the symbol rate is 1000 symbols/sec. There is no carrier in the simulation since the carrier is tracked and removed before baseband combining. The cross-correlation length N (or loop bandwidth B_L) is varied, and the loop SNR, ρ, is obtained from the theory and the simulations. The results are shown in Table 1 and the comparison of the results is shown in Fig. 7.

The simulation results also show that when the noise standard deviation at the input of the closed loop, σ_w, is larger than the one-sided length of the linear region, N_d, ($\sigma_w > N_d$) the loop fails to track. This can be observed in Fig. 7 for ρ smaller than 17 dB.

The SNR degradation due to the closed-loop time-alignment system is also measured through simulation and the results are very close to those obtained from Eq. (23), as shown in Table 2. The parameters in the simulations are as follows: Assume two identical antennas, each with $P_t/N_o = 23.54$ dB-Hz. The modulation index is 70 deg. The sampling rate is 100,000 Hz; the carrier frequency is 3 Hz; the subcarrier frequency is 1000 Hz; and the symbol rate is 200 symbols/sec. The time-alignment loop gain, K, is 0.0625. Simulations are done for two cases, with a perfect time alignment and with the closed-loop time-alignment system. The difference in SNR degradation is due to the closed-loop time-alignment system.

V. Conclusions

The article gives a full description of the design of a closed-loop time-alignment system and a detailed description of each of the elements in the system. It also provides an analysis of loop stability, noise effect, loop bandwidth, loop SNR in terms of the length of the correlation, and SNR degradation due to the loop. The closed-loop time-alignment system described here differs from the RTC in that this system is fully digital and is suitable for software and completely digital hardware implementation. The correlation is performed with floating-point samples of the to-be-combined signals rather than their sign bits. This causes the quantization-noise power to drop from -16.8 to

−58.8 dB in the correlation, assuming 8-bit quantization. The delay can be adjusted in either branch rather than a single branch only. The length of the cross-correlation can be adjusted for different levels of SNR. Because delays are adjusted digitally, an operation to take the nearest integer is needed, which results in a nonlinear system. An analysis of the nonlinear system is presented in the article. The noise analysis of this loop is significantly different from that of the RTC, due to the floating-point cross-correlation and the nonlinearity. The article also gives a well-matched comparison of theoretical and simulated results. The linear region of the relationship between the cross-correlation and the delay is determined. When the standard deviation of the delay estimates exceeds a quarter of the subcarrier period, the relationship between the cross-correlation and the delay is no longer linear, which corresponds to about a 17-dB loop SNR.

Although the time-alignment system described in this article aligns two signals, it can be easily expanded into the alignment of N_a signals, where $N_a > 2$, by aligning the signals pairwise in N layers, where $N = \lceil \log_2 N_a \rceil$. It can be shown that the total number of alignments is $N_a - 1$, which is the same as if we chose the signal with the longest delay as the reference, and adjusted all the delays of the $N_a - 1$ signals one by one as in [1]. Two advantages of aligning the signals pairwise using this scheme are

(1) No reference signal is needed as this reference may need to be changed in time.

(2) The SNRs get stronger after each layer, thus enabling the computation time to be reduced.

The combining degradation still needs to be analyzed.

Some simulation results show that this time-alignment system can be improved by low-pass filtering the to-be-correlated signals before the quadrature correlator. However, further analysis is needed to completely characterize the system and the associated improvement.

Acknowledgments

The author would like to thank Dr. Sam Dolinar, Dr. Sami Hinedi, Larry Howard, Alexander Mileant, Biren Shah, Dr. Marv Simon, and Dr. Steve Townes for their many helpful suggestions and discussions. The author is also very thankful to Samson Million for his assistance in implementing the system.

References

[1] A. Mileant and S. Hinedi, "Overview of Arraying Techniques in the Deep Space Network," *The Telecommunications and Data Acquisition Progress Report 42-104, vol. October–December 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 109–139, February 15, 1991.

[2] R. A. Winkelstein, "Analysis of the Signal Combiner for Multiple Antenna Arraying," *JPL Deep Space Network Progress Report 42-26, vol. January–February 1975*, Jet Propulsion Laboratory, Pasadena, California, pp. 102–118, April 15, 1975.

[3] E. I. Jury, *Theory and Application of the Z-Transform Method*, Malabar, Florida: Robert E. Krieger Publishing Co., 1964.

[4] M. K. Simon and A. Mileant, *Performance of the DSN Baseband Assembly (BBA) Real-Time Combiner (RTC)*, JPL Publication 84-94, Revision 1, Jet Propulsion Laboratory, Pasadena, California, May 1, 1985.

[5] L. D. Howard, "Prototype Real-Time Baseband Signal Combiner," *The Telecommunications and Data Acquisition Progress Report 42-60, vol. September–October 1980*, Jet Propulsion Laboratory, Pasadena, California, pp. 145–151, December 15, 1980.

Table 1. Performance of the time-alignment closed loop.

N	B_L, Hz	$\sigma_w^2(theo)$	$\sigma_y^2(theo)$	ρ_{theo}, dB	ρ_{simu}, dB	$\rho_{simu} - \rho_{theo}$, dB
25	1821.6	1303.6	237.47	14.26	3.17	−11.08
40	1157.1	814.8	150.85	16.23	9.82	−6.41
50	926.6	651.8	120.79	17.20	17.13	−0.07
100	463.7	325.9	60.45	20.20	20.13	−0.07
200	232.2	163.0	30.28	23.20	23.32	0.11
400	116.5	81.5	15.19	26.20	26.40	0.20
800	58.7	40.7	7.64	29.18	29.13	−0.05
1600	29.6	20.4	3.87	32.14	32.32	0.18

Table 2. SNR degradation due to the time-alignment closed loop.

N	B_L, Hz	ρ, dB	$\sigma_y^2(theo)$	$\sigma_y^2(simu)$	D_{theo}, dB	D_{simu}, dB
10,000	0.172	19.21	75.97	83.11	−0.92	−1.11
50,000	0.035	26.18	15.26	15.48	−0.39	−0.38
500,000	0.0036	35.96	1.60	2.45	−0.14	−0.13
5,000,000	4.58×10^{-4}	44.96	0.20	0.00	0.00	0.00

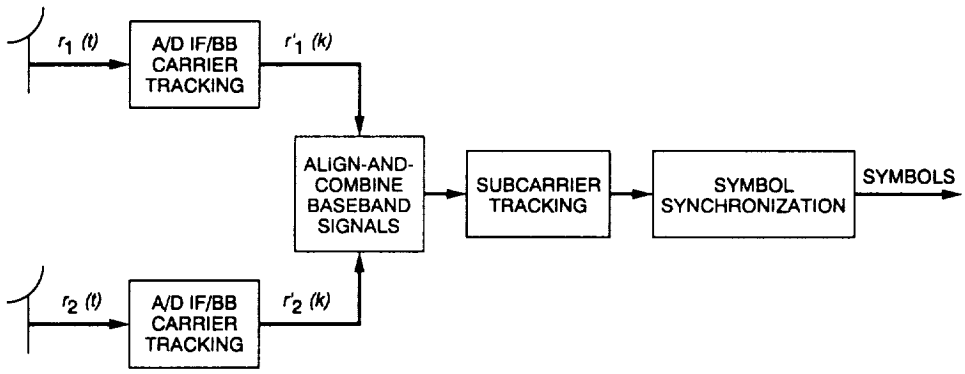


Fig. 1. A diagram of baseband combining of two signals.

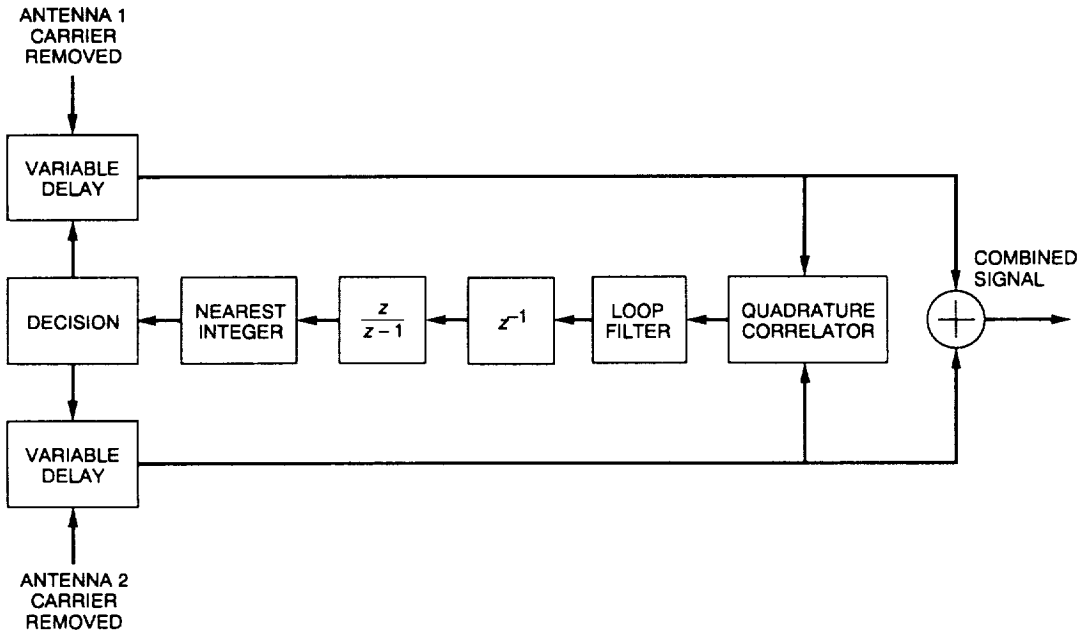


Fig. 2. A closed-loop time-alignment system.

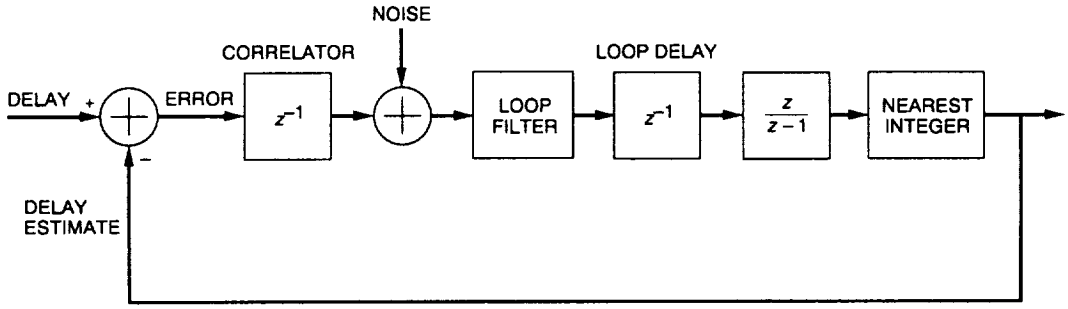


Fig. 3. An equivalent diagram of the closed-loop time-alignment system.

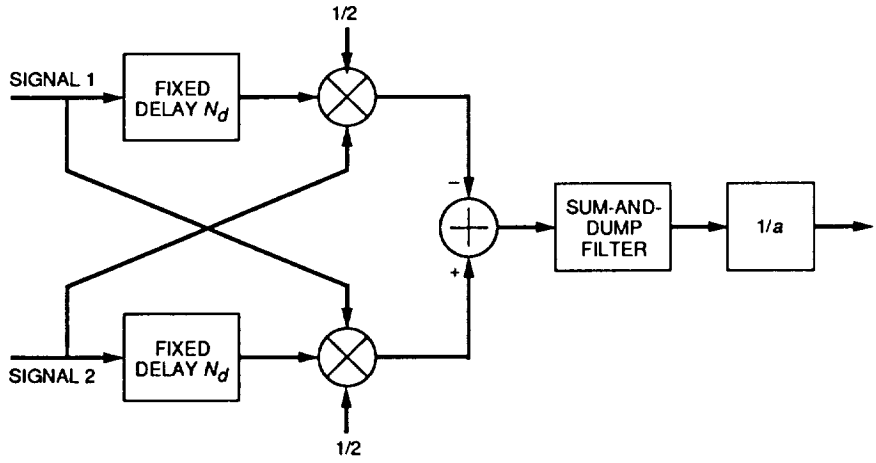


Fig. 4. A diagram of the quadrature correlator.

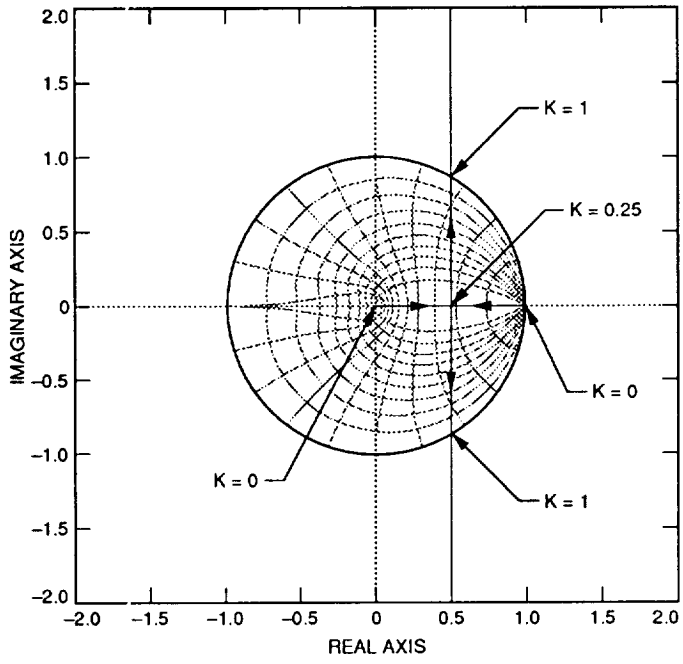


Fig. 5. Root locus for the first-order loop filter.

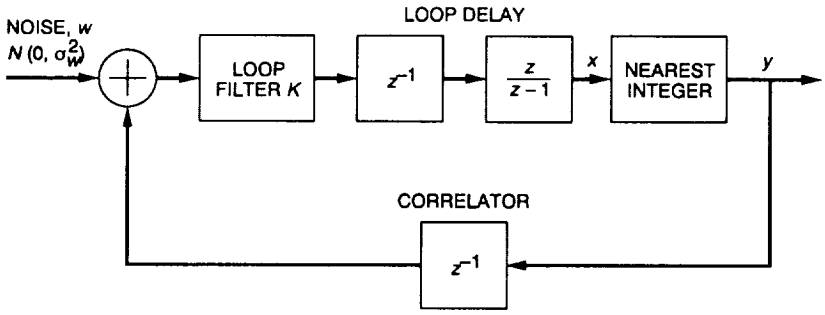


Fig. 6. The equivalent noise diagram.

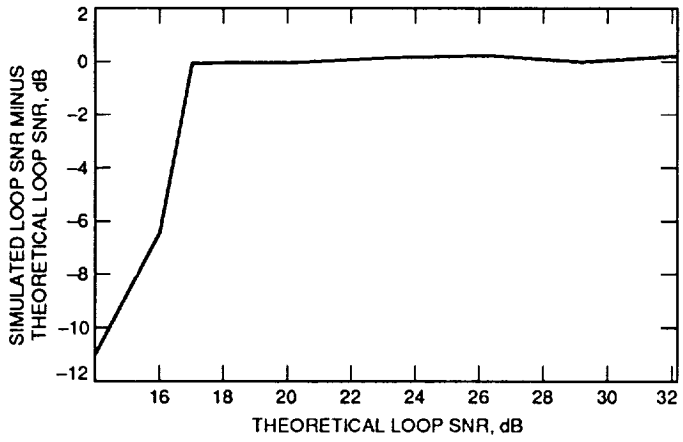


Fig. 7. Comparison of theoretical and simulated loop SNR.

Appendix A

Correlator Output Versus Relative Delay

The following gives the proportionality between the cross-correlator output and the relative delay between the signals received from two antennas. Let N_{sc} be the number of samples in a subcarrier period, N_{sy} the number of samples in a symbol period, and N_d the number of samples in the artificial delay. As stated before, a baseband signal consists of data on a square-wave subcarrier. Assuming that the data and the subcarrier are independent, the autocorrelation of the product of the data times the subcarrier equals the product of the autocorrelation of the data and that of the subcarrier. Considering the linear region $|m| < N_{sc}/2$, the data correlation function is

$$c_{data}(m) = 1 - \frac{|m|}{N_{sy}} \tag{A-1}$$

and the subcarrier correlation function is

$$c_{sc}(m) = 1 - \frac{|m|}{N_{sc}/4} \tag{A-2}$$

The correlation of data times subcarrier is

$$c_{(data)(sc)}(m) = 1 - \frac{|m|}{N_{sc}/4} - \frac{|m|}{N_{sy}} + \frac{m^2}{N_{sy}N_{sc}/4} \tag{A-3}$$

Then correlation curves are delayed or advanced by N_d samples, which is equivalent to the correlations of $s_1(k - N_d)s_2(k)$ or $s_1(k)s_2(k - N_d)$. That is, in the region of $[-N_{sc}/2 + N_d, N_d]$, the correlation function is

$$c_{(s_{1_{delayed}})} = 1 + \frac{m - N_d}{N_{sc}/4} + \frac{m - N_d}{N_{sy}} + \frac{(m - N_d)^2}{N_{sy}N_{sc}/4} \tag{A-4}$$

and in the region of $[-N_d, N_{sc}/2 - N_d]$

$$c_{(s_{2_{delayed}})} = 1 - \frac{m + N_d}{N_{sc}/4} - \frac{m + N_d}{N_{sy}} + \frac{(m + N_d)^2}{N_{sy}N_{sc}/4} \tag{A-5}$$

Finally, taking the difference of these two correlations and dividing the result by 2, one obtains an equation of the cross-correlator output in the linear region:

$$c(m) = \frac{4f_{sc} - R_{sym}}{f_s} m = am \tag{A-6}$$

where f_s is the sampling frequency, f_{sc} is the subcarrier frequency, and R_{sym} is the symbol rate.

It is clear that the longer the linear region, the larger the relative delay that can be tracked. The linear region is longest when the regions $[-N_{sc}/2 + N_d, N_d]$ and $[-N_d, N_{sc}/2 - N_d]$ coincide. That is, when

$$N_d = N_{sc}/4$$

Hence, the cross-correlator is also called a quadrature correlator.

Figure A-1 shows a special case where the number of samples per symbol period is chosen to be $N_{sy} = 10$, the number of samples per subcarrier period is $N_{sc} = 8$, and the number of samples in the delay is $N_d = N_{sc}/4 = 2$ to demonstrate the correlation of random data, a square-wave subcarrier, and their product. Note that N_{sy} and N_{sc} are arbitrarily chosen with a condition that $N_{sy} > N_{sc}$.

The first part of Fig. A-2 shows the correlation of $s_1(k - N_d)$ and $s_2(k) = s_1(k-m)$, which is the correlation of $s_1(k)$ and $s_2(k)$ shifted to the right by N_d samples. Similarly, the correlation of $s_1(k)$ and $s_2(k - N_d) = s_1(k - N_d - m)$ is the correlation of $s_1(k)$ and $s_2(k)$ shifted to the left by N_d samples. The second part of Fig. A-2 takes the difference between these two correlations and divides the result by 2, so that a linear region can be observed from $-N_d$ to N_d, with N_d being 2 in this case. Figure A-3 confirms that the maximum length of the linear region occurs when $N_d T_s$ is chosen to be one-quarter of the subcarrier period.

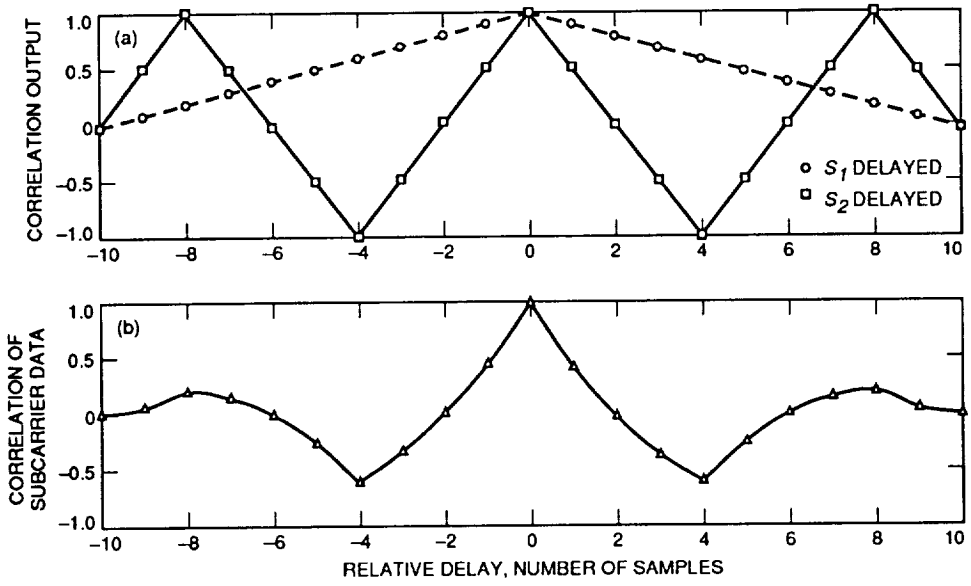


Fig. A-1. Autocorrelations of (a) PN to the square-wave subcarrier, and (b) to their products, $N_{sc} = 8$ and $N_{sy} = 10$.

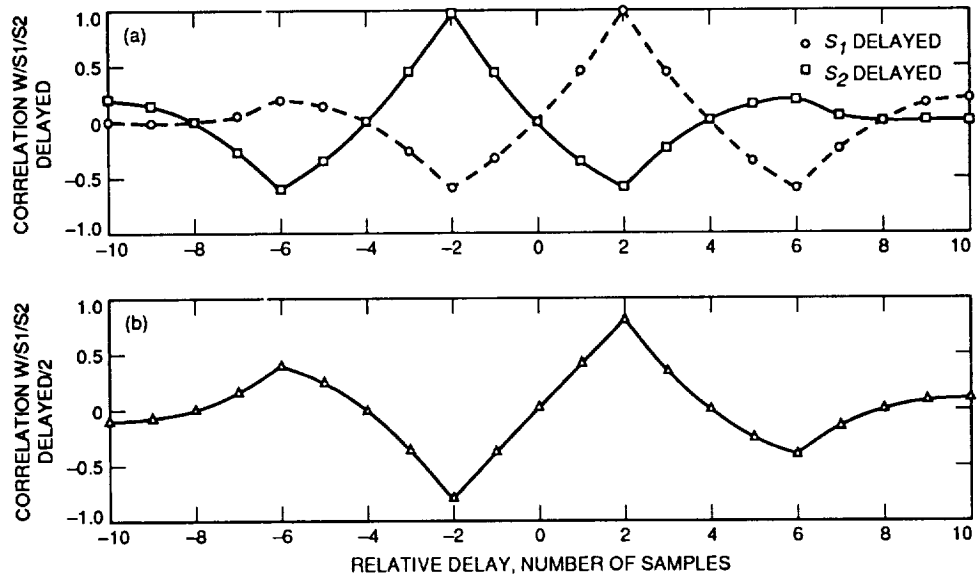


Fig. A-2. The correlation (a) signals S_1 and S_2 delayed and (b) their difference over 2.

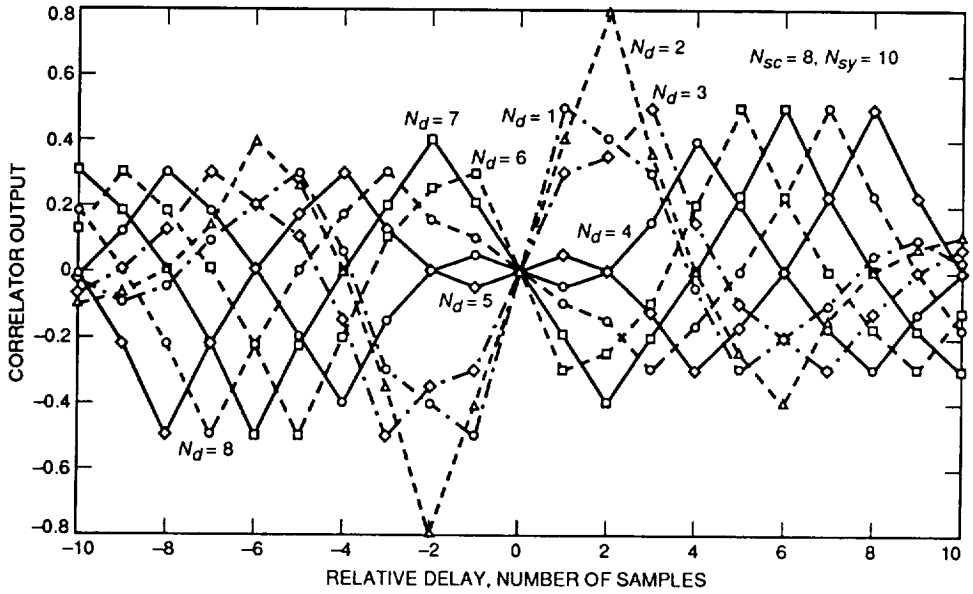


Fig. A-3. Linear-region length in the "S-curve" depends on N_d.

Appendix B

Noise Variance at the Correlator Output

To find the noise variance at the correlator output, we initially find the first and second moments of the correlation. Assuming that the correlation $\{w_m\}$ is stationary (and noting that the index m can be dropped), then its first moment is

$$\begin{aligned} E\{w\} &= \frac{1}{aN}\sum_{k=1}^{N} E\{q(k)\} \\ &= \frac{1}{2aN}\sum_{k=1}^{N} E\{r_1'(k)r_2'(k-N_d) - r_1'(k-N_d)r_2'(k)\} \\ &= \frac{1}{2aN}\sum_{k=1}^{N} E\{s_1(k)s_2(k-N_d) \\ &\quad - s_1(k-N_d)s_2(k)\} \end{aligned} \tag{B-1}$$

The square of the first moment is

$$\begin{aligned} (E\{w\})^2 &= \frac{1}{(2aN)^2}\sum_{k=1}^{N}\sum_{m=1}^{N} E\{s_1(k)s_2(k-N_d) \\ &\quad - s_1(k-N_d)s_2(k)\} \\ &\quad \times E\{s_1(m)s_2(m-N_d) \\ &\quad - s_1(m-N_d)s_2(m)\} \\ &\quad + E\{s_1(k)s_1(m)s_2(k-N_d)s_2(m-N_d)\} \\ &\quad - E\{s_1(k)s_1(m-N_d)s_2(k-N_d)s_2(m)\} \\ &\quad - E\{s_1(k-N_d)s_1(m)s_2(k)s_2(m-N_d)\} \\ &\quad + E\{s_1(k-N_d)s_1(m-N_d)s_2(k) \\ &\quad \times s_2(m)\} \end{aligned} \tag{B-2}$$

The second moment of w is defined as

$$E\{w^2\} = \frac{1}{(2aN)^2}\sum_{k=1}^{N}\sum_{m=1}^{N} E\{q(k)q(m)\} \tag{B-3}$$

Denote the variances of the noises at the correlator inputs as $\sigma_{in_1}^2$ and $\sigma_{in_2}^2$ and the signal powers as P_1 and P_2, respectively. Expanding the product $q(k)q(m)$, one obtains

$$\begin{aligned} E\{w^2\} &= \frac{1}{(2aN)^2}\sum_{k=1}^{N}\sum_{m=1}^{N} [2\sigma_{in_1}^2\sigma_{in_2}^2\delta_{m,k} \\ &\quad + 2\sigma_{in_2}^2 P_1\delta_{m,k} + 2\sigma_{in_1}^2 P_2\delta_{m,k} \\ &\quad - \sigma_{in_2}^2 E\{s_1(k)s_1(k-2N_d)\}\delta_{m,k-N_d} \\ &\quad - \sigma_{in_2}^2 E\{s_1(k+N_d)s_1(k-N_d)\}\delta_{m,k+N_d} \\ &\quad - \sigma_{in_1}^2 E\{s_2(k)s_2(k-2N_d)\}\delta_{m,k-N_d} \\ &\quad - \sigma_{in_1}^2 E\{s_2(k+N_d)s_2(k-N_d)\}\delta_{m,k+N_d} \\ &\quad + E\{s_1(k)s_1(m)s_2(k-N_d)s_2(m-N_d)\} \\ &\quad - E\{s_1(k)s_1(m-N_d)s_2(k-N_d)s_2(m)\} \\ &\quad - E\{s_1(k-N_d)s_1(m)s_2(k)s_2(m-N_d)\} \\ &\quad + E\{s_1(k-N_d)s_1(m-N_d) \\ &\quad \times s_2(k)s_2(m)\}] \end{aligned} \tag{B-4}$$

For a quadrature correlator, the delay N_d is one-quarter of the subcarrier period; thus, two points of the same signal that are $2N_d$ samples apart are one-half of a period apart, which means that they always have the same magnitude and the opposite sign. Therefore, their product is $-P_i$, $i = 1, 2$. Also note that the last four terms in Eq. (B-4) are the mean squared of w, which is subtracted from $E\{w^2\}$ to obtain the variance of w; hence, the variance of w is

$$\sigma_w^2 = \frac{1}{(2a)^2 N}[2\sigma_{in_1}^2\sigma_{in_2}^2 + 4\sigma_{in_2}^2 P_1 + 4\sigma_{in_1}^2 P_2] \tag{B-5}$$

Appendix C

Noise Variance at the Loop Output

The difference equation that describes the closed loop in Fig. 6 is

$$x_k - x_{k-1} = K(w_{k-1} - y_{k-2}) \tag{C-1}$$

Assuming that w and x are stationary zero-mean Gaussian processes, taking the expected value of Eq. (C-1), one obtains

$$E\{y\} = 0$$

Then the second moments of the processes w, x, and y are their variances denoted by σ_w^2, σ_x^2, and σ_y^2, respectively. Now taking the expected value on both sides of Eq. (C-1) squared,

$$E\{x_k^2\} - 2E\{x_k x_{k-1}\} + E\{x_{k-1}^2\} =$$

$$K^2(E\{w_{k-1}^2\} + E\{y_{k-2}^2\}) \tag{C-2}$$

That is

$$2\sigma_x^2 = K^2\sigma_w^2 + K^2\sigma_y^2 + 2E\{x_k x_{k-1}\} \tag{C-3}$$

The last term in Eq. (C-3) can be expressed in terms of the variance of x when multiplying Eq. (C-1) by x_{k-1} and taking the expected value. That is,

$$E\{x_k x_{k-1}\} = E\{x_{k-1}^2\} + KE\{w_{k-1}x_{k-1}\}$$

$$- KE\{x_{k-1}y_{k-2}\} \tag{C-4}$$

To see that $E\{w_{k-1}x_{k-1}\} = 0$, one can multiply Eq. (C-1) by w_k:

$$E\{w_n x_n\} = E\{w_n x_{n-1}\} + KE\{w_n w_{n-1}\}$$

$$- KE\{w_n y_{n-2}\} = 0 \tag{C-5}$$

The last step in the above equation follows since the past outputs are independent of the future inputs when the system is causal and the noises at different times are independent. Going back to Eq. (C-4), since y_{k-2} is the nearest integer of x_{k-2}, it can be expressed through an ϵ_{k-2} as

$$y_{k-2} = x_{k-2} + \epsilon_{k-2}$$

where ϵ_{k-2} represents a quantization error. Substituting the above into Eq. (C-4), one has

$$E\{x_k x_{k-1}\} = \sigma_x^2 - KE\{x_{k-1}x_{k-2}\}$$

$$- KE\{x_{k-1}\epsilon_{k-2}\} \tag{C-6}$$

The last term in Eq. (C-6) is zero since the past output ϵ_{k-2} is independent of the future input x_{k-1}. Applying the stationary property of x and rearranging Eq. (C-6), one obtains

$$E\{x_k x_{k-1}\} = \frac{1}{1+K}\sigma_x^2 \tag{C-7}$$

Substituting Eq. (C-7) into Eq. (C-3), one has

$$\frac{2K}{1+K}\sigma_x^2 - K^2\sigma_y^2 = K^2\sigma_w^2 \tag{C-8}$$

Appendix D

Noise Variance at Loop Output for Linear Cases

Without taking the nearest integer, the linear system theory should be applied. The variance at the output of the loop y equals the variance of the delay estimates due to the noise n, and it is

$$\sigma_y^2 = \frac{1}{|H(1)|^2}\sigma_w^2 I_2$$

where

$$I_2 = \frac{1}{2\pi j}\oint_{|z|=1} H(z)H(z^{-1})\frac{1}{z}dz$$

Using the first-order loop filter, the closed-loop transfer function is

$$H(z) = \frac{Kz}{z^2 - z - K} \tag{D-1}$$

Using the formula given in [3], I_2 is

$$I_2 = \frac{K^2 + K}{2 - K - K^2} \tag{D-2}$$

The variance of y is

$$\sigma_y^2 = \frac{K^2 + K}{2 - K - K^2}\sigma_w^2 \tag{D-3}$$

Appendix E

Signal Amplitude Reduction Due to Misalignment

Let N_{sc}, N_{sy}, and N_{rd} be the number of samples per subcarrier period, symbol duration, and in the relative delay. First, consider the ideal case where the signals are perfectly aligned, and the combined signal amplitude at the output of the sum-and-dump filter is

$$A_{ideal} = 2N_{sy}$$

Then consider adding two misaligned signals. The signal amplitude at the output of the sum-and-dump filter becomes

$$A_{mis} = \frac{1}{2}\left[2\left(\frac{N_{sc}}{2} - N_{rd}\right)\right] + \frac{1}{2}2\frac{N_{sc}}{2} + \left(2\frac{N_{sy}}{N_{sc}} - 1\right)2\left(\frac{N_{sc}}{2} - N_{rd}\right) \quad \text{(E-1)}$$

The amplitude reduction of the misaligned signal at the output of the sum-and-dump filter is

$$\begin{aligned} C_i &= \frac{A_{mis}}{A_{ideal}} \\ &= 1 - 2\frac{N_{sy}}{N_{sc}}\frac{N_{rd}}{N_{sy}} + \frac{1}{2}\frac{N_{rd}}{N_{sy}} \\ &= 1 + \left(\frac{1}{2} - 2\frac{N_{sy}}{N_{sc}}\right)\tau_i \\ &= 1 + \left(\frac{1}{2} - 2\frac{f_{sc}}{R_{sym}}\right)\tau_i \end{aligned} \quad \text{(E-2)}$$

where f_{sc} is the subcarrier frequency, R_{sym} is the symbol rate, and

$$\tau_i = \frac{N_{rd}}{N_{sy}}$$

is the relative delay between the ith signal and the reference signal in terms of a symbol period.

N94-35553

TDA Progress Report 42-117

May 15, 1994

Correlator Data Analysis for the Array Feed Compensation System

B. Iijima and D. Fort
Tracking Systems and Applications Section

V. Vilnrotter
Communications Systems Research Section

The real-time array feed compensation system is currently being evaluated at DSS 13. This system recovers signal-to-noise ratio (SNR) loss due to mechanical antenna deformations by using an array of seven Ka-band (33.7-GHz) horns to collect the defocused signal fields. The received signals are downconverted and digitized, in-phase and quadrature samples are generated, and combining weights are applied before the samples are recombined. It is shown that when optimum combining weights are employed, the SNR of the combined signal approaches the sum of the channel SNRs. The optimum combining weights are estimated directly from the signals in each channel by the Real-Time Block II (RTB2) correlator; since it was designed for very-long-baseline interferometer (VLBI) applications, it can process broadband signals as well as tones to extract the required weight estimates. The estimation algorithms for the optimum combining weights are described for tones and broadband sources. Data recorded in correlator output files can also be used off-line to estimate combiner performance by estimating the SNR in each channel, which was done for data taken during a Jupiter track at DSS 13.

I. Introduction

The advantages of array feed combining for recovering signal-to-noise ratio (SNR) lost to mechanical deformations of large receiving antennas have been described in previous articles [1,2]. Typically, SNR losses become significant when carrier wavelengths smaller than the design tolerance of the reflector are employed. This is the case in the DSN, where Ka-band (33.7-GHz) reception is contemplated with the large 34- and 70-m antennas, whose surfaces are subject to considerable deformation from gravity and wind.

The array feed compensation system currently being evaluated at DSS 13 has been designed to recover SNR losses at both low and high elevations, where losses are most severe. The idea is to collect some of the deflected

signal power using an array of feeds in the focal plane and to optimally recombine it for maximum SNR. In this implementation, combining is carried out on complex baseband samples using complex combining weights estimated in real time from the received signals. The estimates are obtained from 1-bit quantized samples processed by the Real-Time Block II correlator (RTB2) [3] over a specified time interval. The estimated weight vectors are supplied to a combining subsystem via an RS232 interface and used for combining over the following interval, during which time the next weight vector is estimated.

Independent confirmation of experimental results is always desirable, particularly when new concepts or techniques are analyzed. In the context of the array feed compensation system, the correlator can be used to predict combiner performance as well as to obtain real-time weight estimates from the 1-bit quantized bit stream. Direct comparison of the predicted and measured SNRs of the combined signal can be used to verify system performance. In the following sections, a simple expression will be derived for the SNR of the combined signal when accurate combining weights are available, and its implementation with recorded data will be described.

II. Optimum Combining Weights for Independent Noise

Suppose that the antenna points to a distant natural radio source (such as a planet or a quasar) that generates a broadband signal, and that an independent noise waveform is added to each channel of the array. The received RF signal in the kth channel may then be represented by

$$r_k(t) = s_k(t) + n_k(t) \qquad k = 1, 2, \ldots, K \tag{1a}$$

with source signal and background noise components

$$s_k(t) = \sqrt{2} S_k \left[a_c(t) \cos(\omega t + \theta_k) + a_s(t) \sin(\omega t + \theta_k)\right] \tag{1b}$$

$$n_k(t) = \sqrt{2} \left[n_{ck}(t) \cos(\omega t) + n_{sk}(t) \sin(\omega t)\right] \tag{1c}$$

where $a_c(t)$ and $a_s(t)$ are uncorrelated random processes representing the source signal, as are $n_{ck}(t)$ and $n_{sk}(t)$, which represent the background noise. The bandwidths of these random processes are assumed to be narrow compared to the center frequency ω; however, when downconverted to baseband, these processes are considered to be "broadband" signals. Since the source signals in the various channels differ from each other only in amplitude and phase (hence $a_c(t)$ and $a_s(t)$ are independent of k) and the time-varying envelopes are identical, the random processes $s_k(t)$ are correlated. The background noise processes are assumed to be uncorrelated in this analysis, as these consist of noise generated within the receivers plus background radiation arriving from different directions in space. (Actually, correlation coefficients on the order of 0.01 are typical between the feeds, probably due to near-field atmospheric noise; this issue is currently being investigated.) Following baseband downconversion and sampling, the inphase and quadrature samples may be represented by the complex process

$$\tilde{r}_k(i) = \tilde{s}_k(i) + \tilde{n}_k(i) \tag{2}$$

where $\tilde{s}_k(i) = \tilde{S}_k \tilde{a}(i)$, and $\tilde{S}_k = S_k e^{j\theta_k}$. Each component of the complex noise is independent with variance σ_k^2, but we shall assume that the real and imaginary components of $\tilde{a}(i)$ have variance $1/2$ to avoid introducing additional scaling. Thus, the SNR of $\tilde{r}_k(i)$ becomes $S_k^2/2\sigma_k^2$.

Let $\{\tilde{w}_k\}$ be a set of complex weights and form the sum

$$\tilde{z}(i) = \sum_{k=1}^{K} \tilde{r}_k(i) \tilde{w}_k \tag{3}$$

where $\tilde{z}(i)$ is the weighted sum. With

$$\tilde{s}_c(i) = \sum_{k=1}^{K} \tilde{s}_k(i) \tilde{w}_k$$

denoting the combined source signal component, the SNR of the combined sequence can be defined as

$$SNR_C = \frac{\langle |\tilde{s}_c(i)|^2 \rangle}{\mathrm{var}\left\{ \sum_{k=1}^{K} \tilde{w}_k \tilde{n}_k(i) \right\}} \tag{4}$$

where the numerator is the expected value of the magnitude squared of the combined signal, and the denominator is just the variance of the weighted noise sum. As shown in [1,2], the SNR of $\tilde{z}(i)$ is maximized when the complex weights $\{\tilde{w}_k\}$ are selected according to the formula

$$\tilde{w}_k = \frac{\tilde{S}_k^*}{2\sigma_k^2} \tag{5}$$

(The symbol * denotes a complex conjugate.) Substituting these complex combining weights into Eq. (4) yields

$$SNR_C = \frac{\left|\sum_{k=1}^{K} \tilde{S}_k \tilde{w}_k\right|^2}{\sum_{k=1}^{K} 2\left|\tilde{w}_k\right|^2 \sigma_k^2} = \sum_{k=1}^{K} \frac{S_k^2}{2\sigma_k^2} \tag{6}$$

Thus, the SNR of the combined sequence is equal to the sum of the individual channel SNRs when optimum combining weights are used. The problem is that the optimum weights are not known a priori, so they must be estimated and, in the presence of noise, are always subject to error.

The expression for optimum combining weights for a continuous wave source is very similar to the natural radio source case. The continuous wave signal in each channel can be written as

$$s_k(t) = \sqrt{2} S_k \cos\left[\phi_{cw}(t) + \theta_k\right]$$

where $\phi_{cw}(t)$ is the RF signal phase. Following baseband downconversion and sampling, the in-phase and quadrature samples of the continuous wave signal may be represented by the complex process $\tilde{s}_k(i) = \tilde{S}_k e^{j\phi_{BB}}$, where $\tilde{S}_k = S_k e^{j\theta_k}$ and $\phi_{BB}(t)$ is the baseband signal phase. With these definitions, Eqs. (2) through (6) apply in the continuous wave case as well as in the natural radio source case.

III. Array Feed Combining Demonstration Setup

A schematic of the array feed combining demonstration is shown in Fig. 1. The signals in the seven feeds are amplified using high electron mobility transistor low-noise amplifiers (HEMT LNAs) and are sent to the DSS-13 control room after downconversion to IF. (Typically, the IF frequency is in the 250-300 MHz range.) Each of the seven IFs is split in two signals, with one signal fed directly into the Mark III VLBI data-acquisition terminal (Mark III DAT), where it is downconverted to baseband. The second signal goes into a "trombone" before being downconverted to baseband. The trombone is a waveguide whose length can be adjusted so that, at IF, a narrowband signal passes through it 90 deg out of phase from the direct signal, giving rise to in-phase (I) and quadrature (Q) baseband signals for each of the seven feeds. After downconversion, the baseband signals are sent to the digital signal processor (DSP) subsystem for combining and to RTB2, where the combining weights are produced. The baseband signals are sampled at the Nyquist rate and single-bit quantized by the formatter of the Mark III DAT before being sent to RTB2. In the broadband signal case, the IF signal from the central feed is divided again and downconverted using a heterodyne frequency 10 kHz less than that used for the other 14 signals; this frequency-offset signal is also sent to RTB2. The 10-kHz offset ensures that both correlated signal amplitude and phase can be extracted from the correlation product, as in Eqs. (9a) through (9c). Finally, RTB2 produces complex combining weights, which are sent to the DSP where the 14 baseband signals are combined.

The real-time correlator is being used to compute (1) combining weights in real-time and (2) the SNR for the optimally combined array feed signal (performed off-line). The computation of the combining weights is described in Section IV. The computation of the expected optimally combined SNR is described in Section V.

IV. Computing the Combining Weights

The computation of the combining weights is discussed for two signal types: a continuous wave signal, such as a spacecraft carrier wave, and a broadband noise source.

RTB2 has two primary functions: (1) cross-correlating pairs of signal bit streams and (2) extracting tones (approximately sinusoidal signals) from individual signal bit streams. In the cross-correlation process, pairs of signal bit streams are multiplied together with a complex sinusoid and integrated. In the tone-extraction process, individual signal bit streams are multiplied by complex sinusoids and integrated. When weights are computed for a spacecraft carrier wave, tone extraction is used. When weights are computed for a broadband noise source, cross-correlation is used.

A. Computing Combining Weights for a Continuous Wave Source

Combining weights for a continuous wave source are computed as follows: RTB2 multiplies the in-phase signal

bit streams by the complex sinusoid $\exp(-j\phi_m(t))$ and integrates them, where $\phi_m(t)$ is a model for the signal phase, thus producing complex correlation sums for each feed. These have the expectation values [4]

$$\rho_k(t) = \sqrt{\frac{1}{\pi}}\sqrt{\frac{P_{sk}}{P_{nk}}}e^{j\theta_k}e^{j\phi_r(t)} \qquad k = 1,\ldots,7 \qquad (7)$$

where P_{sk} is the signal power for feed k; P_{nk} is the noise power for feed k; θ_k is a phase shift for feed k due to spatial propagation and electronic phase shifts [see Eq. (1b)]; and $\theta_k + \phi_r(t)$ is the residual of the true signal phase relative to the model signal phase, $\phi_m(t)$. The factor of π is a vestige of the single-bit quantized nature of the data going into RTB2. A low SNR assumption has been used in deriving Eq. (7). Three parameters are needed to compute the combining weights: each feed's SNR, phase shift, and average noise power. If the noise power P_{nk} is measured separately, the complex combining weights can be estimated as follows:

$$\tilde{w}_k = \frac{\rho_k^* \rho_1}{|\rho_1|^2}\sqrt{\frac{P_{n1}}{P_{nk}}} \qquad k = 1,\ldots,7 \qquad (8)$$

The overall phase and magnitude of the combining weights for the seven feeds are arbitrary and have been chosen so that the central feed (feed 1) has weight 1; this renders the weights dimensionless. In practice, the feeds have nearly identical noise power, so the last factor typically is ignored.

B. Computing Combining Weights for a Broadband Noise Source

Computation of combining weights for an unresolved far-field broadband noise source is somewhat more complicated than for continuous waves. Unlike the continuous-wave case, the single-bit signals from the feeds must be correlated with each other to measure the common noise signal from the source, leading to a requirement for additional information to get the SNR of each channel. Although not modeled in the derivation of the optimal weights above, an additional complication arises from the existence of low-level background signals common to the feeds (primarily near-field atmospheric noise) that must be accounted for when computing the weights. This is particularly important when observing weak sources, where the weight estimates could be corrupted by background noise correlation between the feeds.

RTB2 cross-correlates the signal from the frequency-offset central feed signal (described in Section III) with each of the seven in-phase signal bit streams. That is, each of the seven in-phase signal bit streams is multiplied by the frequency-offset bit stream and integrated. The product stream will contain a 10-kHz signal from the source (introduced by the frequency offset described earlier). Before integration, the product stream is multiplied by a 10-kHz complex sinusoid so that the source signal will integrate coherently. The results of the cross-correlation are correlation sums for each of the seven feeds with expectation values [5]

$$\rho_1 = A_1 e^{j\theta_1} e^{-j\theta_{freq\ offset}} \qquad (9a)$$

$$\rho_k = A_k e^{j\theta_k} e^{-j\theta_{freq\ offset}} + \rho_{bk} \qquad k = 2,\ldots,7 \qquad (9b)$$

where

$$A_k = \frac{a}{\pi}\sqrt{\frac{P_{sk}P_{s1}}{P_{Tk}P_{T1}}} \qquad k = 2,\ldots,7 \qquad (9c)$$

and $k = 1$ is the central feed. In feed k, P_{sk} is the source signal power and P_{Tk} is the total signal power (source power plus background noise power). The constant a equals 1.176 and is a vestige of the three-level sinusoid used by RTB2 in the cross-correlation process; $\theta_{freq\ offset}$ is the phase shift for the frequency-offset central feed signal; and ρ_{bk} is the correlation of the background noise in the central feed with the background noise in feed k. A low SNR assumption has been used in the derivation of Eq. (9c). The term ρ_1 is the correlation of the center feed signal with itself offset by 10 kHz; A_1 contains no SNR information since it results from the product of a single-bit quantized bit stream with itself.

In order to compute the combining weights, the background noise correlation ρ_{bk} must be removed from the correlation sums. Even after doing so, phase offset information is complete, but SNR information is not. These two problems are resolved by performing correlation and central-feed temperature measurements both on source and 84 mdeg off source in cross-elevation. The background noise correlation is nearly identical at angularly close points, so by performing the cross-correlation off source, ρ_{bk} can be measured and subsequently removed when back on source. Also, by measuring the central feed temperature on source and off source, the central feed SNR can be measured. With this information, it is then possible to compute the SNR in each feed.

The array feed combining experiments thus far have been performed during boresight tracks, in which the elevation boresights alternate with cross-elevation boresights.

Thus, on-source and slightly off-source pointing alternate regularly. The correlation ρ_{bk} is measured while off source and is used to calibrate the subsequent on-source correlation. This yields the calibrated on-source correlation sums

$$\rho_k^{cal} = A_k e^{j\theta_k} e^{-j\theta_{freq\ offset}} \qquad k = 1, \ldots, 7 \qquad (10)$$

Combining this with measurements of the central feed SNR and noise voltage measurements, one can compute the combining weights

$$\tilde{w}_k = \frac{\rho_k^{cal*} \rho_1}{\left|\rho_k^{cal} \rho_1\right|} \sqrt{\frac{P_{sk}/P_{nk}}{P_{s1}/P_{n1}}} \sqrt{\frac{P_{n1}}{P_{nk}}} \qquad k = 1, \ldots, 7 \qquad (11)$$

where the SNR in outer feed k is computed from

$$\frac{P_{sk}}{P_{nk}} = \frac{(P_{T1}/P_{s1})(\pi/a)^2 \left(\left|\rho_k^{cal}\right|^2 - 2\sigma_c^2\right)}{1 - (P_{T1}/P_{s1})(\pi/a)^2 \left(\left|\rho_k^{cal}\right|^2 - 2\sigma_c^2\right)} \qquad k = 2, \ldots, 7 \qquad (12)$$

and where σ_c^2 is a correlation sum variance, negligible for a strong source, which is subtracted to remove a positive bias on the square magnitude $\left|\rho_k^{cal}\right|^2$ (see Appendix). The overall magnitude and phase of the weights are arbitrary and have been chosen to set $\tilde{w}_1$ equal to 1. In practice, the feeds have nearly identical noise power, so that the last factor in Eq. (11) is typically ignored.

V. Computation of Optimal Combined SNR

In Eq. (6) it is shown that, when the optimal weights are used, the SNR for the combined signal is the sum of the individual feed SNRs. By combining the direct radiometric measurement of the central feed SNR and the calibrated correlation sums ρ_k^{cal}, one can compute the combined SNR expected for optimal weighting. Correlation sums were written to a postcorrelation record file and processed off-line to produce this combined SNR.

The above computations have been carried out for a Jupiter boresight track on 1993 day-of-year (DOY) 335 using the 34-m antenna at DSS 13. Plotted in Fig. 2 are the central feed SNR and the optimal combined SNR, the sum of the central feed SNR plus the six ring-feed SNRs as computed from Eq. (12). Also plotted are modifications to the SNR plot, eliminating the effects of elevation-dependent noise power and signal attenuation.

The elevation dependence in the noise power was obtained directly from the off-source power measurements. The elevation-dependent signal loss was estimated using station weather data (temperature, pressure, and humidity) during tracking, computing a zenith attenuation coefficient, and applying this value to an atmospheric loss model. Note that even at a 42.5-deg elevation the SNR of the combined channels (SNR_C) exceeds that of the central channel (SNR_1) by about 0.5 dB. (SNR_1 has been effectively "smoothed" on a greater time scale than the SNR_C curve.) The corrected combined signal-to-noise ratio SNR_C** decreases by only 0.54 dB at low elevations, while the corrected central channel loses about 1.12 dB of SNR at 9 deg. (This 34-m antenna is actually much better at Ka-band than the 70-m antennas used by the DSN, which suffer as much as a 5-dB SNR loss over the same elevation range [6]; hence, much greater SNR recovery can be expected for a 70-m antenna.) Thus, combining tends to provide uniform SNR over a range of elevations.

The rms scatter of the estimates of the combined SNR is roughly 0.035 dB. This is largely due to the inherent limit on accuracy imposed by the signal strength, the noise power in each channel, and the integration time. The variance of the estimates of the combined SNR (derived in the Appendix) is roughly

$$(\sigma_{SNR_C})^2 = \frac{3\pi^2}{2a^2 B T_{integration}} \frac{SNR_C - SNR_1}{SNR_1} \qquad (13)$$

where B is the recorded bandwidth and $T_{integration}$ is the integration time, which is comparable with the results shown in Fig. 2.

VI. Summary and Conclusions

A technique has been described for estimating the optimum combining weights for the array feed compensation system, both for received tones and broadband signals, observed in the presence of additive noise. In the derivation of the optimum combining weights, the noise components in the various channels were assumed to be independent, even though correlations on the order of 1 percent are routinely observed. These low-level correlations are believed to be primarily due to near-field atmospheric radiation; their presence could introduce a large error into the weight estimates, particularly when weak broadband sources are observed, i.e., when ρ_{bk} is comparable to A_k in Eq. (9b). Therefore, the noise correlations are measured slightly off source and subtracted from ρ_k before computing the combining weights. The combining weights thus

obtained can then be used for combining the signals from all of the feeds. Modifications to the combining weights to account for noise correlations are currently being examined.

The ability to compute channel SNRs based on the correlation coefficients is useful for estimating the SNR of the combined signal, which is simply the sum of the channel SNRs when accurate combining weights are available. (If the combining weights are in error, a loss in combined SNR will occur.) The above procedure has been carried out for a Jupiter track, from which we conclude that on the 34-m antenna at DSS 13 the combined channel loses only 0.54 dB in efficiency as the elevation varies from 43 deg down to 9 deg, after correction for elevation-dependent losses. This result confirms the ability of the array feed combining system to improve the efficiency response of large DSN antennas. The technique will be applied to subsequent tracks in order to accumulate data on combiner performance with a variety of sources under various conditions.

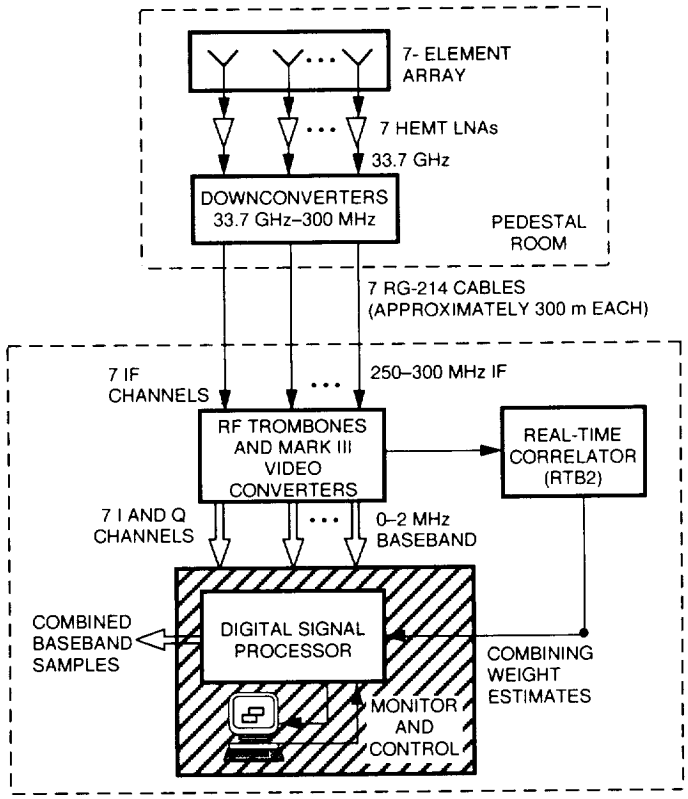


Fig. 1. Array feed combining system.

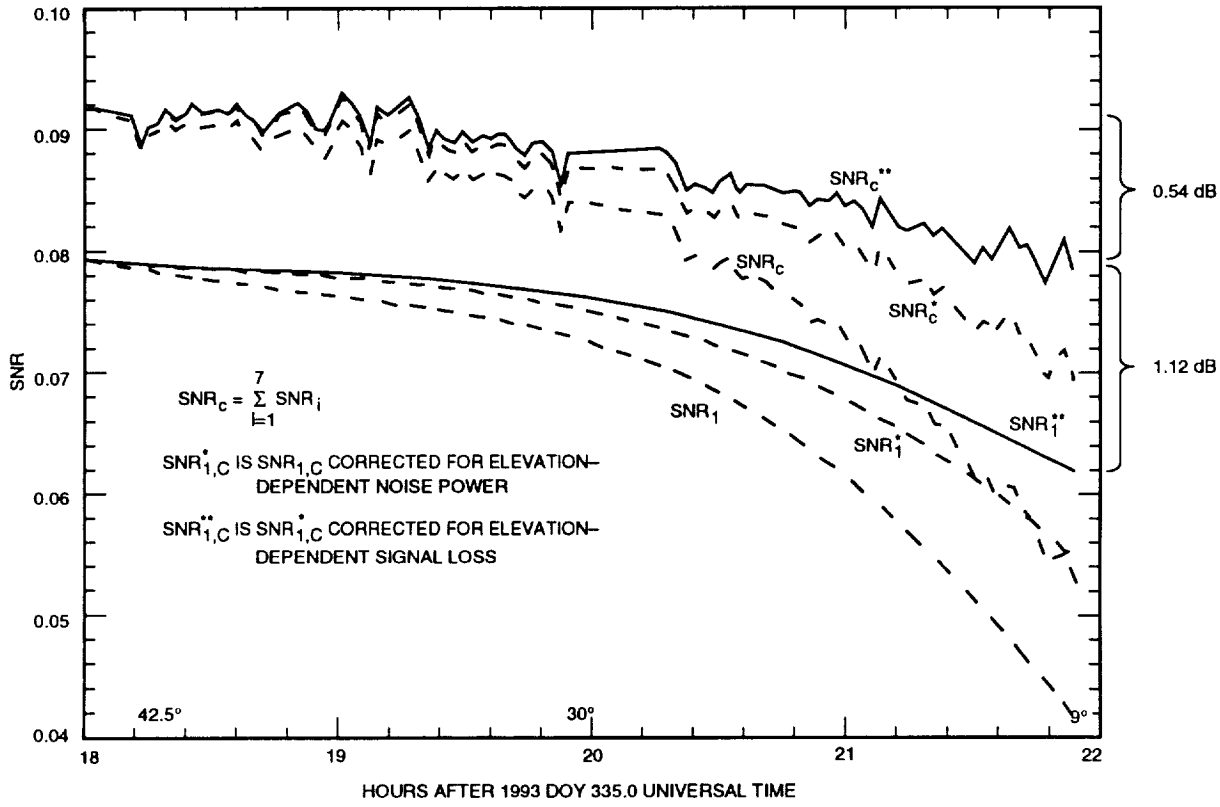


Fig. 2. SNR estimates for central (SNR_1) and combined (SNR_C) channels.

Appendix

Correlation Sum, Amplitude, and SNR Variances

The complex correlation sum ρ can be written as a two-component vector, $\rho = (c_0 + \delta c, s_0 + \delta s)$, where (c_0, s_0) is the expectation value of ρ, and δc and δs are deviations with variances $\langle \delta c^2 \rangle = \langle \delta c^2 \rangle = \sigma_c^2$ where $\langle\ \rangle$ indicates ensemble average. The correlation sum variance then can be written [7]

$$\sigma_c^2 = \frac{3}{8BT_{integration}} \tag{A-1}$$

where B is the recorded bandwidth and $T_{integration}$ is the integration time. The expected value of $|\rho|^2$ is

$$\langle |\rho|^2 \rangle = c_0^2 + s_0^2 + 2\sigma_c^2 = A^2 + 2\sigma_c^2 \tag{A-2}$$

where A is the amplitude in Eq. (9c) when ρ is the calibrated correlation sum in Eq. (10), so an unbiased estimator for A^2 is $|\rho|^2 - 2\sigma_c^2$. This is used in deriving Eq. (12). The variance of $|\rho|^2$ is

$$(\sigma_{|\rho|^2})^2 = 4A^2\sigma_c^2 \tag{A-3}$$

This is used in the derivation of combined SNR uncertainty below.

The estimate of the total SNR in the outer feeds is approximately

$$SNR_O = \sum_{k=2}^{7} \left(\frac{\pi}{a}\right)^2 A_i^2 \frac{1}{SNR_1} \tag{A-4}$$

and its variance is

$$(\sigma_{SNR_O})^2 = \sum_{k=2}^{7} \left[\left(\frac{\pi}{a}\right)^2 \frac{1}{SNR_1}\right]^2 (\sigma_{|\rho_i|^2})^2 \tag{A-5}$$

Combining Eqs. (A-1), (A-3), and (A-5) gives

$$(\sigma_{SNR_O})^2 = \frac{3\pi^2}{2a^2 BT_{integration}} \frac{SNR_O}{SNR_1}$$

which, when rewritten, is the optimal combined SNR variance given in Eq. (13).

References

[1] V. A. Vilnrotter, E. R. Rodemich, and S. J. Dolinar, Jr., "Real-Time Combining of Residual Carrier Array Signals Using ML Weight Estimates," *IEEE Transactions on Communications*, vol. 40, no. 3, pp. 604–615, March 1992.

[2] V. A. Vilnrotter and E. R. Rodemich, "A Digital Combining-Weight Estimation Algorithm for Broadband Sources With the Array Feed Compensation System," *The Telecommunications and Data Acquisition Progress Report 42-116, vol. October–December 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 86–97, February 15, 1994.

[3] C. Edwards, D. Rogstad, D. Fort, L. White, and B. Iijima, "The Goldstone Real-Time Connected Element Interferometer," *The Telecommunications and Data Acquisition Progress Report 42-110, vol. April–June 1992*, Jet Propulsion Laboratory, Pasadena, California, pp. 52–62, August 15, 1992.

[4] J. B. Thomas, "The Tone Generator and Phase Calibration in VLBI Measurements," *DSN Progress Report 42-44, January and February 1978*, Jet Propulsion Laboratory, Pasadena, California, pp. 63–74, April 15, 1978.

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

[6] M. Gatti, M. J. Klein, and T. B. H. Kuiper, "32-GHz Performance of the DSS-14 70-Meter Antenna: 1989 Configuration," *The Telecommunications and Data Acquisition Progress Report 42-99, vol. July–September 1989*, Jet Propulsion Laboratory, Pasadena, California, pp. 206–219, November 15, 1989.

[7] S. T. Lowe, *Theory of Post-Block II VLBI Observable Extraction*, JPL Publication 92-7, Jet Propulsion Laboratory, Pasadena, California, July 15, 1992.

Trace-Shortened Reed–Solomon Codes

R. J. McEliece
California Institute of Technology, Department of Engineering

G. Solomon[1]

Reed–Solomon (RS) codes have been part of standard NASA telecommunications systems for many years. RS codes are character-oriented error-correcting codes, and their principal use in space applications has been as outer codes in concatenated coding systems [6, Section 5.4.4]. However, for a given character size, say m *bits, RS codes are limited to a length of, at most,* 2^m. *It is known in theory that longer character-oriented codes would be superior to RS codes in concatenation applications, but until recently no practical class of "long" character-oriented codes had been discovered. In 1992, however, Solomon [4,5], discovered an extensive class of such codes, which are now called* trace-shortened Reed–Solomon (TSRS) codes. *In this article, we will continue the study of TSRS codes. Our main result is a formula for the dimension of any TSRS code, as a function of its error-correcting power. Using this formula, we will give several examples of TSRS codes, some of which look very promising as candidate outer codes in high-performance coded telecommunications systems.*

I. Construction Summary

In this section, we will summarize the construction of trace-shortened Reed–Solomon (TSRS) codes, and state our main result (Theorem 1), which is a formula for the binary dimension of an arbitrary TSRS code. In Section II, we will give some numerical examples. In Section III, we will state and prove a theorem (Theorem 2) that is more general than Theorem 1. Finally, in Section IV, we will summarize our results and list several open problems.

We begin with a summary of TSRS codes, as introduced in [4,5]. For any length n of the form $n = 2^m - 1$, any desired minimum distance $d \leq n - 1$, and any positive integer $\mu \leq m$, there is a TSRS code of length n and minimum distance d over the symbol alphabet $V(2^{m-\mu})$, the space of binary $(m - \mu)$-tuples. TSRS codes are constructed using properties of the Galois field $GF(2^m)$. The field $GF(2^{m-\mu})$ does not come into play in the construction, and so TSRS codes are not linear over the symbol field $GF(2^{m-\mu})$. However, they are linear over $GF(2)$, and the symbol-wise cyclic shift of any codeword is also a codeword. Our main result (Theorem 2) is a formula which allows the easy calculation of the binary dimension

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

of any TSRS code. We give several numerical examples, which show that TSRS codes are in some cases "almost maximum distance separable (MDS)," even though they are much longer than any true MDS code can be.

We begin with an ordinary Reed–Solomon (RS) code, as originally defined [3]. That is, with m and k_0 fixed, we consider the set $\mathcal{P}_m^{k_0}$ of polynomials of degree $k_0 - 1$ or less over the field $GF(2^m)$, and for each polynomial $P \in \mathcal{P}_m^{k_0}$, we define a length $n = 2^m - 1$ codeword $C(P)$ as follows:

$$C(P) = (P(1), P(\alpha), \ \ldots, \ P(\alpha^{n-1})) \tag{1}$$

where α is a primitive nth root of unity in $GF(2^m)$. The set of all such codewords, i.e., $\{C(P) : P \in \mathcal{P}_m^{k_0}\}$, is an (n, k_0, d) RS code over $GF(2^m)$, where $n = 2^m - 1$ and $d = n - k_0 + 1$.

The words in the above-constructed RS codes are vectors of length n over the field $GF(2^m)$. The next step in our construction is to "expand" each codeword into a length nm binary vector by representing each codeword symbol as a vector of length m over $GF(2)$. The particular coordinate basis we shall use for this representation is the basis which is dual to the "natural" basis $\{1, \alpha, \ \ldots, \ \alpha^{m-1}\}$. With respect to this dual basis, which we denote by $\{\beta_0, \beta_1, \ \ldots, \ \beta_{m-1}\}$, an element $x \in GF(2^m)$ has representation $x = x_0\beta_0 + \cdots + x_{m-1}\beta_{m-1}$, where the components x_h, which are elements of $GF(2)$, are given by the formula

$$x_h = \mathrm{Tr}_1^m(x\alpha^h) \qquad \text{for } h = 0, 1, \ \ldots, \ m-1 \tag{2}$$

In Eq. (2), and throughout the article, we use the symbol "Tr" to denote the trace operator. The super- and subscripts denote the subfields involved. Thus $\mathrm{Tr}_1^m(\xi)$, the trace of ξ from $GF(2^m)$ to $GF(2^1)$, represents the $GF(2)$-linear mapping from $GF(2^m)$ to $GF(2^1)$, given by

$$\mathrm{Tr}_1^m(\xi) = \xi + \xi^2 + \cdots + \xi^{2^{m-1}} \tag{3}$$

Similarly, if d is a divisor of m, $\mathrm{Tr}_d^m(\xi)$ denotes the trace of ξ from $GF(2^d)$ to $GF(2^m)$, defined as the $GF(2^d)$-linear mapping from $GF(2^m)$ to $GF(2^d)$, given by

$$\mathrm{Tr}_d^m(\xi) = \xi + \xi^{2^d} + \xi^{2^{2d}} + \cdots + \xi^{2^{(f-1)d}} \tag{4}$$

where $f = m/d$. Throughout the article, we will freely use the basic properties of the trace operator and the dual basis representation, which are summarized in [2, Chapter 8].

In any event, we call this "expanded" code the bit-mapped version of the RS code.

Now we can define the trace-shortened Reed–Solomon codes. Given the bit-mapped RS code defined above, for any integer μ in the range $0 \le \mu \le m$, we define the TSRS code of index μ to be the set of bit-mapped RS codewords for which the first μ binary components of each codeword symbol are zero. That is, if $C = (C_0, \ \ldots, \ C_{n-1})$ is a codeword in the original RS code, C is in the corresponding TSRS code of index μ if and only if

$$\mathrm{Tr}_1^m(\alpha^h C_i) = 0 \begin{cases} \text{for } h = 0, 1, \ \ldots, \ \mu - 1 \\ \text{for } i = 0, 1, \ \ldots, \ n - 1 \end{cases} \tag{5}$$

If we delete the μ guaranteed-zero binary components from each bit-mapped RS symbol, the TSRS code defined by Eq. (5) becomes a code of length n whose codeword components lie in the vector space $V(2^{m-\mu})$ of binary $(m-\mu)$-tuples. We note that while $V(2^{m-\mu})$ is not a field, it is nevertheless a group, viz, the elementary abelian group of order $2^{m-\mu}$ [1, Section 3.3]. Since this code is, by definition, a subcode of the parent RS code, its minimum distance is at least d. The bitwise sum of any two codewords satisfying Eq. (5) also satisfies Eq. (5), and so the code is a group code over the elementary abelian group $V(2^{m-\mu})$. Also, any (symbolwise) cyclic shift of any codeword in the TSRS code is also a codeword, so the code is a cyclic group code. The determination of the binary dimension of the code is somewhat challenging, and is the main result of this article. In the next paragraph, we will discuss our formula for this dimension.

As we have seen, the TSRS code of index μ over $GF(2^m)$ is a subgroup of the group $V(2^{m-\mu})^n$. The order of the code, i.e., the number of codewords, need not be a power of $2^{m-\mu}$, but since the code is linear over $GF(2)$, it must be a power of 2, say 2^{K_μ}. The following theorem gives a simple way to calculate K_μ. Before stating it, however, we need to introduce so-called cyclotomic cosets of the field $GF(2^m)$, which are the cycles of the permutation $i \to 2i \bmod (2^m - 1)$ on the set $\{0, 1, \ldots, 2^m - 2\}$ [2, Chapter 7]. We shall denote the size of the jth cyclotomic coset by d_j. For example, if $m = 4$, there are five such cyclotomic cosets:

j	C_j	d_j
0	(0)	1
1	(1, 2, 4, 8)	4
3	(3, 6, 12, 9)	4
5	(5, 10)	2
7	(7, 14, 13, 11)	4

Theorem 1. Denote by e_j the number of integers in the set $\{0, 1, \ldots, k_0 - 1\}$ which lie in the jth cyclotomic coset of $GF(2^m)$. Then the binary dimension of the index μ trace-shortened (n, k_0) RS code is given by the formula

$$K_\mu = \sum_j \max(me_j - \mu d_j, 0) \tag{6}$$

In Section II, we will give two extended numerical examples of TSRS codes. In Section III, we will give our proof of Theorem 1. Finally in Section IV, we will make some concluding remarks, and list several open problems about TSRS codes.

II. Examples

For our first example, we begin with the $(15, 9, 7)$ RS code over the field $GF(16)$. Here $m = 4$, $n = 15$, $k_0 = 9$. In the table below, we list the cyclotomic cosets of $GF(16)$, together with the numbers d_j and e_j, which are needed to apply Theorem 1. (For clarity, we list the numbers in the set $\{0, 1, \ldots, 8\}$ in boldface.)

j	C_j	d_j	e_j
0	**(0)**	1	1
1	**(1, 2, 4, 8)**	4	4
3	**(3, 6**, 12, 9)	4	2
5	**(5**, 10)	2	1
7	**(7**, 14, 13, 11)	4	1

From this table and using Eq. (6), the binary dimension K_μ is given by the formula

$$K_\mu = (4 - \mu)_+ + (16 - 4\mu)_+ + (8 - 4\mu)_+ + (4 - 2\mu)_+ + (4 - 4\mu)_+$$

where x_+ is short for $\max(x, 0)$. Thus we have

$$K_0 = 4 + 16 + 8 + 4 + 4 = 36$$

$$K_1 = 3 + 12 + 4 + 2 + 0 = 21$$

$$K_2 = 2 + 8 + 0 + 0 + 0 = 10$$

$$K_3 = 1 + 4 + 0 + 0 + 0 = 5$$

In other words, by "trace-shortening" the parent $(15, 9, 7)$ RS code, we obtain the codes over $V(8)$, $V(4)$, and $V(2)$, as shown in the table below. (In the following table, we extend the usual (n, k, d) notation for linear codes to the nonlinear TSRS codes by letting k denote the "pseudodimension" of the code, defined as $k_\mu = K_\mu/(m - \mu)$.)

Code parameters	Symbol group
$(15, 9, 7)$	$V(16)$
$(15, 7, 7)$	$V(8)$
$(15, 5, 7)$	$V(4)$
$(15, 5, 7)$	$V(2)$

Since the Singleton bound implies that any code with $n = 15$ and $d = 7$ must have $k \leq 9$, it follows that the $(15, 7, 7)$ code over $V(8)$ is close to optimal. The $(15, 5, 7)$ code over $V(4)$ is not as good as the $(15, 6, 7)$ generalized Bose–Chaudhari–Hocquenghem (BCH) code, but we could have obtained a code with the same parameters by starting with an alternate RS code. Finally, the $(15, 5, 7)$ code over $V(2)$ has the same parameters as the $(15, 5, 7)$ BCH code.

As our second example, we start with the $(31, 27, 5)$ RS code over $GF(32)$. The distribution of the numbers in the set $\{0, 1, \ldots, 26\}$ is shown in the following table:

j	C_j	d_j	e_j
0	**(0)**	1	1
1	**(1, 2, 4, 8, 16)**	5	5
3	**(3, 6, 12, 24, 17)**	5	5
5	**(5, 10, 20, 9, 18)**	5	5
7	**(7, 14**, 28, **25, 19)**	5	4
11	**(11, 22, 13, 26, 21)**	5	5
15	**(15**, 30, 29, 27, **23)**	5	2

Thus we have from Theorem 1,

$$K_0 = 5 + 25 + 25 + 25 + 20 + 25 + 10 = 135$$

$$K_1 = 4 + 20 + 20 + 20 + 15 + 20 + 5 = 104$$

$$K_2 = 3 + 15 + 15 + 15 + 10 + 15 + 0 = 73$$

$$K_3 = 2 + 10 + 10 + 10 + 5 + 10 + 0 = 47$$

$$K_4 = 1 + 5 + 5 + 5 + 0 + 5 + 0 = 21$$

This leads to the following list of codes:

Code parameters	Symbol group
$(31, 27, 5)$	$V(32)$
$(31, 26, 5)$	$V(16)$
$(31, 24.33, 5)$	$V(8)$
$(31, 23.5, 5)$	$V(4)$
$(31, 21, 5)$	$V(2)$

This list contains some interesting codes. In particular, the $(31, 26, 5)$ code over $V(16)$ is optimal, and in fact is "nearly" MDS, despite the fact that it is twice as long as any MDS code over a 16-symbol alphabet.

Let us conclude this section with a few bigger examples, omitting the detailed computations. Beginning with a $(511, 478, 34)$ RS code over $GF(2^9)$, for which the parity-check polynomial is $h(x) = \prod_{j=1}^{478}(1 - \alpha^i x)$, and trace-shortening it with $\mu = 1$, one obtains a $(511, 474, 34)$ TSRS code over the symbol group $V(2^8)$. Similarly, beginning with a $(511, 470, 42)$ RS code over $GF(2^9)$, with $h(x) = \prod_{j=1}^{470}(1 - \alpha^i x)$, again taking $\mu = 1$, one obtains a $(511, 465, 42)$ code over $V(2^8)$. Preliminary calculations performed by Dr. Fabrizio Pollara of the Communications Systems Research Section indicate that when concatenated with the NASA standard $(7, 1/2)$ convolutional code, both of these TSRS codes give an overall performance which is superior to that of the standard $(255, 223)$ RS code.

III. Proof of Theorem 1

In this section we will state and prove a theorem which is slightly more general than Theorem 1.

We begin with the field $F = GF(2^m)$, a positive integer n which is a divisor of $2^m - 1$, and a primitive nth root of unity in F, say α. Let J be a subset of $\{0, 1, \ldots, n-1\}$, with $|J| = k_0$. We then define the code $\mathcal{C}_J$ to be the (n, k_0) cyclic code over F, with check polynomial $h(x) = \prod_{j \in J}(1 - \alpha^j x)$. Equivalently, $\mathcal{C}_J$ consists of all vectors $\mathbf{C} = (C_0, C_1, \ldots, C_{n-1})$ of the form

$$C_i = \sum_{j \in J} c_j \alpha^{ij} \qquad i = 0, 1, \ldots, n-1 \tag{7}$$

where (c_j), for $j \in J$, is an arbitrary set of elements of F, indexed by J. For future reference, we denote the minimum distance of the code $\mathcal{C}_J$ by d_J. Note that if the elements of J form an arithmetic progression modulo n, whose increment is relatively prime to n, then $d_J = n - k_0 + 1$ by the BCH argument, and the code $\mathcal{C}_J$ is a (generalized) Reed–Solomon code. For example, if $n = 2^m - 1$ and $J = \{0, 1, \ldots, k_0 - 1\}$, the code $\mathcal{C}_J$ is the same as the RS code defined in Eq. (1).

We now define the μth-order trace-shortening of $\mathcal{C}_j$, denoted by $\mathcal{C}_J^\mu$, to be the set of codewords $\mathbf{C} \in \mathcal{C}_J$ such that

$$\mathrm{Tr}_1^m(\alpha^h C_i) = 0 \quad \begin{cases} h = 0, 1, \ldots, \mu - 1 \\ i = 0, 1, \ldots, n - 1 \end{cases} \tag{8}$$

In words, $\mathcal{C}_J^\mu$ consists of all codewords of $\mathcal{C}_J$ for which the first μ binary components (with respect to the basis for $GF(2^m)$ over $GF(2)$, which is dual to the basis $\{1, \alpha, \ldots, \alpha^{m-1}\}$) of each codeword symbol are zero. If we ignore these μn guaranteed-zero components, the code $\mathcal{C}_J^\mu$ becomes a code of length n over the set $V(2^{m-\mu})$ of binary $(m - \mu)$-tuples.

We denote the minimum symbol distance of the code $\mathcal{C}_J^\mu$ by d_J^μ. Since every codeword in $\mathcal{C}_J^\mu$ is also a codeword in the parent code $\mathcal{C}_j$, the minimum symbol distance of $\mathcal{C}_J^\mu$ cannot be less than that of $\mathcal{C}_J$, i.e., $d_J^\mu \geq d_J$. This simple bound is all we will have to say about the distance properties of the code $\mathcal{C}_J^\mu$. Our main results concern the *size* of the code.

Because the parent code $\mathcal{C}_J$ is a linear, cyclic code over $GF(2^m)$, it follows that the code $\mathcal{C}_J^\mu$ is closed under the operations of addition of any two codewords, and takes the (symbolwise) cyclic shift of any individual codeword. However, since the field $GF(2^{m-\mu})$ does not come into play, $\mathcal{C}_J^\mu$ cannot be a linear code. Nevertheless, we will define a pseudodimension for $\mathcal{C}_J^\mu$.

Since, as observed, the sum of any two codewords from $\mathcal{C}_J^\mu$ is another codeword, $\mathcal{C}_J^\mu$ is a linear code over $GF(2)$. Let us denote its $GF(2)$-dimension by K_μ. (Thus $K_0 =$

$k_0 \times m$, $K_m = 0$.) Then the pseudodimension of $\mathcal{C}_J^\mu$ is defined as

$$k_\mu = \frac{1}{m-\mu} K_\mu \tag{9}$$

Thus if $|\mathcal{C}_J^\mu|$ denotes the number of codewords in $\mathcal{C}_J^\mu$, we have that

$$|\mathcal{C}_J^\mu| = 2^{(m-\mu)k_\mu}$$

The main result of this article (Theorem 2, below), is the determination of K_μ, and hence also k_μ, for $\mu = 0, 1, \ldots, m$. To state the result, we need to define the modulo n cyclotomic cosets.

Let n be an odd positive integer. If i and j are integers in the range $0 \le i \le n-1$, and if $2^s i \equiv j \pmod{n}$ for some integer s, we say that i and j are conjugate modulo n. It is easy to see that conjugation modulo n is an equivalence relation on the set $\{0, 1, \ldots, n-1\}$, and so the set $\{0, 1, \ldots, n-1\}$ is partitioned into a number of disjoint equivalence classes, which are called the modulo n cyclotomic cosets. Alternatively, the cyclotomic coset containing j, which we will denote by Γ_j, can be described explicitly as the set $\{j, 2j, \ldots, 2^{d-1}j\}$, where d is the least positive integer such that $2^d j \equiv j \pmod{n}$. The integer d is called the degree of j, written $d = \deg(j)$. In what follows, we will denote the cardinality of Γ_j by d_j. It is easy to see that every element of Γ_j has degree d_j, and that d_j is a divisor of n. Finally, we denote by I_n the set consisting of the smallest integers in each cyclotomic coset.

Example 1. Let $n = 15$. A short calculation shows that there are five cyclotomic cosets modulo 15; indeed, we have $I_{15} = \{0, 1, 3, 5, 7\}$, and

$$\begin{array}{ll} \Gamma_0 = (0) & d_0 = 1 \\ \Gamma_1 = (1, 2, 4, 8) & d_1 = 4 \\ \Gamma_3 = (3, 6, 12, 9) & d_3 = 4 \\ \Gamma_5 = (5, 10) & d_5 = 2 \\ \Gamma_7 = (7, 14, 13, 11) & d_7 = 4 \end{array}$$

Thus $\deg(0) = 1$, $\deg(1) = \deg(2) = \deg(4) = \deg(8) = 4$, etc. ❐

Now we can give our formula for the binary dimensions of the codes $\mathcal{C}_J^\mu$. For each $j \in I_n$, we define $J_j = J \cap \Gamma_j$, and $e_j = |J_j|$. (Compare to Theorem 1.)

Theorem 2. The binary dimension K_μ of the code $\mathcal{C}_J^\mu$ is given by the formula

$$K_\mu = \sum_{j \in I} (m e_j - \mu d_j)_+ \tag{10}$$

where $(x)_+ = \max(x, 0)$.

To prove Theorem 2, we need several Lemmas. If n is a divisor of $2^m - 1$, let $P(x)$ be a polynomial of degree $n-1$ in the indeterminate x, with coefficients in $F = GF(2^m)$:

$$P(x) = \sum_{j=0}^{n-1} P_j x^j, \qquad P_j \in F \tag{11}$$

Now define the polynomial $\mathcal{P}(x)$ as follows:

$$\mathcal{P}(x) = \mathrm{Tr}_1^m(P(x)) \bmod (x^n - 1)$$

$$= \sum_{j=0}^{n-1} \mathcal{P}_j x^j \tag{12}$$

where in Eq. (12) it is understood that $\mathrm{Tr}_1^m(\xi) = \xi + \xi^2 + \cdots + \xi^{2^{m-1}}$, as in Eq. (3).

Example 2. Let $m = 4$, $n = 5$. If $P(x) = P_0 + P_1 x + P_2 x^2 + P_3 x^3 + P_4 x^4$, then $\mathcal{P}(x) = P(x) + P(x)^2 + P(x)^4 + P(x)^8 \bmod (x^5 - 1) = 1 \times (P_0 + P_0^2 + P_0^4 + P_0^8) + x \times (P_1 + P_3^2 + P_4^4 + P_2^8) + x^2 \times (P_2 + P_1^2 + P_3^4 + P_4^8) + x^3 \times (P_3 + P_4^2 + P_2^4 + P_1^8) + x^4 \times (P_4 + P_2^2 + P_1^4 + P_3^8)$. Thus,

$$\mathcal{P}_0 = P_0 + P_0^2 + P_0^4 + P_0^8$$

$$\mathcal{P}_1 = P_1 + P_3^2 + P_4^4 + P_2^8$$

$$\mathcal{P}_2 = P_2 + P_1^2 + P_3^4 + P_4^8$$

$$\mathcal{P}_3 = P_3 + P_4^2 + P_2^4 + P_1^8$$

$$\mathcal{P}_4 = P_4 + P_2^2 + P_1^4 + P_3^8$$ ❐

Lemma 1. Let $P(x)$ be a polynomial of degree $n-1$, as defined in Eq. (11). Then $\mathrm{Tr}_1^m(P(x)) = 0$ for all $x \in \{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ if and only if $\mathcal{P}_j = 0$ for $j = 0, 1, \ldots, n-1$.

Proof: Since $x^n = 1$ for all $x \in \{1, \alpha, \ldots, \alpha^{n-1}\}$, it follows from Eq. (12) that $\mathrm{Tr}_1^m(P(x)) = 0$ for all $x \in \{1, \alpha, \ldots, \alpha^{n-1}\}$ iff $\mathcal{P}(x) = 0$ for all $x \in \{1, \alpha, \ldots, \alpha^{n-1}\}$. But $\deg \mathcal{P}(x) \leq n-1$, so that $\mathcal{P}(x)$ must be identically 0. ❐

The next lemma gives an explicit formula for the coefficients $\mathcal{P}_j$ of $\mathcal{P}(x)$.

Lemma 2. For $j \in \{0, 1, \ldots, n-1\}$, if $d = \deg(j)$, then

$$\mathcal{P}_j = \mathrm{Tr}_d^m \left(\sum_{g \in \Gamma_j} P_g^{2^i} \right) \tag{13}$$

Note: For a fixed j, for each $g \in \Gamma_j$, there exists a unique integer i in the set $\{0, 1, \ldots, d-1\}$ such that $2^i g \bmod n = j$. The summation in Eq. (13) is to be understood to be over all pairs (g, i) such that $g \in \Gamma_j$, $i \in \{0, 1, \ldots, d-1\}$, and $2^i g \bmod n = j$. We will observe this same convention in summations like that in Eq. (13) in the remainder of the article.

Example 3. With $n = 15$ and $j = 10$, we have from Example 1 $\Gamma_{10} = \{5, 10\}$, and so Lemma 2 implies that $\mathcal{P}_{10} = \mathrm{Tr}_2^4(P_5^2 + P_{10}) = (P_5^2 + P_{10}) + (P_5^2 + P_{10})^4 = P_5^2 + P_{10} + P_5^4 + P_{10}^4$. ❐

Proof of Lemma 2: By definition (Eq. (12)),

$$\mathcal{P}(x) = \mathrm{Tr}_1^m(P(x)) \bmod x^n - 1$$

$$= \sum_{g=0}^{n-1} \mathrm{Tr}_1^m(P_g x^g) \bmod x^n - 1$$

Since the exponents appearing in the expansion of $\mathrm{Tr}_1^m(P_g x^g)$ are exactly those which are modulo n conjugates of g, it follows that, in calculating the coefficient of x^j in $\mathcal{P}(x)$, only those indices g which are conjugates of j, i.e, elements of Γ_j, need be considered. In other words, the coefficient of x^j in $\mathcal{P}(x)$ equals the coefficient of x^j in

$$\mathrm{Tr}_1^m \left(\sum_{g \in \Gamma_j} P_g x^g \right) \bmod x^n - 1 \tag{14}$$

Now we need to invoke the fact that, if d is a divisor of m, then

$$\mathrm{Tr}_1^m(\xi) = \mathrm{Tr}_1^d(\mathrm{Tr}_d^m(\xi)) \tag{15}$$

[2, Theorem 8.2]. Combining Eqs. (14) and (15), we find that the coefficient of x^j in $\mathcal{P}(x)$ is in fact the coefficient of x^j in

$$\mathrm{Tr}_1^d \left(\mathrm{Tr}_d^m \sum_{g \in \Gamma_j} P_g x^g \right) \bmod x^n - 1 \tag{16}$$

Now since each element of Γ_j has degree d, it follows that $\mathrm{Tr}_d^m(P_g x^g) = x^g \mathrm{Tr}_d^m(P_g) \bmod x^n - 1$ for each $g \in \Gamma_j$, so that Expression (16) becomes

$$\mathrm{Tr}_1^d \sum_{g \in \Gamma_j} x^g \mathrm{Tr}_d^m(P_g) \tag{17}$$

Recalling that $\mathrm{Tr}_1^d(\xi) = \xi + \xi^2 + \cdots + \xi^{2^{d-1}}$, we see that if i is the unique index in the range $0 \leq i \leq d-1$ such that $2^i g = j$, the gth term in the sum in Eq. (17) contributes exactly

$$\mathrm{Tr}_d^m(P_g)^{2^i} = \mathrm{Tr}_d^m(P_g^{2^i})$$

to the coefficient of x^j. But this is exactly what Eq. (13) says. ❐

Corollary. If j_1 and j_2 are conjugate modulo n, then $\mathcal{P}_{j_1}$ and $\mathcal{P}_{j_2}$ are conjugates in $GF(2^m)$. More precisely, if j has degree d, and if $s \in \{0, 1, \ldots, d-1\}$, then

$$\mathcal{P}_{2^s j \bmod n} = \mathcal{P}_j^{2^s}$$

Proof: If we use Lemma 2 to compute $\mathcal{P}_j^{2^s}$, we find

$$\mathcal{P}_j^{2^s} = \left(\mathrm{Tr}_d^m \sum_{\substack{g \in \Gamma_j \\ 2^i g \bmod n = j}} P_g^{2^i} \right)^{2^s}$$

$$= \mathrm{Tr}_d^m \sum_{\substack{g \in \Gamma_j \\ 2^i g \bmod n = j}} P_g^{2^{i+s}}$$

$$= \mathrm{Tr}_d^m \sum_{\substack{g \in \Gamma_j \\ 2^k g \bmod n = 2^s j}} P_g^{2^k}$$

$$= \mathcal{P}_{2^s j} \qquad \square$$

Example 4. For $m = 4$ and $n = 5$, we have from the Corollary, with $j = 1$ and $s = 0, 1, 2, 3$, that

$$\mathcal{P}_{2^0 \bmod 5} = \mathcal{P}_1 = \mathcal{P}_1$$

$$\mathcal{P}_{2^1 \bmod 5} = \mathcal{P}_2 = \mathcal{P}_1^2$$

$$\mathcal{P}_{2^2 \bmod 5} = \mathcal{P}_4 = \mathcal{P}_1^4$$

$$\mathcal{P}_{2^3 \bmod 5} = \mathcal{P}_3 = \mathcal{P}_1^8$$

These relationships can be verified directly by referring to Example 2. $\square$

Now we are prepared to begin the proof of Theorem 2. In effect, we wish to count the number of sets $(c_j)_{j \in J}$ such that Eq. (8) holds. If we substitute the formula given in Eq. (7) into Eq. (8), we find that $(c_j)_{j \in J}$ defines a word in the TSRS code $\mathcal{C}_J^\mu$ if and only if

$$\mathrm{Tr}_1^m \left(\sum_{j \in J} \alpha^h c_j x^j \right) = 0 \quad \begin{cases} h = 0, 1, \ \ldots, \ \mu - 1 \\ x \in \{1, \alpha, \ \ldots, \ \alpha^{n-1}\} \end{cases} \tag{18}$$

Now, for $h = 0, 1, \ \ldots, \ \mu - 1$, we define the polynomial $P_h(x)$ as

$$P_h(x) = \sum_{j \in J} \alpha^h c_j x^j$$

Thus Eq. (18) holds if and only if $\mathrm{Tr}_1^m(P_h(x)) = 0$ for all $x \in \{1, \alpha, \ \ldots, \ \alpha^{n-1}\}$, for all $h = 0, 1, \ \ldots, \ \mu - 1$. By Lemma 1, this will be true if and only if $\mathcal{P}_{h,j} = 0$ for all $h = 0, 1, \ \ldots, \ \mu - 1$ and all $j \in J$, where $\mathcal{P}_{h,j}$ is the coefficient of x^j in the polynomial $\mathcal{P}_h(x) = \mathrm{Tr}_1^m(P_h(x)) \bmod (x^n - 1)$ (Eq. (12)).

Now by Lemma 2, if $d = \deg(j)$, the coefficient $\mathcal{P}_{h,j}$ is given by the formula

$$\mathcal{P}_{h,j} = \mathrm{Tr}_d^m \left(\sum_{g \in J_j} c_g^{2^i} \alpha^{h 2^i} \right) \tag{19}$$

where in Eq. (19), $J_j = \Gamma_j \cap J$.

In summary, a set $(c_j)_{j \in J}$ of elements from $GF(2^m)$ corresponds to a codeword in $\mathcal{C}_J^\mu$ if and only if $\mathcal{P}_{h,j}$, as defined in Eq. (19), is zero, for all $h = 0, 1, \ \ldots, \ \mu - 1$ and all $j \in J$. However, by the Corollary to Lemma 2, conjugate j's correspond to conjugate $\mathcal{P}_{h,j}$'s, and so if $\mathcal{P}_{h,j} = 0$ for *one* element j of a given cyclotomic coset, it will be zero for all other elements of the coset as well. Thus in "solving" the equations $\mathcal{P}_{h,j} = 0$, it is sufficient to restrict j to lie in the set I_n, consisting of the least element of each cyclotomic coset. Thus if we want to count the number of sets $(c_j)_{j \in J}$ corresponding to codewords in the TSRS code $\mathcal{C}_J^\mu$, we get one set of equations of the form $\mathcal{P}_{h,j} = 0$ for $h = 0, 1, \ \ldots, \ \mu - 1$ for each modulo n cyclotomic coset, i.e., each $j \in I_n$.

To simplify the notation, for each $g \in J_j$, where $j \in I_n$, we define $x_g = c_g^{2^i}$, where according to our convention i is the unique index such that $2^i g \bmod n = j$. Note that since the mapping $\xi \rightarrow \xi^{2^i}$ is one to one, the c_g's can be uniquely recovered from the x_g's. For the remainder of the proof, we shall focus on the problem of determining when a set $(x_g)_{g \in J}$ corresponds to a codeword in $\mathcal{C}_J^\mu$. By the foregoing discussion and Eq. (19), this will be true if and only if

$$\mathrm{Tr}_d^m \left(\sum_{g \in J_j} x_g \alpha^{h 2^i} \right) = 0 \qquad \text{for } h = 0, 1, \ \ldots, \ \mu - 1 \tag{20}$$

for all $j \in I_n$. Since a set of equations of the form of Eq. (20) involves only those variables x_g corresponding to g's in a fixed cyclotomic coset, it follows that if the number of solutions to Eq. (20) is denoted by N_j, then the total number of codewords in the code $\mathcal{C}_J^\mu$ is simply $\prod_{j \in I_n} N_j$.

Let Γ be one of the cyclotomic cosets modulo n, with $|\Gamma| = d$, and let E be a subset of Γ, with $|E| = e$. Let $(x_g)_{g\in E}$ be e variables taking values in the field $GF(2^m)$, which satisfy the μ simultaneous linear equations

$$\mathrm{Tr}_d^m\left(\sum_{g\in E} x_g \alpha^{h2^g}\right) = 0 \qquad \text{for } h = 0, 1, \ldots, \mu - 1 \quad (21)$$

Theorem 3. The set of solutions $(x_g)_{g\in E}$ to Eq. (21) is a vector space over $GF(2^d)$, of dimension $(ef-\mu)_+$, where $f = m/d$. The number of solutions is therefore $2^{d(ef-\mu)_+}$.

Theorem 3, combined with the previous discussion, completes the proof of Theorem 2, since it implies that, for each $j \in I_n$, the number of solutions to Eq. (20) is $2^{d_j(e_jf_j-\mu)_+}$, where $f_j = m/d_j$, which is the same as $2^{(me_j-\mu d_j)_+}$, as asserted in Theorem 2.

Proof of Theorem 3: The fact that the set of solutions to Eq. (21) is a vector space over $GF(2^d)$ follows from the fact that Tr_d^m is a linear mapping from $GF(2^m)$ to $GF(2^d)$, i.e, that if x and x' are elements of $GF(2^m)$, and if λ and λ' are elements of $GF(2^d)$, then

$$\mathrm{Tr}_d^m(\lambda x + \lambda' x) = \lambda\,\mathrm{Tr}_d^m(x) + \lambda'\,\mathrm{Tr}_d^m(x) \qquad (22)$$

Using Eq. (22), it is easy to see that if (x_g) is one solution to Eq. (21), and if (x'_g) is another, and if λ and λ' are elements of $GF(2^d)$, then $(\lambda x_g + \lambda' x'_g)$ is also a solution to Eq. (21). In the remainder of the proof, we will show that the $GF(2^d)$ dimension of the solution space to Eq. (21) is $(ef-\mu)_+$.

To simplify notation, let $q = 2^d$, so that $2^m = q^f$, and let $\tau = \mathrm{Tr}_d^m$. If now $(x_g)_{g\in E}$ is an arbitrary vector from $GF(2^m)^e \cong GF(q^f)^e$, define, for $h = 0, 1, \ldots, \mu - 1$,

$$y_h = \sum_{g\in E} x_g \alpha^{h2^g} \qquad (23)$$

and

$$z_h = \tau(y_h) \qquad (24)$$

Now let T be the $GF(q)$-linear mapping from $GF(q^f)^e = GF(q)^{ef}$ to $GF(q)^\mu$, defined by

$$T : (x_g)_{g\in E} \to (z_h)_{h=0,\ \ldots,\ \mu-1}$$

We will show that the rank of the mapping T is $\min(ef, \mu)$, so that the nullity of T is $(ef-\mu)_+$, and this will complete the proof. We will consider the cases $ef \geq \mu$ and $ef < \mu$ separately.

Case 1: $ef \geq \mu$. We need to show that $\mathrm{rank}(T) = \mu$. This is equivalent to showing that there is no set of nonzero (λ_h)'s such that

$$\sum_{h=0}^{\mu-1} \lambda_h z_h = 0 \qquad (25)$$

for all vectors (x_g) in $GF(q^f)^e$. If Eq. (25) *is* true, then from Eq. (24), we have

$$\tau\left(\sum_{h=0}^{\mu-1} \lambda_h y_h\right) = 0 \qquad (26)$$

for all (x_g)'s. The inner sum in Eq. (26) is, by Eq. (23),

$$\sum_{h=0}^{\mu-1} \lambda_h y_h = \sum_{h=0}^{\mu-1} \lambda_h \sum_{g\in E} x_g \alpha^{h2^g}$$

$$= \sum_{g\in E} x_g \sum_{h=0}^{\mu-1} \lambda_h \alpha^{h2^g}$$

It follows then that

$$\tau\left(\sum_{g\in E} x_g \beta_g\right) = 0$$

for all (x_g), where $\beta_g = \sum_{h=0}^{\mu-1} \lambda_h \alpha^{h2^g}$. But it is easy to see that this can hold if and only if $\beta_g = 0$ for all $g \in E$. In summary, then, Eq. (25) will be true for all (x_g)'s if and only if

$$\sum_{h=0}^{\mu-1} \lambda_h \alpha^{h2^g} = 0 \qquad \text{for all } g \in E \qquad (27)$$

Next, define the polynomial $L(x)$ as

$$L(x) = \sum_{h=0}^{\mu-1} \lambda_h x^h$$

Then Eq. (27) says $L(\alpha^{2^i}) = 0$ for all $g \in E$, so that $L(x)$ has $|E| = e$ distinct roots in the set $\{\alpha, \alpha^2, \ldots, \alpha^{2^{d-1}}\}$. However, since the coefficients of $L(x)$ are in $GF(q)$, then every $GF(q)$ conjugate of a root of $L(x)$ is also a root. But conjugation with respect to $GF(q)$ is the mapping $\beta \rightarrow \beta^q = \beta^{2^d}$. But since a is a primitive root, each element of $GF(2^m) = GF(2^{df})$ of the form α^{2^i} with $0 \leq i \leq d-1$ has exactly f $GF(2^d)$ conjugates, viz, α^{2^k} for $k = i, i+d, \ldots, i+d(f-1)$. Hence by taking conjugates of the original e roots of $L(x)$, we obtain ef roots. But since $ef \geq \mu$ by assumption, and since $\deg L(x) \leq \mu - 1$, it follows that the coefficients of $L(x)$, i.e, the λ_h's, are all zero. This means that no nontrivial relationship of the form of Eq. (25) can hold, which completes the proof in Case 1.

Case 2: $ef < \mu$. In this case we need to show that $\text{rank}(T) = ef$. But clearly $\text{rank}(T) \leq ef$, since the $GF(q)$-dimension of the space of all (x_g)'s is ef. However, by the argument in Case 1, the first ef components of (z_g) are linearly independent, and so $\text{rank}(T) \geq ef$ as well. Hence $\text{rank}(T) = ef$, as asserted. ❐

IV. Summary and Conclusions

In this article, we have introduced an extensive class of symbol-oriented error-correcting codes, which have properties much like those of Reed–Solomon codes, without, however, suffering from the major drawback of RS codes, viz, an intrinsic limitation on codeword length. As subcodes of RS codes, these codes can be decoded by any RS decoding algorithm. However, the study of these codes is in its infancy, and we therefore close with a list of unsolved problems related to TSRS codes.

(1) Our selection of the representation of an element from $GF(2^m)$ as a binary m-tuple was more or less arbitrary. If another representation is used, will a code of larger dimension result?

(2) Devise an efficient encoding algorithm for an arbitrary TSRS code, or at least a large class of them.

(3) Determine the conditions under which TSRS codes are systematic over the symbol alphabet $V(2^{m-\mu})$. (A necessary condition for this to be so is that the pseudodimension k_μ be an integer. But in [4], it was shown by example that this condition is not sufficient.)

(4) Study the combinatorial optimality of TSRS codes. For example, devise bounds on the cardinality of a (n, k, d) code over a q-letter alphabet when q is a fixed fraction of n, say λn, as $n \rightarrow \infty$.

(5) Investigate the relationship between TSRS codes and generalized BCH codes, i.e., BCH codes whose symbol field is $GF(2^d)$ and whose locator field is $GF(2^m)$, where d is a divisor of m.

(6) Compare the distance properties of TSRS codes to algebraic geometry codes with approximately the same values of n, k, and q.

(7) Do TSRS codes meet (or exceed) the Gilbert-Varshamov bound?

References

[1] M. Hall, *The Theory of Groups*, New York: Macmillan, 1959.

[2] R. J. McEliece, *Finite Fields for Computer Scientists and Engineers*, Boston: Kluwer, 1986.

[3] I. S. Reed and G. Solomon, "Polynomial Codes over Certain Finite Fields," *J. Soc. Indus. Appl. Math.*, vol. 8, pp. 300–304, 1960.

[4] G. Solomon, "Nonlinear, Nonbinary Cyclic Group Codes," *The Telecommunications and Data Acquisition Progress Report 42-108, vol. October–December 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 84–95, February 15, 1992.

[5] G. Solomon, "Nonlinear, Nonbinary Cyclic Group Codes," *Proceedings of the 1993 International Symposium on Information Theory*, San Antonio, Texas, p. 192, January 17–22, 1993.

[6] J. H. Yuen, ed., *Deep Space Telecommunications Engineering*, JPL Publication 82–76, Jet Propulsion Laboratory, Pasadena, California, July 1982.

N94-35555

TDA Progress Report 42-117 May 15, 1994

Maximum-Likelihood Soft-Decision Decoding of Block Codes Using the A* Algorithm

L. Ekroot and S. Dolinar
Communications Systems Research Section

The A algorithm finds the path in a finite depth binary tree that optimizes a function. Here, it is applied to maximum-likelihood soft-decision decoding of block codes where the function optimized over the codewords is the likelihood function of the received sequence given each codeword. The algorithm considers codewords one bit at a time, making use of the most reliable received symbols first and pursuing only the partially expanded codewords that might be maximally likely. A version of the A* algorithm for maximum-likelihood decoding of block codes has been implemented for block codes up to 64 bits in length. The efficiency of this algorithm makes simulations of codes up to length 64 feasible. This article details the implementation currently in use, compares the decoding complexity with that of exhaustive search and Viterbi decoding algorithms, and presents performance curves obtained with this implementation of the A* algorithm for several codes.*

I. Introduction

The A* algorithm is an artificial intelligence algorithm for finding the path in a graph that optimizes a function. Nilsson [1, pp. 72–88] describes the algorithm as a heuristic graph-search procedure and shows that the algorithm terminates in an optimal path. The A* algorithm has been used to implement full maximum-likelihood soft decoding of linear block codes by Han et al. [2–4].

The A* algorithm explores the codewords one bit at a time using the most reliable information first, pursuing the most likely codewords first, and ruling out suboptimal codewords as soon as possible. The details of the A* algorithm are covered in Section II. The implementation discussed here has most of the features recommended in [2] and has allowed code performance simulations for codes up to length 64 in reasonable amounts of time.

The codes considered here are binary linear codes. An (N, K) code has 2^K codewords each of length N bits. The binary symbols are transmitted over a communications channel and corrupted by channel noise. An additive noise channel is one in which each received symbol can be described as the sum of the signal and noise. The deep space channel is accurately described as an additive white Gaussian noise channel, meaning that the additive noise for each symbol is an independent random variable distributed according to a zero-mean Gaussian with variance

σ^2. Therefore, the received symbols, r_i for $i = 1, \ldots, N$, that form the received sequence $\boldsymbol{r}$ are continuously valued and are called soft symbols.

Binary antipodal signaling transmits the bits as signals of equal magnitude, but different signs. For this article, the ith bit, $b_i \in \{0, 1\}$, of the binary codeword $\boldsymbol{b}$ is transmitted as $c_i = (-1)^{b_i}$. That is, a 0 is transmitted as $+1$, and a 1 is transmitted as -1. Thus, the energy in a transmitted codeword $\boldsymbol{c}$ is given by $\sum_{i=1}^{N} c_i^2 = \sum_{i=1}^{N}((-1)^{b_i})^2 = N$ independently of the codeword sent. The signal-to-noise ratio (SNR) is E_b/N_0, where $E_b = N/K$ is the energy per information bit, and $N_0/2 = \sigma^2$ is the white Gaussian noise two-sided spectral density. Thus, the SNR is given by

$$SNR = 10 \log_{10} \left(\frac{N}{2\sigma^2 K} \right) \text{(dB)}$$

The hard-limited symbol h_i is the transmitted signal value ± 1 nearest to the received symbol r_i; it is given by

$$h_i = \text{sgn}\ (r_i) = \frac{r_i}{|r_i|}$$

where h_i equals 1 in the zero-probability case of $r_i = 0$. If the received symbols are individually hard-limited before decoding, some information is lost from each symbol. For instance, soft symbols like 0.01 and 1.01 both hard limit to $+1$, yet the received value 1.01 is much more likely to have been transmitted as $+1$ than is the received value 0.01. The decoder performance is improved by approximately 2 dB [5, pp. 404–407] if, instead of the hard-limited symbols, the soft symbols r_i are used for what is then called soft-decision decoding.

A. Maximum-Likelihood Decoding

One soft-decision decoding technique is called maximum-likelihood soft-decision decoding. It decodes a received sequence to the codeword $\boldsymbol{c}^*$ that maximizes the likelihood of the received soft symbols.

It is convenient to think of the codewords, which are length N sequences of ± 1's, and the received sequence $\boldsymbol{r}$ as points in N-dimensional space. Assuming an additive white Gaussian noise channel, the codeword $\boldsymbol{c}^*$ that maximizes the likelihood of the received sequence $\boldsymbol{r}$ is the one that minimizes the Euclidean distance between the received word $\boldsymbol{r}$ and the codeword $\boldsymbol{c}$.

The codeword that is closest to the received word can be found by exhaustively checking all possible codewords, or by cleverly seeking out the one that minimizes the distance. The first technique is called exhaustive search maximum-likelihood decoding. For an (N, K) code, there are 2^K codewords to check, making an exhaustive search prohibitive for most interesting codes. Viterbi decoding the block code on a trellis can give better decoding performance with a fixed number of calculations. Techniques such as the A* algorithm that use a heuristic search to find the minimizing codeword can significantly reduce the calculations needed for decoding, especially at a high SNR. Nilsson [1] shows that the A* algorithm terminates with the optimal path; Han [2] shows how to apply the technique to maximum-likelihood decoding; and this article describes the algorithm as we have implemented it.

B. Linear Codes as Trees

Define $\mathcal{C}$ to be an (N, K) linear code with 2^K length N binary codewords $\boldsymbol{b} \in \mathcal{C}$. The generator matrix G for the code is a $K \times N$ matrix of 0's and 1's that has as rows linearly independent codewords. Given K information bits in a row vector $\boldsymbol{x}$, the corresponding binary codeword $\boldsymbol{b}$ is defined to be $\boldsymbol{b} = \boldsymbol{x}G$. For a systematic code, the K information bits are directly visible in the codewords. For the systematic codes referred to here, the first K columns of G form an identity matrix, and the codewords can be divided into information bits, in the first K positions, and parity bits, in the last $N - K$ positions.

1. Example. Consider the shortened (6,3) Hamming code with the generator matrix

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

The information sequences and the corresponding binary codewords and transmitted codewords are listed in Table 1.

To apply a heuristic tree search algorithm to the decoding of a block code, the code is thought of as a binary tree with 2^K depth N leaves where each path from the root to a leaf corresponds to a codeword. Figure 1 shows the representation of the (6,3) shortened Hamming code as a tree with solid and dashed edges used to represent 0's and 1's, respectively, or equivalently, transmitted $+1$'s and -1's. A node of level l in the tree is defined by the path $\boldsymbol{p}^l = p_1 p_2, \ldots, p_l$ from the root to that node with binary components p_i. For example, the path $\boldsymbol{p}^2 = 10$ specifies the level 2 node designated by "o" in Fig. 1.

If the codewords of a systematic code live in a binary tree, the tree is full through level $K-1$, that is, every node branches. This is because any length K sequence of 0's and 1's can be an information sequence. Since the parity bits are determined by the information bits, every level K node has only one descendant path that continues to level N.

II. Algorithm Description

The A* algorithm for soft decoding block codes searches a binary tree for the length N path that minimizes a function. On any given iteration, it uses a heuristic to expand the node that is likely to yield the optimal path and eliminate any nodes that can only have suboptimal descendants. The method by which nodes are selected for expansion or eliminated from consideration uses an underestimate of the function to be minimized, called a heuristic function. The heuristic function at a node must lower bound the function to be minimized for all paths that pass through that node.

For maximum-likelihood soft-decision decoding, the function that is minimized over all codewords is the Euclidean distance between a codeword and the received word. It is equivalent to minimizing the square of the Euclidean distance:

$$s(\boldsymbol{r}, \boldsymbol{c}) = \sum_{i=1}^{N} (r_i - c_i)^2$$

For the algorithm to find the minimizing codeword, the heuristic function at a node must be less than or equal to the actual squared distance of any full-length path that passes through that node (or equivalently any codeword that is prefixed by $\boldsymbol{p}^l$, the path that defines the node).

The minimum distance over all codewords that begin with the path $\boldsymbol{p}^l$ is lower bounded by the minimum distance over all length N binary sequences that begin with the path $\boldsymbol{p}^l$. This can be made explicit for distance squared by

$$\min_{\{\boldsymbol{b} \in \{0,1\}^N | b_i = p_i,\ i=1,2,\ldots,l\}} s(\boldsymbol{r}, \boldsymbol{c}) \leq$$

$$\min_{\{\boldsymbol{b} \in \mathcal{C} | b_i = p_i,\ i=1,2,\ldots,l\}} s(\boldsymbol{r}, \boldsymbol{c})$$

where $c_i = (-1)^{b_i}$. The minimum distance over all length N sequences that begin with the path $\boldsymbol{p}^l$ is achieved by the sequence that begins with $\boldsymbol{p}^l$ and continues with binary symbols consistent with the hard-limited received symbols. The heuristic function is the squared distance from the received sequence to either the codeword defined by the path $\boldsymbol{p}^K$, if the node is at level $l = K$, or the sequence that begins with the path $\boldsymbol{p}^l$ and is completed by symbols consistent with the hard-limited symbols, if $l < K$.

A. Fundamentals of the Algorithm

The A* algorithm keeps an ordered list of possible nodes to operate on. Associated with each node is the path $\boldsymbol{p}^l$ that identifies the node, the value of the heuristic function, and an indicator of whether the node represents a single codeword. The value of the heuristic function determines the order of the list of nodes and, therefore, guides the search through the tree.

When the A* algorithm begins the search, the root of the tree is the only node on the list. The algorithm expands a node if it might yield a codeword with the minimum distance from the soft received symbols, and eliminates from the list nodes that are too far from the received symbols to have the maximum-likelihood codeword as a descendant. The algorithm terminates when the node at the top of the list represents a single codeword. That codeword is the maximum-likelihood codeword.

At each iteration, the node on the top of the list, which has the smallest value of the heuristic function, is expanded. It is taken off the list, and the two possible ways of continuing the path are considered as nodes to put back on the list. Each new node is placed back on the list provided that its heuristic function is not greater than the actual squared distance for a completed codeword. If the node expanded is at level $K-1$, the two level K children specify codewords, and the heuristic function for each child node is the actual squared distance between the codeword and the received word. These codewords are called candidate codewords. When a node that defines a codeword is placed back on the list, all nodes below it on the list are deleted.

B. Features That Improve the Efficiency

The features described in this section are not necessary to guarantee maximum-likelihood soft-decision decoding, but they improve the efficiency. Sorting the bit positions to take advantage of reliable symbols reduces the number of nodes expanded. Using a simplification of the heuristic function reduces the number of computations during each node expansion. These features have been implemented and make the A* decoder a practical tool for decoding.

1. Sorting by Reliability. If the bit positions corresponding to the more reliable received symbols are expanded first, then the search will be directed more quickly to close candidate codewords. The nearer a symbol is to 0, the less reliable it is because it is almost equally far from both +1 and −1. Similarly, the greater the magnitude of the received symbol, the more reliable that symbol is. To take advantage of the reliable symbols, the received symbols are reordered in descending order of magnitude, and the code symbols are reordered equivalently. Since the A* algorithm defines the code by the $K \times N$ generator matrix G, reordering the code symbols is equivalent to reordering the columns of the generator matrix.

This implementation of the A* algorithm sorts the received symbols by reliability, reorders the columns of the generator matrix in the same way, and then tries to row reduce the generator matrix so that it is systematic. However, if it encounters a column, among the first K columns, that is linearly dependent on previous columns, it moves the offending column and corresponding received symbol to the end, and continues the row reduction.

Typically, the number of nodes expanded while decoding a received word is significantly reduced by sorting the symbols before starting the decoding process. The increase in efficiency from sorting first was found, for the shorter codes like the Golay (24,12) code, to outweigh the cost of sorting and row reducing the generator matrix. For the larger codes, like the quadratic residue (QR) (48,24) code, decoding without sorting was so much more time-consuming that it was not a reasonable option to run comparison tests. Sorting was adopted as a standard feature.

2. Alternative Function to Minimize. Every soft symbol r_i is at least as far away from the codeword symbol c_i as it is from the hard-limited symbol h_i. The squared distance, $s(\boldsymbol{r},\boldsymbol{c})$, can be written as the sum of the square of the distance to the hard-limited symbols, $s(\boldsymbol{r},\boldsymbol{h})$, and an amount $a(\boldsymbol{r},\boldsymbol{c})$ that is nonzero only when there is at least one symbol c_i that does not equal the hard-limited symbol h_i, as follows:

$$s(\boldsymbol{r},\boldsymbol{c}) = \sum_{i=1}^{N}(r_i - c_i)^2$$

$$= \sum_{i=1}^{N}(r_i - h_i + h_i - c_i)^2$$

$$= \sum_{\substack{i=1 \\ h_i=c_i}}^{N}(r_i - h_i)^2 + \sum_{\substack{i=1 \\ h_i\neq c_i}}^{N}(r_i - h_i + 2h_i)^2$$

$$= \sum_{\substack{i=1 \\ h_i=c_i}}^{N}(r_i - h_i)^2$$

$$+ \sum_{\substack{i=1 \\ h_i\neq c_i}}^{N}\left[(r_i - h_i)^2 + 4h_i(r_i - h_i) + 4h_i^2\right]$$

$$= \sum_{i=1}^{N}(r_i - h_i)^2 + 4\sum_{\substack{i=1 \\ h_i\neq c_i}}^{N} h_i r_i \tag{1a}$$

$$= \sum_{i=1}^{N}(r_i - h_i)^2 + 4\sum_{\substack{i=1 \\ h_i\neq c_i}}^{N} |r_i| \tag{1b}$$

$$= s(\boldsymbol{r},\boldsymbol{h}) + 4a(\boldsymbol{r},\boldsymbol{c}) \tag{1c}$$

where Eq. (1a) follows by recombining the summations of like terms, Eq. (1b) uses $h_i r_i = \operatorname{sgn}(r_i) r_i = |r_i|$, and Eq. (1c) introduces an alternative function,

$$a(\boldsymbol{r},\boldsymbol{c}) = \sum_{\substack{i=1 \\ \operatorname{sgn}(r_i)\neq c_i}}^{N} |r_i|$$

Since the first term in Eq. (1c) does not depend on the codeword $\boldsymbol{c}$, it is constant over the minimization, and

$$\min_{\boldsymbol{c}} s(\boldsymbol{r},\boldsymbol{c}) = s(\boldsymbol{r},\boldsymbol{h}) + 4\min_{\boldsymbol{c}} a(\boldsymbol{r},\boldsymbol{c})$$

Maximum-likelihood decoding of the received sequence can be done by finding the codeword that minimizes either $s(\boldsymbol{r},\boldsymbol{c})$ or $a(\boldsymbol{r},\boldsymbol{c})$.

Because each term of $a(\boldsymbol{r},\boldsymbol{c})$ is either zero or $|r_i|$, based on a comparison, it is simpler to calculate than $s(\boldsymbol{r},\boldsymbol{c})$, which for each i requires a difference and a square. The alternative function $a(\boldsymbol{r},\boldsymbol{c})$ is used by this implementation of the A* algorithm.

3. Example Revisited. Consider the shortened (6,3) Hamming code and the received sequence $\boldsymbol{r} = (0.05, -1.3,$

1.1, 0.8, −0.25, 0.6). Reordering the received vector by magnitude gives $\boldsymbol{r}' = (-1.3, 1.1, 0.8, 0.6, -0.25, 0.05)$. The reordered generator matrix in systematic form is

$$G' = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

and a binary codeword $\boldsymbol{b}$ in the original code corresponds to a sorted codeword $\boldsymbol{b}'$. Figure 2 shows the tree explored by the A* algorithm when the sorted code is used. Each node is labeled with the value at that node of the heuristic function using the alternative function. The 3 expanded nodes are each designated by a "o"; the 2 candidate codewords are each designated by a "Δ"; and the 12 edges explored are the edges of the tree that are shown. The nodes with paths 0, 11, and 101 are dropped from the list when the candidate word, with alternative function 0.65, is put on the list. The search promptly terminates because the top node on the list defines a candidate codeword, namely $\boldsymbol{b}' = 100111$. Unshuffling $\boldsymbol{b}'$ gives the maximum-likelihood decoded codeword to be $\boldsymbol{b} = 110011$. For comparison, Fig. 3 shows the larger tree explored by the algorithm when the symbols are not sorted.

C. Verification of the Decoder

The decoding results of the A* algorithm were compared to the results of two exhaustive search decoder implementations. The (24,12) Golay code was used for this test since it has only $2^{12} = 4096$ codewords, making it feasible to get timely results from an exhaustive search decoder. First, the software decoded the received sequences using both A* and exhaustive search, and compared the results internally. Second, a couple hundred received sequences were decoded by both the A* software and an independent exhaustive search decoder written in APL. The results showed that both exhaustive search and A* decoders decoded the same noisy vectors to the same codewords.

The software to implement the A* algorithm has been written in C and run on several Sun platforms. Since integers on these processors are 32 bits long, the software to implement the A* algorithm has been constrained to linear codes with 64 or fewer bits per codeword by using two 32-bit integers for each codeword. Because of this implementation detail, it was important to confirm that the A* software properly decodes codes longer than length 32. Most interesting codes with lengths over 32 bits take a prohibitively long time to decode exhaustively. A test code with length N greater than 32 and one with more than 32 information bits were devised so they could be readily decoded by other means. The code with length greater than 32 was created by repeating the parity bits of the Golay (24,12) code. This formed a (36,12) code that was no more difficult to exhaustively decode than was the Golay (24,12) code. After debugging and testing, the decoder decoded 500 codewords consistent with the exhaustive decoder results. Next, a simple (34,33) code, consisting of 33 information bits and 1 overall parity bit, was tested on 200 noisy received words. This code was selected because a maximum-likelihood decoder is easy to write, and an APL program was used to verify that the 200 test words decoded consistently.

D. Operational Details

To analyze the performance of either the algorithm or a code, data are taken by running the software with different input parameters. For a given run, the software takes as input the generator matrix for the code, the SNR, the seed for the random number generator, and the number of words to decode. It returns the average number of nodes expanded, the average number of candidate codewords, and the number of word errors that occurred. Sometimes a system call from inside the program was used to provide the amount of central processing unit (CPU) time consumed during a run. These decoding runs ranged in size from hundreds to tens of thousands of decoded received sequences. The codes that have been examined include a Bose–Chaudhuri–Hocquenghem (BCH) (63,56) code, a quadratic residue (48,24) code, a Golay (24,12) code, a BCH (31,10) code, and a Reed–Muller (32,16) code. The data from multiple runs are combined carefully to give the results in the following sections.

III. Algorithm Performance

The intricacy of the A* algorithm makes it difficult to describe the number of calculations necessary to decode a received sequence. Possible measures of the number of calculations describe the search size for the A* algorithm and include the number of candidate codewords, the number of expanded nodes, and the number of edges searched in the tree. The number of edges E in the search tree is given by

$$E = 2X + (N - K)C$$

where X is the number of nodes expanded including the root and C is the number of candidate codewords. Because the search size for the A* algorithm varies from one

received sequence to the next, the averages of these numbers over many received sequences are used for comparison. Section III.A explains simulation timing results that show that the search size and the time to decode are related linearly. Section III.B shows how the average size of the A* search tree depends on the signal-to-noise ratio. Section III.C introduces other maximum-likelihood decoders that are used for comparisons in Section III.D.

A. Time to Decode Versus Search Size

The average amount of time it takes to decode received sequences reflects both the computational overhead for each sequence decoded and the computations for each part of the search tree. Analyzing the time to decode requires that all of the timing data be taken on the same computer and that the accuracy of the timing data be sufficient to perform comparisons. The system call used to generate the timing information for a run was accurate to within a second, which is too coarse to study data on individual decoded sequences, but sufficient for data on ensembles of decoded sequences.

The relationship between the indicators of search size introduced earlier and decoding time may be observed in the data from many runs for the quadratic residue (48,24) code on a Sparc 10 Model 30 workstation. The average decoding times versus the average numbers of candidate codewords and expanded nodes are shown in Fig. 4, along with a weighted linear fit[1] to the data. Although the data display a small amount of statistical variability, the time to decode displays a nearly linear relationship to the indicators of search size.

The relationship between the average numbers of expanded nodes and candidate codewords for the (48,24) code in Fig. 5 is well described as linear.

B. Search Size Versus SNR

The average size of the tree that the A* algorithm searches is a function of the SNR for the received sequences. For each of the codes studied, the average numbers of candidate codewords, expanded nodes, and search tree edges are shown versus SNR in Figs. 6, 7, and 8, respectively. The total number of sequences decoded for each point in these figures is shown in Table 2.

Not surprisingly, for extremely high SNR, the A* algorithm typically finds only two candidate words, along the way expands K nodes, and therefore has a search tree with $E = 2K + 2(N - K) = 2N$ edges.

For low SNR, the soft symbols are predominantly noise, but the A* algorithm still expands a mere fraction of the nodes in the tree, especially as it bases early decisions on the symbols that contribute most to the final choice. Figures 6 and 7 show that simulation data of the average numbers of candidate words and expanded nodes are almost constant for SNRs below -4 dB. Since the number of edges is calculated from the number of nodes and candidate words, the average number of edges also levels off for low SNRs, as seen in Fig. 8.

The algorithm was also tested for each code with no signal at all, i.e., an SNR of $-\infty$. Table 3 shows that for each code the average number of expanded nodes and the average number of candidate words during A* decoding were comparable to the numbers for SNRs below -4 dB. It also shows the number of words decoded to obtain these averages.

C. Other Maximum-Likelihood Decoders

Many other maximum-likelihood soft-decision decoding algorithms use a fixed number of calculations to decode any received sequence, independent of the SNR.

1. Exhaustive Search and Full Tree Search. An exhaustive search decoder calculates the distance between the received sequence and each codeword in the code individually, and returns the codeword with the minimum distance. For an (N, K) block code, an exhaustive search technique computes the distances from the received sequence to all 2^K codewords. If exhaustive search is cast in terms of a graph with one edge for each bit in each codeword, the number of edges for an exhaustive search is $N2^K$, independent of the SNR.

A slightly more efficient technique to compute the distances to all the codewords is to use the full code tree. Here the squared distance from the received sequence to a codeword, at a leaf, is the sum along the path to that leaf of the squared distances from each received symbol to the symbol associated with each edge. For an (N, K) code, the number of edges in the full tree is $2^{K+1} - 2 + (N - K)2^K$. This technique checks all 2^K leaves but has fewer edges than an exhaustive search.

2. Viterbi Decoding of Block Codes. To apply soft-decision Viterbi decoding to a block code, the code is represented as a trellis. Wolf [6] and Massey [7] introduce a minimal code trellis for block codes. McEliece [8] shows

[1] The number of decoded words in each run was included in the line fitting process to account for the variation in accuracy between data from large and small runs.

a simple technique for constructing the minimal trellis for a given code, and also shows that it is optimal for Viterbi decoding complexity. A Viterbi decoder for a code on a trellis uses a constant number of calculations and comparisons independent of signal-to-noise ratio. The Viterbi decoding complexity for a given trellis for a given code can be measured by the number of edges in the trellis.

An (N, K) code has a minimal trellis that can be constructed from the generator matrix. Different permutations of a code may have different minimal trellises. There are codes, such as cyclic codes, for which the minimal trellis has the most edges compared with the minimal trellises for other permutations of the code. The permutation that gives the most edges in the minimal trellis is the worst permutation of the code. The number of edges in the minimal trellis for the worst permutations is no more than $2^{M+2} - 4 + 2^M(N - 2M)$ where $M = \min(K, N - K + 1)$. Other permutations give smaller minimal trellises.

D. Comparisons

The search size for the A* algorithm depends on the received sequence, and the average search size depends on the SNR. The averages found with no signal present are used for comparison with the maximum-likelihood decoders that use a fixed number of calculations.

Table 4 shows the number of edges used for an exhaustive search for each code and for the full code tree, both of which are greater than the average number of edges in the A* search tree shown for the case of no signal in Table 3. The average number of edges in the search tree for the A* algorithm is presented for comparison to the number of edges in the trellis for Viterbi decoders. Table 4 also shows the number of edges in some special trellises for the codes where the numbers are known or bounds have been calculated.

Consider the Golay (24,12) code. An exhaustive search explores $(24)2^{12} = 98,304$ edges. The full tree has $2^{13} - 2 + (12)2^{12} = 57,342$ edges. The number of edges in the minimal trellis for the worst-case permutation of the Golay $(24, 12)$ code is $2^{14} - 4 = 16,380$. The number of edges in the minimal trellis for the best permutation of the Golay code is 3580 [8]. By using enhancements on a certain trellis for the Golay code, Forney [9, p. 1183] reduces the number of binary operations of a decoder to 1351. This decoder is mentioned as interesting, but the number of binary operations is not directly comparable to edge counts for the A* algorithm. Note that at low SNRs the A* algorithm on average expands 62 nodes, checks 29 candidate codewords, and has a search tree with 469 edges.

Less is known about the best trellises for the quadratic residue (48,24) code. A worst-case permutation produces a minimal trellis with $2^{26} - 4 = 67,108,860$ edges. A better permutation results in a minimal trellis with 925,692 edges. The A* algorithm search tree has on average 34,429 edges when no signal is present.

IV. Future Enhancements

There are several elements affecting the efficiency of the software, including the initial computations to set up the search, the size of the search, and the number of computations for each part of the search. With greater understanding of the algorithm come more ideas for improving the software to reduce at least one of these elements. Either using the minimum distance of the code to determine if the search can be successfully terminated before the list is exhausted or improving the heuristic function will reduce the search size. Ideas like reducing the complexity of sorting the received symbols for each decoding will trim the number of initial setup computations, but will increase the search size by an undetermined amount.

A. Escaping When a Candidate Is Definitely Closest

If the angle between the received word and a candidate codeword, as viewed from the origin, is less than half the minimum angle between any two codewords, then that candidate codeword is the closest one to the received word. One of the suggestions in [2] is to calculate the angle between the received word and each candidate codeword as it is found. If this angle indicates that the candidate is the closest codeword, then declare that the decoding is complete, and exit the algorithm.

This feature has not been implemented at this writing, but is expected to reduce the size of the tree searched.

B. Improving the Heuristic Function

Each parity bit is a linear function of a subset of the information bits. If a node in the tree is deep enough to specify all the information bits for a particular parity bit, then any codeword passing through that node will have the same value for that parity bit. The heuristic function could use this parity bit to improve the distance underestimate for all nodes at that depth, and thus reduce the size of the tree searched by the algorithm.

C. Sorting Fewer Symbols

To simplify the sorting of symbols by reliability and the necessary row reductions to the generator matrix, it may

be beneficial to sort only the information symbols or to sort only a few of the most reliable symbols.

Sorting only the information symbols in the received word greatly simplifies the production of a systematic generator matrix for the reordered code bits. Since the columns corresponding to information bits are columns of an identity matrix, they are linearly independent, and exchanging rows is all it takes to return the generator matrix to systematic form. Thus, the row reduction portion of the algorithm is simplified.

Another possibility for simplifying sorting is to reorder only the n most reliable linearly independent symbols and not to reorder the rest. In such a design, there are no more than $N!/(N-n)!$ reorderings of the columns, and, hence, this many systematic generator matrices. For small values of n such as 1 or 2, it may be acceptable to store and retrieve systematic generator matrices for each of these reorderings.

V. Code Performance

The A* algorithm that we have implemented has been very useful for simulating code performance. Figure 9 shows the probability of word error versus SNR for the BCH (63,56), Reed-Muller (32,16), BCH (31,10), Golay (24,12), and quadratic residue (48,24) codes. The error bars are one standard deviation of an average of m independent Bernoulli trials. Specifically, the estimated standard deviation is $\sigma = \sqrt{p(1-p)/m}$, where p is the estimate of the probability of word error at a given SNR, and m is the number of decodings done at that SNR.

VI. Conclusions

The application of the A* algorithm to maximum-likelihood soft-decision decoding allows for efficient simulation of code performance. The A* algorithm finds the codeword that maximizes the likelihood of the received word given the codeword. This is equivalent to minimizing either the Euclidean distance between the received word and a codeword or the alternative function described in Section II.B.2. The use of a heuristic function constrains the search to only a subtree of the code's finite binary tree. The heuristic function underestimates the function being minimized in order to ensure that the subtree contains the optimal path.

The size of the tree searched by the A* algorithm, as described by the numbers of nodes expanded, candidate words, and edges, is a good indicator of the complexity for decoding that received sequence. Since the search size depends on the received sequence, the average search size as a function of signal-to-noise ratio is used for comparison. The search tree is smallest for a high SNR, where the algorithm goes straight to the maximum-likelihood codeword, and larger at a low SNR, where the searched portion of the tree is still much smaller than the full code tree. At a low SNR, the average size of the A* search tree is also smaller than the fixed trellis size of a good Viterbi decoder.

For research applications, simulations using this implementation can provide data on code performance, such as word error rate, for comparisons to theoretical results, such as bounds, and for testing other predictors of code performance.

Acknowledgments

The authors would like to thank A. Kiely, R. McEliece, and F. Pollara for many helpful discussions.

References

[1] N. J. Nilsson, *Principles of Artificial Intelligence*, Palo Alto, California: Tioga Publishing Co., 1980.

[2] Y. S. Han, C. R. P. Hartmann, and C.-C. Chen, *Efficient Maximum-Likelihood Soft-Decision Decoding of Linear Block Codes Using Algorithm A**, Technical Report SU-CIS-91-42, School of Computer and Information Science, Syracuse University, Syracuse, New York, December 1991.

[3] Y. S. Han and C. R. P. Hartmann, *Designing Efficient Maximum-Likelihood Soft-Decision Decoding Algorithms for Linear Block Codes Using Algorithm A**, Technical Report SU-CIS-92-10, School of Computer and Information Science, Syracuse University, Syracuse, New York, June 1992.

[4] Y. S. Han, C. R. P. Hartmann, and C.-C. Chen, "Efficient Priority-First Search Maximum-Likelihood Soft-Decision Decoding of Linear Block Codes," *IEEE Transactions on Information Theory*, vol. 39, no. 5, pp. 1514–1523, September 1993.

[5] S. Benedetto, E. Biglieri, and V. Castellani, *Digital Transmission Theory*, Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1987.

[6] J. K. Wolf, "Efficient Maximum Likelihood Decoding of Linear Block Codes Using a Trellis," *IEEE Transactions on Information Theory*, vol. IT-24, no. 1, pp. 76–80, January 1978.

[7] J. L. Massey, "Foundations and Methods of Channel Coding," *Proceedings of the International Conference on Information Theory and Systems*, NTG-Fachberichte vol. 65, Berlin, pp. 148–157, September 18–20,1978.

[8] R. J. McEliece, "The Viterbi Decoding Complexity of Linear Block Codes," to be presented at IEEE ISIT'94, Trondheim, Norway, June 1994.

[9] G. D. Forney, Jr., "Coset Codes — Part II: Binary Lattices and Related Codes," *IEEE Transactions on Information Theory*, vol. IT-34, no. 5, pp. 1152–1187, September 1988.

Table 1. Information sequences and the corresponding binary and transmitted codewords for the (6,3) shortened Hamming code.

Information sequence x	Binary codeword b	Transmitted codeword c
0 0 0	0 0 0 0 0 0	1 1 1 1 1 1
0 0 1	0 0 1 0 1 1	1 1 −1 1 −1 −1
0 1 0	0 1 0 1 0 1	1 −1 1 −1 1 −1
0 1 1	0 1 1 1 1 0	1 −1 −1 −1 −1 1
1 0 0	1 0 0 1 1 0	−1 1 1 −1 −1 1
1 0 1	1 0 1 1 0 1	−1 1 −1 −1 1 −1
1 1 0	1 1 0 0 1 1	−1 −1 1 1 −1 −1
1 1 1	1 1 1 0 0 0	−1 −1 −1 1 1 1

Table 2. The number of codewords decoded and used to generate Figs. 6 through 9.

SNR, dB	BCH (31,10)	QR (48,24)	Golay (24,12)	Reed-Muller (32,16)	BCH (63,56)
−7	40,000	6200	6400	6400	6100
−6	40,000	6200	6400	6400	6100
−5	40,000	6200	6400	6400	5900
−4	40,000	6200	6400	6400	5900
−3	40,000	6200	6400	6400	5900
−2	642,100	47,400	58,900	6400	33,300
−1	923,200	70,600	71,700	19,200	57,200
0	1,204,000	83,500	84,500	32,000	83,700
1	1,483,000	207,900	385,000	279,300	279,600
2	3,362,100	409,800	465,000	359,100	468,900
3	5,014,000	1,042,500	1,102,500	995,000	1,036,000
4	6,212,500	1,571,000	2,113,000	1,200,000	1,256,000
5	7,418,000	2,472,500	6,055,000	1,400,000	1,476,000
6	21,242,000	8,048,000	16,330,000	7,200,000	6,512,000
7	12,400,000	8,337,000	8,400,000	8,400,000	7,547,000

Table 3. The number of codewords decoded at an SNR of negative infinity, and the average numbers of expanded nodes, candidate codewords, and edges in the search tree for the A* algorithm.

A* results at SNR of negative infinity	BCH (31,10)	QR (48,24)	Golay (24,12)	Reed-Muller (32,16)	BCH (63,56)
Number of decoded words	89,400	29,900	90,000	59,600	35,600
Average number of expanded nodes	181.16	2705.54	62.45	244.66	79.59
Average number of candidate codewords	120.96	1209.07	28.65	113.34	12.28
Average number of edges in search tree	2902.56	34,428.71	468.68	2302.77	245.12

Table 4. Indicators of decoding complexity for some maximum-likelihood decoding techniques that use a fixed search size or number of calculations for decoding.

Other decoder sizes for comparison to A*	BCH (31,10)	QR (48,24)	Golay (24,12)	Reed-Muller (32,16)	BCH (63,56)
Number of edges in a minimal trellis for the worst permutation of the code	15,356	67,108,860	16,380	262,140	13,052
Number of edges in a minimal trellis for a better permutation of the code		925,692			
Number of edges in a minimal trellis for the best permutation of the code		$\geq$860,156	3580		
Number of binary operations in a trellis with enhancements for the Golay (24,12) code [9, p. 1183]			1351		
Number of edges in a full code tree	23,550	436,207,614	57,342	1,179,646	6.5×10^{17}
Number of edges in an exhaustive search	31,744	805,306,368	98,304	2,097,152	4.5×10^{18}

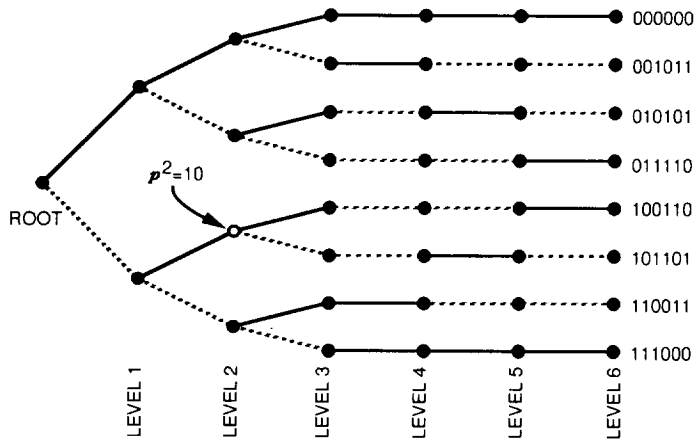


Fig. 1. A binary tree representation of the (6,3) shortened Hamming code.

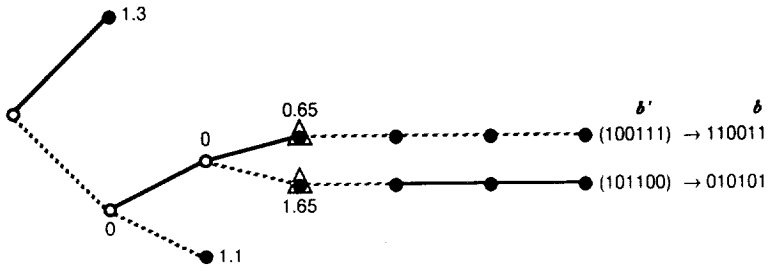


Fig. 2. The tree explored by the A* algorithm when the bit positions are sorted to take advantage of the more reliable symbols first. Expanded nodes are designated by o, and nodes defining candidate codewords are designated by Δ.

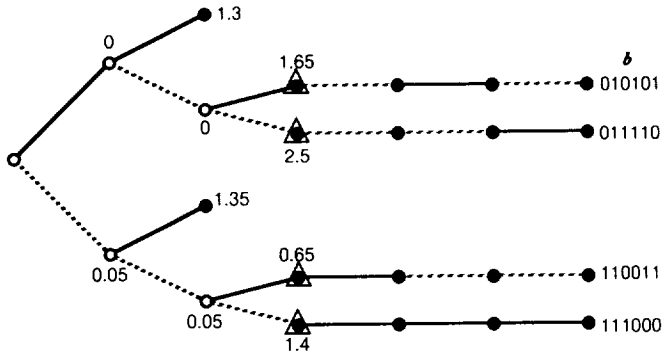


Fig. 3. The tree explored by the A* algorithm when the bit positions are not sorted. Expanded nodes are designated by o, and nodes defining candidate codewords are designated by Δ.

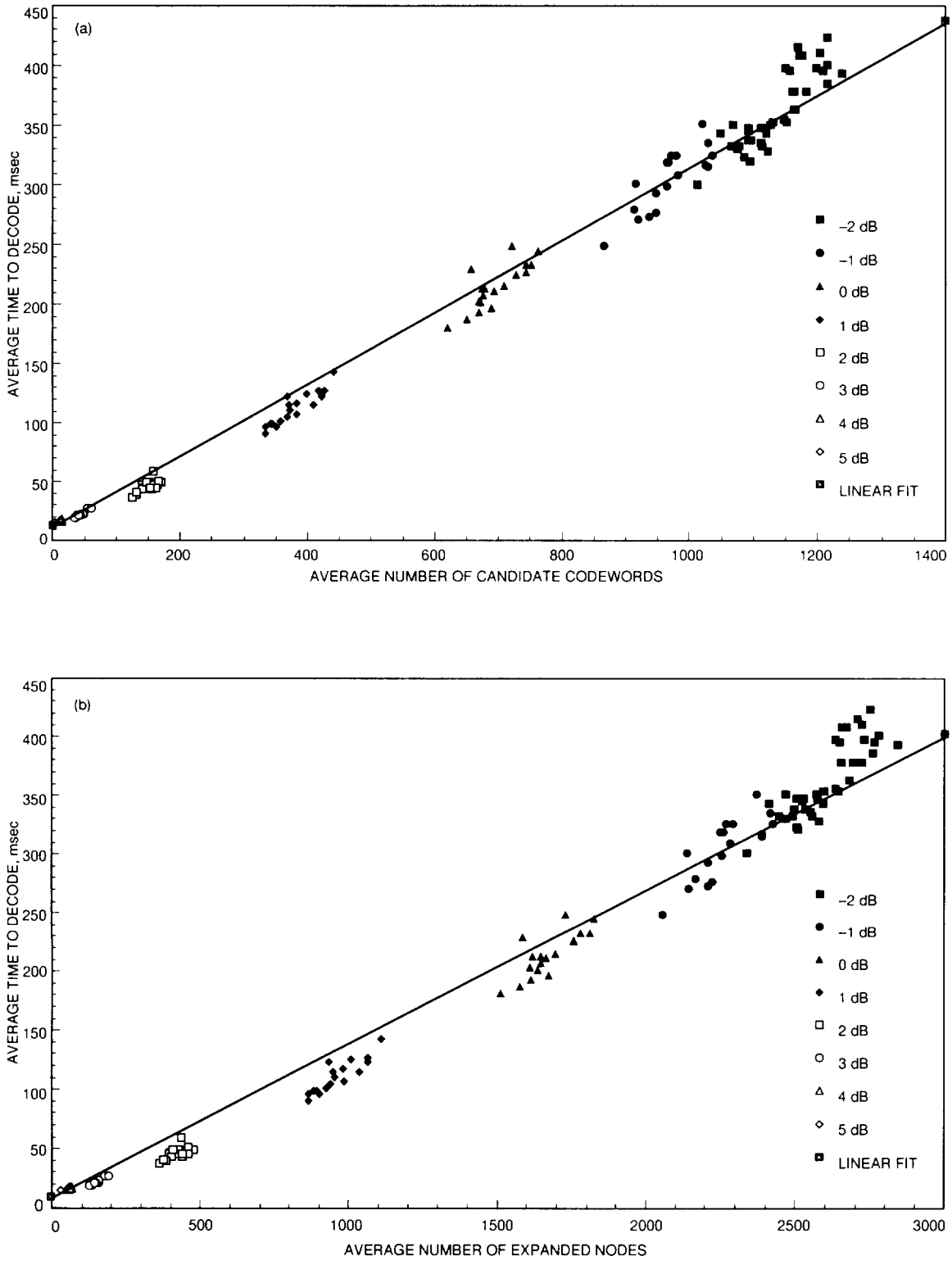


Fig. 4. Scatter plots of average CPU time per decoded word versus (a) average number of candidate codewords per decoded word and (b) average number of expanded nodes per decoded word, for the quadratic residue (48,24) code on a Sparc 10 Model 30 workstation.

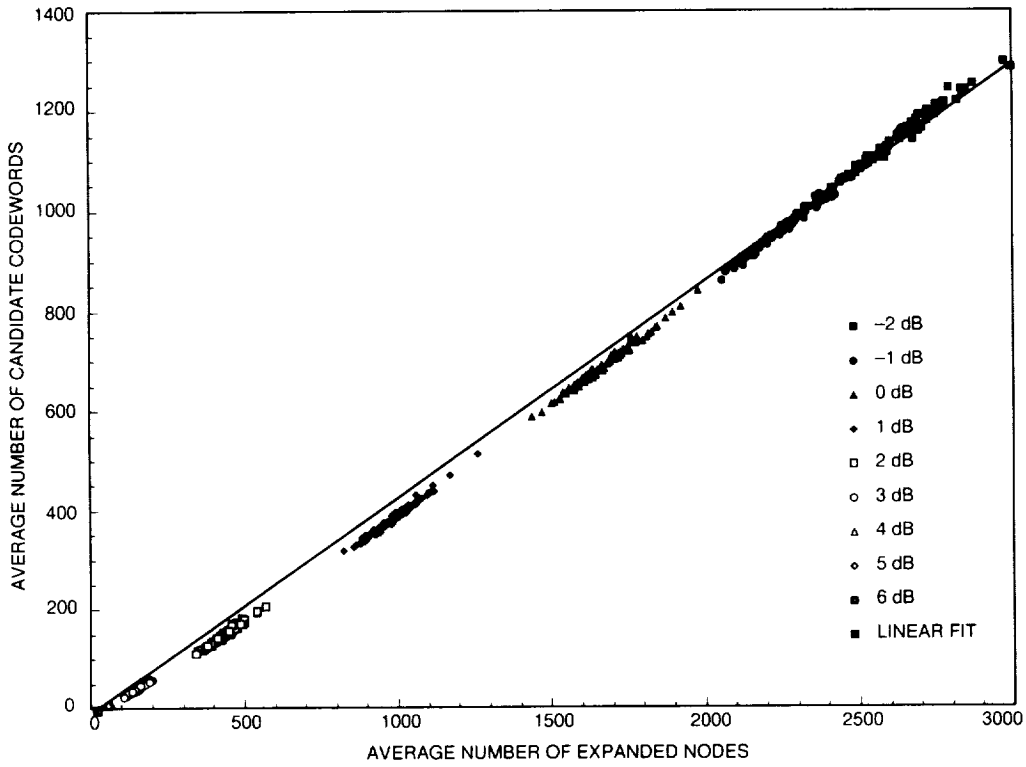


Fig. 5. A scatter plot of the average number of candidate codewords versus the average number of expanded nodes for the quadratic residue (48,24) code, shown with a weighted linear fit.

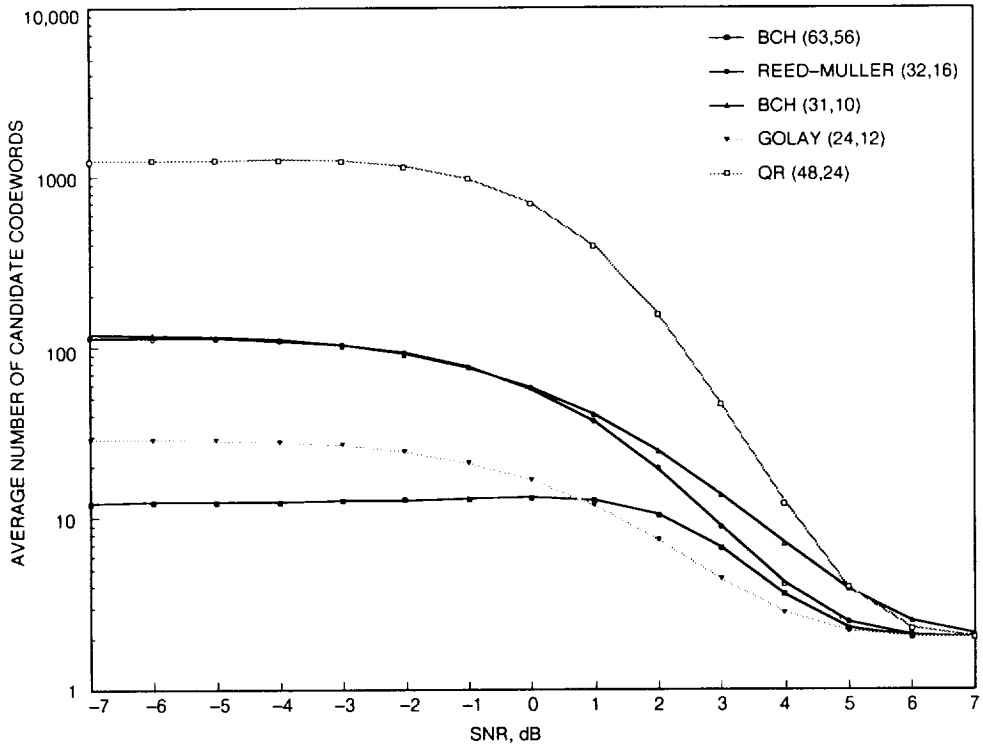


Fig. 6. The average number of candidate codewords for several codes as a function of SNR.

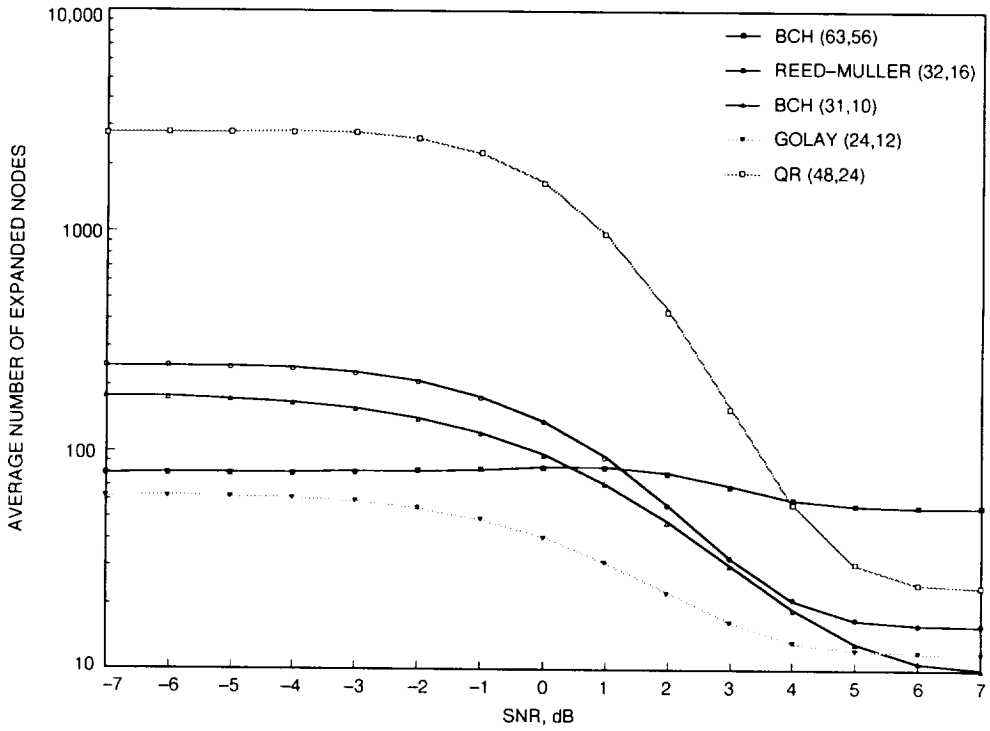


Fig. 7. The average number of expanded nodes for several codes as a function of SNR.

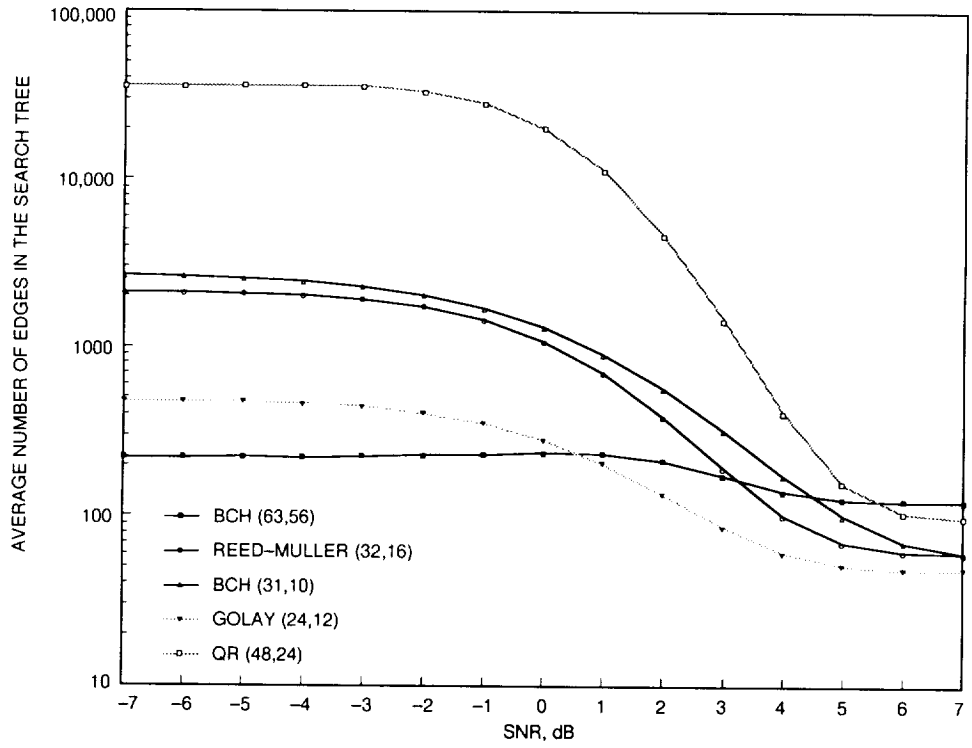


Fig. 8. The average number of edges in the search tree for several codes as a function of SNR.

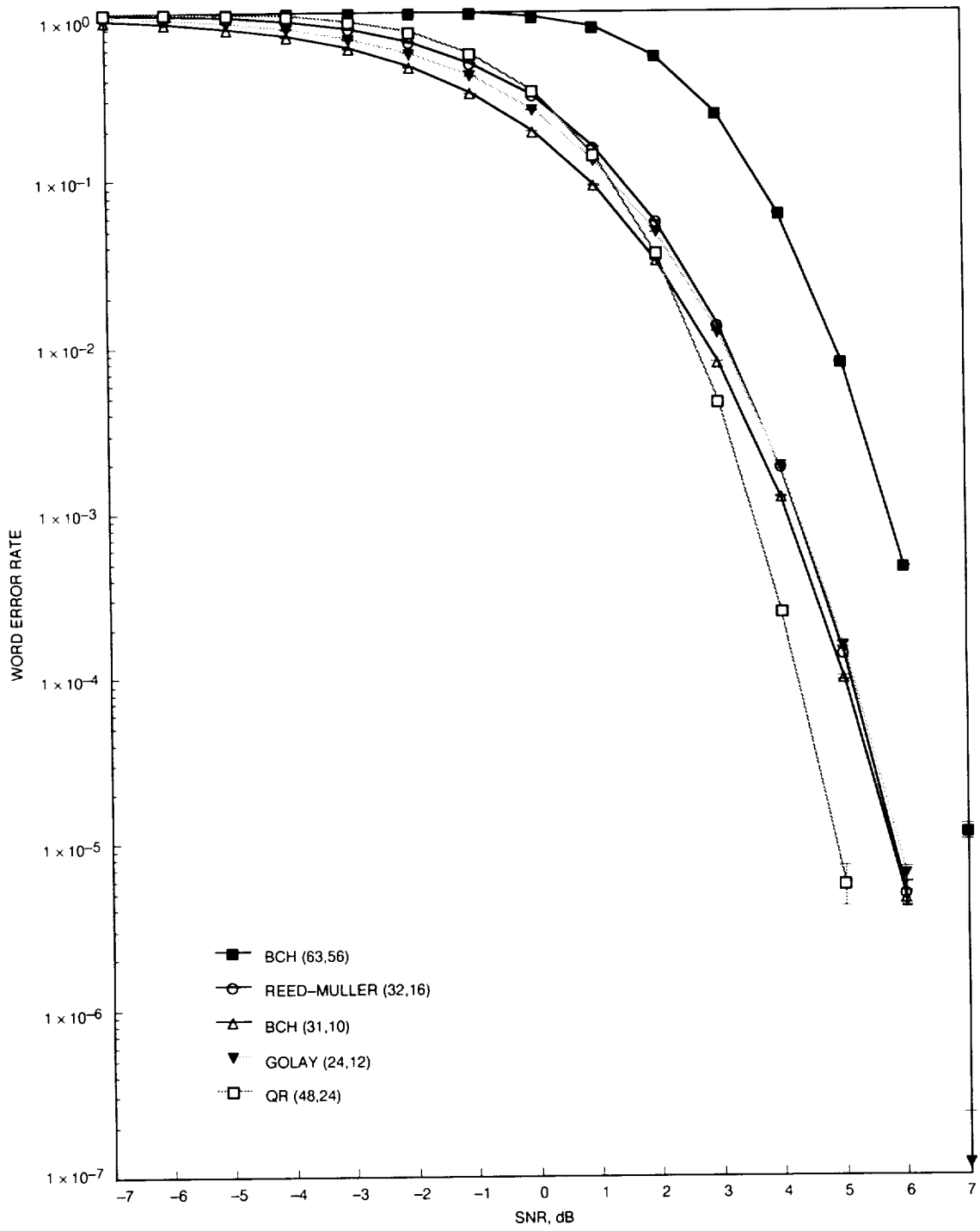


Fig. 9. Probability of word error versus SNR for the BCH (63,56), Reed–Muller (32,16), BCH (31,10), Golay (24,12), and quadratic residue (48,24) codes. The error bars are $\pm\sigma$, one standard deviation.

TDA Progress Report 42-117

May 15, 1994

Bit-Wise Arithmetic Coding for Data Compression

A. B. Kiely
Communications Systems Research Section

This article examines the problem of compressing a uniformly quantized independent and identically distributed (IID) source. We present a new compression technique, bit-wise arithmetic coding, that assigns fixed-length codewords to the quantizer output and uses arithmetic coding to compress the codewords, treating the codeword bits as independent. We examine the performance of this method and evaluate the overhead required when used block-adaptively. Simulation results are presented for Gaussian and Laplacian sources. This new technique could be used as the entropy coder in a transform or subband coding system.

I. Introduction

Many lossy data compression systems consist of separate decorrelation and entropy coding stages. In such schemes, the source data are transformed by some technique (e.g., discrete cosine transform or subband coding) with the goal of producing decorrelated components. Each component is independently scalar quantized and the quantizer output is losslessly compressed. Frequently each component is modeled as a sequence of IID random variables. This model motivates the topic of this article: block-adaptive compression of uniformly quantized samples from an IID source.

The traditional approach to the problem of efficiently encoding the quantizer output symbols is to use a variable-length code, assigning shorter codewords to the more probable symbols. The well-known Huffman coding technique gives the optimal such assignment. This method, however, has some performance limitations. Since each source sample must be mapped to a codeword of length at least 1, the rate of a Huffman code can never be less than 1 bit/sample, no matter how small the entropy. The redundancy of Huffman codes, the difference between rate and entropy, has been studied and bounded by many researchers, e.g., [4]. A common approach to reducing the redundancy at low entropy is to combine Huffman coding with zero-runlength encoding or to encode groups of symbols rather than individual symbols. In this article, we present an alternative solution.

Another problem with Huffman codes is that in block-adaptive situations there may be significant overhead costs (the extra symbols required to identify to the decoder the code being used) [1,5]. Finally, if the buffer overflows (say, if the source is much less compressible than anticipated), we may be forced to discard source samples.

With these problems in mind, we introduce a new technique, bit-wise arithmetic coding. The solution we propose is to assign a *fixed-length* binary codeword to each output symbol in such a way that a zero is more likely than a one in every codeword bit position. We then take the codewords corresponding to several adjacent quantizer output symbols and use a binary arithmetic encoder[1] to

[1] The term "binary encoder" is intended as shorthand for "binary-input binary-output encoder."

encode the first codeword bit for each of these symbols. We repeat this procedure for each group of codeword bits, treating each group independently from the others. This technique can be thought of as a simple progressive transmission scheme using an arithmetic coder to independently encode each level of detail. It turns out that this technique is often surprisingly efficient, despite the fact that interbit dependencies are ignored.

In Section II, we define the uniform quantizer parameters. In Section III, we analyze the block-adaptive binary arithmetic encoder that will be used as part of the bit-wise arithmetic encoding. We make use of this analysis in Section IV, where we examine in detail the the bit-wise arithmetic encoding procedure. We present performance results in Section IV.C.

II. Uniform Quantizer

For several reasons we limit our investigation to the uniform quantizer, not the least of which is simplicity of implementation and analysis. The uniform quantizer often outperforms the Lloyd–Max quantizer in terms of rate-distortion performance (see, e.g., [2]); more important, if we do not know the source statistics a priori, it may be difficult to design a more suitable quantizer. The proposed new method could also be used with a nonuniform quantizer, but the analysis would be less tractable.

A continuous source with probability density function (pdf) $f(x)$ and variance σ^2 is quantized by a uniform quantizer having b bits and bin width $\delta\sigma$, as in Fig. 1. The quantizer output is an index i in the range $1 - 2^{b-1} \leq i \leq 2^{b-1}$, identifying which of the 2^b intervals contains the source sample. A source sample lying in $[T_{i-1}, T_i)$ has reconstruction point $i\delta\sigma$, where the quantizer thresholds are $T_i = (i + \frac{1}{2})\delta\sigma$, for $|i| \leq 2^{b-1} - 1$, and $T_{2^{b-1}} = \infty$, $T_{-2^{b-1}} = -\infty$. We could obtain a lower distortion by using reconstruction points that are equal to the center of mass [with respect to $f(x)$] of each interval, but since we wish to use this quantizer in adaptive situations, we cannot generally compute this quantity.

Note that the quantizer is asymmetric: There is a reconstruction point at the origin, and since there is an even number of points, there is an "extra" reconstruction point, which we arbitrarily choose to place on the positive side. The obvious alternative, a symmetric quantizer that has no reconstruction point at the origin, results in poor performance when δ is large relative to σ^2.

Let p_i denote the probability that the quantizer output is index i, and let d_i denote the contribution to the mean square error (MSE) from the interval $[T_{i-1}, T_i)$:

$$p_i = \int_{T_{i-1}}^{T_i} f(x)dx$$
$$d_i = \int_{T_{i-1}}^{T_i} (x - i\sigma\delta)^2 f(x)dx \tag{1}$$

The MSE is equal to $\sum_i d_i$. It will be convenient to let $\mathcal{P}$ denote the discrete distribution $\{p_i\}_{i=1-2^{b-1}}^{2^{b-1}}$.

In Figs. 2(a) and (b) we plot the resulting rate-distortion curves (computed analytically) for Gaussian and Laplacian sources over a wide range of δ. Note that the curves show optimal performance when the uniform quantizer is used, rather than the theoretical rate-distortion limit, i.e., the rate shown is the entropy of $\mathcal{P}$. The large range of δ causes the peculiar loop in the curve: when $\delta 2^b$ becomes too small relative to σ^2, the performance is poor because the overload bin probabilities become large while their reconstruction points are too close to zero. This situation should be easily avoidable in practice, as we would expect the source range to be finite and known in advance because of hardware constraints.

In general we will assume

$$p_{-i} = p_i, \qquad 0 \leq i \leq 2^{b-1} - 2 \tag{2}$$

$$p_{1-2^{b-1}} = p_{2^{b-1}} + p_{2^{b-1}-1} \tag{3}$$

$$p_i \text{ is nonincreasing in } |i| \tag{4}$$

These conditions are true when $f(x)$ is symmetric about $x = 0$ and nonincreasing with $|x|$, and δ is not unreasonably small.

For sufficiently large δ, the rate-distortion curve is virtually independent of b, as can be seen in Figs. 2(c) and (d), where we plot rate-distortion curves for $b = 4$ and $b = 8$ for Gaussian and Laplacian sources. Increasing b has the effect of increasing the useful range of δ and lengthening the useful portion of the rate-distortion curve.

III. Block-Adaptive Binary Arithmetic Encoding

In this section, we analyze the operation of the block-adaptive binary arithmetic encoder that will be used as part of the bit-wise arithmetic encoding procedure.

A. Binary Arithmetic Encoder Operation

It is well known that a binary arithmetic encoder that is well-tuned to the source can achieve a rate quite close to the source entropy. Our goal in this section is to determine the performance we can expect from an encoder that may not be well-tuned. For more details on arithmetic coding see [6,7,8].

A binary arithmetic encoder with parameter P, the anticipated probability of a zero, maps an N-length sequence of bits s into an interval $[l, r) \subset [0, 1]$ whose width is

$$r - l = P^{NF}(1 - P)^{N(1-F)} \tag{5}$$

where F is the fraction of bits in s that are zero. Ideally, we would like to have $P = F$, but this might not always be possible.

Example 1. Suppose $P = 5/13$, $s = 01$. Initially, $[l, r) = [0, 1)$. We divide this interval into $[0, 5/13)$, $[5/13, 1)$. Note that this first interval has width of $P = 5/13$ of the total interval. On receiving $s_1 = 0$, we assign $[l, r) = [0, 5/13)$ because the symbol with anticipated probability of $5/13$ was received. Again we divide $[0, 5/13)$ into $[0, 25/169)$, $[25/169, 5/13)$. After receiving $s_2 = 1$, we assign $[l, r) = [25/169, 5/13)$. Note that this assignment satisfies Eq. (5). ❐

A K-bit output sequence from the encoder maps to an interval $[i2^{-K}, (i + 1)2^{-K})$ for some $0 \leq i < 2^K - 1$. For our application, since N is known to the decoder, the encoder must specify the largest interval of this form that is contained in $[l, r)$. That is, the encoder must use as few bits as possible to identify to the decoder a sequence beginning with s.

Example 2. Continuing Example 1, after calculating $[l, r) = [25/169, 5/13)$, the encoder must find an interval of the form $[i2^{-K}, (i+1)2^{-K})$ such that $[i2^{-K}, (i+1)2^{-K}) \subseteq [l, r)$. One such interval is $[1/4, 3/8)$, which corresponds to output sequence 010 (because this interval is equal to the set of numbers having binary expansion beginning with 0.010). We will verify in Example 3 that this is in fact the encoder output sequence.

Knowing P, the decoder realizes that $[1/4, 3/8) \subseteq [525/2197, 5/13)$ and that this latter interval corresponds to input sequence 011. Since the decoder also knows that $N = 2$, it takes the first two bits, giving output 01, which is in fact the encoder input sequence. ❐

We continue with the derivation of the encoder rate. We can write $[l, r) = [j2^{-J} + L2^{-J}, j2^{-J} + R2^{-J})$ where $0 \leq L \leq 1/2 \leq R \leq 1$ for integers j and J. That is,

$$[l, r) \subseteq [j2^{-J}, (j + 1)2^{-J}) \tag{6}$$

for maximum J and some j. The first J bits of the output sequence map to the interval $[j2^{-J}, (j + 1)2^{-J})$, and the remaining bits correspond to a subinterval of $[L, R)$. Comparing interval widths in Eq. (6), we find that $2^{-J} \geq r - l = P^{NF}(1 - P)^{N(1-F)}$, which implies $J \geq Nh(P, F)$ where

$$h(P, F) \triangleq -F \log_2 P - (1 - F) \log_2(1 - P)$$

which is a line tangent to the binary entropy function $\mathcal{H}(F)$ at the point $F = P$. If we are very lucky, then $[j2^{-J}, (j + 1)2^{-J}) = [l, r)$ exactly and we are done encoding, in which case the rate is J/N, so the encoder rate R_{arith} (the number of output bits divided by the number of input bits) satisfies $R_{\text{arith}} \geq h(P, F)$.

Usually we will not be so lucky, and we must send additional bits. In the worst case, $L \neq 0$ and $R \neq 1$, in which case we can assign these final bits to be 10^n, which corresponds to $[1/2, 1/2 + 2^{-n-1})$.

Example 3. Continuing Example 2, we can write $[l, r) = [25/169, 5/13) = [0 \times 2^{-1} + (50/169) \times 2^{-1}, 0 \times 2^{-1} + (10/13)2^{-1})$, i.e., $j = 0, J = 1, L = 50/169 < 1/2 < R = 10/13$. The first output bit is 0 (corresponding to $j = 0$). The remaining bits, 10^n (corresponding to $[1/2, 1/2 + 2^{-n-1})$) must map to a subinterval of $[L, R)$, which implies $n = 1$, so the output sequence is 010. ❐

The value n must be sufficiently large that $1/2 + 2^{-n-1} \leq R$, which implies

$$2^{-n-1} \leq R - 1/2 \leq R - L = 2^J P^{NF}(1 - P)^{N(1-F)}$$

making use of Eq. (5) and the fact that $2^{-J}(R - L) = r - l$. Alternatively, we may use 01^n instead of 10^n if this gives smaller n. Consequently,

$$n \leq \left\lceil -\log_2 \left[2^J P^{NF}(1 - P)^{N(1-F)} \right] \right\rceil - 1$$

$$= \lceil Nh(P, F) \rceil - J - 1$$

The total number of encoder output bits is $J + 1 + n$, so we have

$$h(P,F) \leq R_{\text{arith}} \leq \frac{\lceil Nh(P,F) \rceil}{N} \tag{7}$$

which suggests the close approximation

$$R_{\text{arith}} \approx h(P,F) + \frac{1}{2N} \tag{8}$$

B. The Effect of Finite Precision

The bound of Eq. (7) is valid when the encoder can perform arithmetic with arbitrary precision. In any practical system, however, we are limited in our ability to represent the interval endpoints l, r. Whenever $l \geq 1/2$ or $r \leq 1/2$, we can transmit an output bit and rescale the interval (reassign l and r) in the obvious way so that we can make the most of the available resolution. Failing to do this would degrade performance and severely limit the length of input sequences that can be encoded. A consequence of this rescaling is that $l \in [0, 1/2)$ and $r \in (1/2, 1]$, which means that at 9 bits of resolution, which is the resolution used for all simulations and discussions that follow, the values of l and r used by the encoder are always multiples of 2^{-10}.

The effect of the finite resolution is that P, the anticipated probability of a zero, is almost never represented exactly.[2] In fact, P_{eff}, the effective value of P, varies as the algorithm progresses. At each iteration, the interval between l and r is divided into units of length 2^{-10}, and the number of these units is equal to the resolution available to represent P. For example, in the worst case, $l = 1/2 - 2^{-10}$ and $r = 1/2 + 2^{-10}$, producing $P_{\text{eff}} = 1/2$. Note that we can never allow P_{eff} to be zero or one because this would result in an input symbol having no effect on the interval, making decoding impossible. Unfortunately, bounds on rate derived from bounds on P_{eff} are uselessly weak.

Because of the finite resolution, the final interval width is not exactly $P^{NF}(1-P)^{N(1-F)}$ as it is for the arbitrary precision system, but rather $\prod_{i:s_i=0} P_{\text{eff}_i} \prod_{j:s_j=1} (1 - P_{\text{eff}_j})$. Thus, the rate depends not only on P, F, and the resolution available, but also on the particular input sequence since P_{eff_i} depends on $s_1, s_2, \ldots, s_{i-1}$. The consequence is that Eq. (8) is a slightly optimistic estimate of the rate.

Let k equal the number of extra encoded bits resulting from the limited resolution, when compared to the sequence length predicted by Eq. (8), so that

$$R_{\text{arith}} = h(P,F) + \frac{(k+1/2)}{N}$$

At 9-bit resolution, simulations indicate that k has an expected value of approximately 0.32 bits, and standard deviation of 1.01, nearly independent of P, F, and N.[3] Thus, as an approximation to the rate we use

$$R_{\text{arith}} \approx h(P,F) + 0.82/N \tag{9}$$

for a system with 9-bit resolution. Increasing the resolution should have the effect of reducing the variance and expected value of k. Figure 3 gives an example of encoder performance when P is fixed.

C. Overhead

We would like to use binary arithmetic encoding block-adaptively to transmit a sequence of N bits, i.e., the encoder output sequence is preceded by overhead bits that identify to the decoder the value of P being used. By using $\log_2 N$ bits of overhead, we could specify $P = F$ exactly, but by using fewer bits we can exchange accuracy for lower overhead. In this section, we will explore this tradeoff, showing how to find the optimal number of overhead bits.

Assume for now that we have a fixed number of overhead bits m. We select an ordered set of $M = 2^m$ probabilities $\{\rho_1, \rho_2, \ldots, \rho_M\}$, known to the encoder and decoder in advance, that can be used as values for P. We first show how to find the optimal assignment of these probability points.

Given the set $\{\rho_1, \rho_2, \ldots, \rho_M\}$, the encoder would choose to use the ρ_i that minimizes the rate, i.e., the ρ_i that minimizes $h(\rho_i, F)$. As illustrated in Fig. 4, this amounts to using line segments to approximate the binary entropy function. Let t_i denote the solution to $h(\rho_i, t_i) = h(\rho_{i+1}, t_i)$ and define $\Delta_i \stackrel{\Delta}{=} h(\rho_i, t_i) - \mathcal{H}(t_i)$, which is the redundancy (additional rate) at this "corner" point t_i. Clearly if we choose the ρ_i's to minimize $\max_i \Delta_i$, then we minimize the maximum redundancy over all possible values of F.

Suppose that for some i, we have $\Delta_{i-1} > \Delta_i$, as in Fig. 4. Then by decreasing ρ_i, we can decrease Δ_{i-1} and increase Δ_i. From this argument we can see that the assignment of the ρ_i is optimal when Δ_i is the same for all

[2] There are some trivial exceptions to this, such as when $P = 1/2$.

[3] When P is very close to 0 or 1, the encoder rate is less well behaved. For example, if $P < 2^{-10}$, then clearly P_{eff} will always exceed P. We ignore these extreme cases.

i. We call this optimal quantity $\Delta^*(M)$. Thus, we want to find M, the minimum number of line segments required to approximate the binary entropy function to within Δ^* everywhere.

Given $\Delta^*(M)$, we find that $\rho_1 = 1 - 2^{-\Delta^*(M)}$. Given ρ_i, we find t_i by solving

$$h(\rho_i, t_i) - \mathcal{H}(t_i) = \Delta^*(M)$$

Given t_i, we find ρ_{i+1} by solving

$$h(\rho_{i+1}, t_i) - \mathcal{H}(t_i) = \Delta^*(M)$$

i.e., we solve the same equation in the other direction. This procedure can be used to find the optimal set of ρ_i and t_i given $\Delta^*(M)$, or in an iterative procedure to compute $\Delta^*(M)$. We can also take advantage of the fact that the optimal ρ_i must be symmetric about 1/2. In Table 1, $\Delta^*(M)$ is given for several values of M.

To find the relationship between Δ^* and M for large M, let $w(\rho)$ denote the spacing between adjacent probability points in the neighborhood of ρ so that the "corner" point is $t \approx \rho + w/2$. At this point,

$$\Delta^* \approx h(\rho, \rho + w/2) - \mathcal{H}(\rho + w/2) \approx \frac{w^2}{8\rho(1-\rho)\ln 2}$$

using the first three terms of the Taylor series expansion of $\mathcal{H}(\rho)$. Now $1/w(\rho)$ is the density of probability points, so

$$M \approx \int_0^1 \frac{1}{w(\rho)} d\rho \approx \frac{1}{\sqrt{8\Delta^* \ln 2}} \int_0^1 \frac{d\rho}{\sqrt{\rho(1-\rho)}}$$

$$= \frac{\pi}{\sqrt{8\Delta^* \ln 2}}$$

thus for large M,

$$\Delta^*(M) \approx \left(\frac{\pi^2}{8 \ln 2}\right) M^{-2} \tag{10}$$

In Fig. 5 we show the exact and approximate relationship between M and Δ^*.

The encoding operation is straightforward. The encoder computes F by counting the number of zeros in the input sequence. The largest integer I satisfying $t_{I-1} < F$ gives the optimal ρ_I, which is the parameter used in the encoding procedure. The encoder uses m bits to identify I, followed by the arithmetic encoder output sequence. This procedure guarantees that

$$h(\rho_I, F) \leq \mathcal{H}(F) + \Delta^*(2^m)$$

The rate (including overhead) required to block-adaptively transmit a sequence of N bits is approximately

$$\mathcal{H}(F) + m/N + \Delta^*(2^m) + 0.82/N \tag{11}$$

where $\mathcal{H}(F)$ comes from the source uncertainty; m/N comes from the m overhead bits used to identify ρ_I; $\Delta^*(2^m)$ is a result of using m bits to specify F approximately instead of using $\log_2 N$ bits to specify F exactly (this amounts to a worst-case assumption); and $0.82/N$ comes from the finite resolution of the encoder [see Eq. (9)].

Given N, the optimal number of overhead bits is

$$m^*(N) = \min_m{}^{-1}\{\mathcal{H}(F) + 0.82/N + \Delta^*(2^m) + m/N\}$$

$$= \min_m{}^{-1}\{\Delta^*(2^m) + m/N\} \tag{12}$$

which is tabulated in Table 2.

Using Eqs. (10) and (12), we find that for large m, the optimal value of m satisfies

$$0 = \frac{\partial}{\partial m}[\Delta^*(2^m) + m/N] \approx \frac{1}{N} - \pi^2 2^{-2-2m}$$

which gives

$$m^*(N) \approx \frac{1}{2}\log_2 N + \log_2 \pi - 1 \tag{13}$$

and

$$\Delta^*\left(2^{m^*(N)}\right) \approx \frac{1}{2N \ln 2} \tag{14}$$

Substituting into Eq. (11), we find that the rate of a block-adaptive binary arithmetic encoder, including overhead, is approximately

$$\mathcal{H}(F) + \frac{1}{N}\left[\frac{1}{2}\log_2 N + 0.18 + \log_2 \pi + \frac{1}{2\ln 2}\right] \qquad (15)$$

IV. Bit-Wise Arithmetic Coding

A. Operation of the Bit-Wise Arithmetic Encoder

We now use the results of Section III to analyze the performance of bit-wise arithmetic encoding. First we outline the encoding procedure.

Each quantizer output symbol is mapped to a b-bit codeword. The first bit indicates the sign of the quantizer reconstruction point.[4] The remaining bits are assigned to quantizer levels in increasing lexicographic order as we move away from the origin. Figure 6 illustrates this mapping for $b = 4$. Because of Eqs. (2)–(4), a zero will be more likely than a one in every bit position.

Codewords corresponding to N adjacent source samples are grouped together. The N sign bits of the codeword sequence are encoded using the block-adaptive binary arithmetic encoder analyzed in Section III. Then the N next-most-significant bits are encoded, and so on. This can be viewed as a simple progressive transmission system—each subsequent codeword bit gives a further level of detail about the source. Each bit sequence is encoded independently—at the ith stage the arithmetic coder calculates (approximately) the *unconditional* probability that the ith codeword bit is a zero.

The obvious loss is that we lose the benefit of interbit dependency. For example, the probability that the second bit is a zero is not generally independent of the value of the first bit, though the encoding procedure acts as if it were. Huffman coding does not suffer from this loss, which we examine in Section IV.B.

The advantage is that for many practical sources, this technique has lower redundancy than Huffman coding, because the arithmetic coder is not required to produce an output symbol for every input symbol. Also this scheme is relatively simple[5] and has some advantages in terms of overhead: because the number of codewords is 2^b, the overhead of block-adaptive Huffman coding increases exponentially in b unless we are able to cleverly exploit additional information about the source [5]. By contrast, the overhead required for bit-wise arithmetic encoding increases linearly in b because the codeword bits are treated independently.

Another advantage is that this technique gives us a simple means of handling situations where we are rate constrained (or equivalently, buffer-constrained): We simply encode the blocks of N bits until the allocated rate is exhausted (or the buffer is full). The distortion is automatically reduced for "more compressible" sources—when the most significant bits can be efficiently encoded, we are able to send additional (less-significant) bits, so the encoder resolution increases automatically. This would mean, for example, that a block having 6-bit resolution might be followed by a block having only 8-bit resolution. The advantage is that we hope to prevent any sample from having zero-bit resolution.

The obvious question is whether the gains offset the losses. Given δ, b, and $f(x)$, we can use Eq. (1) to compute $\mathcal{P}$, the distribution on the quantizer output symbols, and use this result to compute π_i, the probability that the ith bit is a zero. For example, for $b = 4$,

$$\begin{bmatrix}\pi_1\\ \pi_2\\ \pi_3\\ \pi_4\end{bmatrix} =$$

$$\begin{bmatrix}1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0\\ 0&0&0&0&1&1&1&1&1&1&1&1&0&0&0&0\\ 0&0&1&1&0&0&1&1&1&1&0&0&1&1&0&0\\ 0&1&0&1&0&1&0&1&1&0&1&0&1&0&1&0\end{bmatrix}$$

$$\times \begin{bmatrix}p_{-7}\\ p_{-6}\\ \vdots\\ p_8\end{bmatrix} = \begin{bmatrix}1&1&1&1&1&1&1&1&1\\ 1&2&2&2&1&0&0&0&0\\ 1&2&1&0&1&2&1&0&0\\ 1&1&1&1&1&1&1&1&0\end{bmatrix} \begin{bmatrix}p_0\\ p_1\\ \vdots\\ p_8\end{bmatrix}$$

using the symmetries of the p_i described in Section II. It turns out that $\pi_1 = (1 + p_0)/2$, $\pi_b = \pi_1 - p_{2^{b-1}}$. The rate obtained is (assuming for the moment that we have an idealized system where we may neglect overhead)

$$\sum_{i=1}^{b} \mathcal{H}(\pi_i)$$

[4] Or, to be precise, the first bit indicates whether the reconstruction point is positive.

[5] To the extent that a binary arithmetic coder is simple.

For the Gaussian and Laplacian sources, this result can be compared to Huffman coding in Figs. 7 and 8. Over the useful range of the quantizer, the rate of the bit-wise arithmetic coding scheme is quite close to the entropy.

B. The Redundancy of Bit-Wise Transmission

We have already examined some of the sources that contribute to the rate of bit-wise arithmetic encoding: overhead bits, finite precision arithmetic, and the use of approximate rather than exact representation of F. We now examine the most obvious redundancy component: the added rate that results from treating each bit independently.

Let β_i be the random variable that is equal to the value of the ith bit of the codeword corresponding to the quantizer output. The source entropy is then the entropy of the β_i, $H(\beta_1\beta_2\cdots\beta_b)$, but the independent encoding of the bits results in a rate of $H(\beta_1)+H(\beta_2)+\cdots H(\beta_b)$. Thus, the redundancy is

$$\begin{aligned}\mathcal{R} &= \sum_{i=1}^{b} H(\beta_i) - H(\beta_1\beta_2\cdots\beta_b) \\ &= \sum_{i=1}^{b} H(\beta_i) - [H(\beta_1) + H(\beta_2|\beta_1) \\ &\quad + \cdots H(\beta_n|\beta_{n-1}\beta_{n-2}\cdots\beta_1)] \\ &= \sum_{i=2}^{b} I(\beta_i, \beta_{i-1}\beta_{i-2}\cdots\beta_1) \qquad (16)\end{aligned}$$

where I is the mutual information function. So, for example, if $b=2$, the redundancy is equal to the mutual information between β_1 and β_2.

Of course, $\mathcal{R}\geq 0$, with equality if and only if the β_i are independent. This bound is tight—zero redundancy occurs, for example, if the p_i are distributed according to the two-sided exponential distribution $p_i=\alpha\theta^{|i|}$, which is a distribution suggested in [3] as a model for certain real-world sources. The exponential distribution is close to the distribution obtained for a Laplacian source, which explains why bit-wise arithmetic coding works well for this source.

A greedy assignment of bits results in almost exactly the same mapping from quantizer output symbol to codeword. In the first bit position, assign a 0 to the 2^{b-1} indices having the largest p_i, and a 1 to the others. In the second bit position, among the quantizer output indices having the same value in the first bit position, assign a 0 to the 2^{b-2} indices having the largest p_i, and so on. In this manner, the sign bit is the last codeword bit.

It should be noted that other codeword assignments, for example, assigning codewords so that Hamming weight is strictly nonincreasing in $|i|$, can sometimes give lower redundancy. Unless the distribution of the p_i is known a priori, we cannot in general determine the optimal assignment. The codeword assignment proposed here has the advantages of symmetry and usefulness from a progressive transmission or buffer-constrained standpoint.

If we relax the assumptions of Eqs. (2)–(4), then the pathological case $p_0=p_{2^{b-1}}=1/2$ gives the maximum possible redundancy of $\mathcal{R}_{\max}=b-1$. We would like to have a tight upper bound on redundancy when we maintain Eqs. (2)–(4). For $b=2$, it is simple to verify that the maximum redundancy occurs when $\mathcal{P}=\{p_{-1},p_0,p_1,p_2\}=\{1/3,1/3,1/3,0\}$, which gives redundancy of $\log_2 3-4/3\approx 0.252$. For larger b, it becomes more difficult to determine analytically what distribution gives maximum redundancy.

Example 4. If $b=3$, the restrictions of Eqs. (2)–(4) imply that any valid distribution $\mathcal{P}$ can be written as a convex combination of the (mostly) uniform distributions

$$\mathcal{P}_1=\{0,0,0,1,0,0,0,0\}$$

$$\mathcal{P}_2=\{0,0,1/3,1/3,1/3,0,0,0\}$$

$$\mathcal{P}_3=\{0,1/5,1/5,1/5,1/5,1/5,0,0\}$$

$$\mathcal{P}_4=\{1/7,1/7,1/7,1/7,1/7,1/7,1/7,0\}$$

$$\mathcal{P}_5=\{1/7,1/7,1/7,1/7,1/7,1/7,1/14,1/14\}$$

Thus, we wish to maximize the redundancy function of Eq. (16) over the convex hull of $\{\mathcal{P}_1,\mathcal{P}_2,\mathcal{P}_3,\mathcal{P}_4,\mathcal{P}_5\}$. Simulations suggest that the maximum occurs at $\mathcal{P}_3$, which gives redundancy of $2/5[5\log_2 5-3\log_2 3-6]\approx 0.342$. Unfortunately, $\mathcal{R}$ is not convex $\cup$ in $\mathcal{P}$, so we are not certain that for arbitrary b maximum redundancy occurs for some uniform distribution.

We conjecture that for any b, the maximum redundancy subject to Eqs. (2)–(4) occurs for some uniform $f(x)$. If

this conjecture is correct, then to find a bound on redundancy we examine redundancy in the limit $b \to \infty$. Consider what happens when the quantizer range is $[-1/2, 1/2]$ and $f(x)$ is uniform over $[-W/2, W/2]$, for some $0 < W < 1$, as in Fig. 9. The lines and spaces in the figure replace the codeword assignment matrix: the lines denote a one in the corresponding bit position and the gaps denote a zero (compare to Fig. 6). Note that in the limit the codeword assignment is symmetric about $x = 0$.

For a large fixed b, $\mathcal{R}_u^b$, the redundancy for a uniform distribution is

$$\mathcal{R}_u^b(W) = \sum_{i=1}^{b} \mathcal{H}(\pi_i) - H(\mathcal{P}) = \sum_{i=1}^{b} \{\mathcal{H}(\pi_i) - 1\} - \log_2 W$$

using the fact that $H(\mathcal{P}) = \log_2(W2^b) = b + \log_2 W$, which is a consequence of the uniform distribution. Given i, $1 - \pi_i$ is equal to the sum of the lengths of the darkened line segment portions divided by W. We ignore the sign bit because $\pi_{\text{sign}} = 1/2$, so this bit makes no contribution to the redundancy in the limit. Examining Fig. 9, we can see that the interval $[-W/2, W/2]$ will always contain either an integer number of line segments or an integer number of gaps. Thus, either π_i or $1 - \pi_i$ will be equal to $\lfloor 1/2 + W2^{i-1} \rfloor / (W2^i)$, so $\mathcal{H}(\pi_i) = \mathcal{H}\left(\lfloor 1/2 + W2^{i-1} \rfloor / W2^i\right)$ and in the limit the redundancy is

$$\mathcal{R}_u^\infty(W) = -\log_2 W - \sum_{i=1}^{\infty} f_i(W) \tag{17}$$

where

$$f_i(W) \triangleq 1 - \mathcal{H}\left(\frac{\lfloor 1/2 + W2^{i-1} \rfloor}{W2^i}\right)$$

Note that we can limit our analysis to the case where $W > 1/2$ without loss of generality, because $\mathcal{R}_u^\infty(W/2^n) = \mathcal{R}_u^\infty(W)$ for any integer n. Figure 10 shows $\mathcal{R}_u^\infty(W)$, and several of the f_i are shown in Fig. 11. If $W = j/2^n$ for integer j and n, then only the first n terms in the summation of Eq. (17) are nonzero.

The function $\mathcal{R}_u^\infty(W)$ attains a maximum of approximately 0.34544 near $W^* \approx 0.610711$. It is difficult to determine an analytic expression for W^* or $\mathcal{R}_u^\infty(W^*)$, in part because the first derivative of $\mathcal{R}_u^\infty(W)$ is discontinuous at infinitely many points in $[1/2, 1]$. Given any $\mathcal{R}' < \mathcal{R}_u^\infty(W^*)$, we can find a uniform $f(x)$ and finite b producing redundancy $\mathcal{R}'$ or higher. We conjecture that in general

$$\mathcal{R} < \mathcal{R}_u^\infty(W^*) \approx 0.34544$$

is a tight upper bound for any pdf producing $\mathcal{P}$ satisfying Eqs. (2)–(4). It is interesting that this bound is independent of b, while without any restrictions on $\mathcal{P}$ the redundancy can be as large as $b - 1$.

C. Performance

Including all of the overhead effects, using Eqs. (11), (13), and (14), the rate of the bit-wise arithmetic coder is approximately

$$\begin{aligned} R_{\text{bit-arith}} &\approx H(\beta_1\beta_2\cdots\beta_b) + \mathcal{R} \\ &\quad + b\left[\Delta^*\left(2^{m^*(N)}\right) + \frac{m^*(N) + 0.82}{N}\right] \\ &\approx H(\beta_1\beta_2\cdots\beta_b) + \mathcal{R} \\ &\quad + \frac{b}{N}\left[\frac{1}{2\ln 2} + 0.18 + \frac{1}{2}\log_2 N + \log_2 \pi\right] \\ &\approx H(\beta_1\beta_2\cdots\beta_b) + \mathcal{R} + \frac{b}{N}\left[2.55 + \frac{1}{2}\log_2 N\right] \end{aligned} \tag{18}$$

In Fig. 12, we plot theoretical and simulated rate-distortion curves for the bit-wise arithmetic coder applied to Gaussian and Laplacian sources. Bit-wise arithmetic coding performs particularly well on the Laplacian source. Note that the increase in rate when we decrease N from 512 to 256 is rather small. This is not surprising considering Eq. (18). The use of smaller N implies faster adaptability to a nonstationary source, and also gives advantages when using a small buffer.

D. Possible Refinements

There are a few tricks that we might use to further reduce the rate in a practical system:

(1) The arithmetic encoder could adaptively estimate the probability rather than simply count the number of zeros in a sequence. This might improve adaptivity to nonstationary sources, in addition to reducing overhead.

(2) The relative frequency of zeros in a sequence, F, must be a multiple of $1/N$, so not all "corners" (i.e., the t_i in Fig. 4) can be reached. Keeping this fact in mind, we could adjust the ρ_i values to obtain a slight improvement in performance: lower $\Delta^*(m)$ and perhaps lower $m^*(N)$. The rate reduction would probably be minuscule except at very small N.

(3) Additional savings may be obtained by considering the variations in rate due to the finite precision of the encoder. When F is near a corner, we could (time and complexity permitting) encode the sequence using the two nearest ρ_i to see which produces a shorter sequence.

(4) We could make the ρ_i more dense in the regions that are more probable. For example, if Eqs. (2)–(4) are satisfied, then the probability of a zero will always be higher than the probability of a one, so we could require most of the ρ_i to be less than 1/2.

(5) Since the decoder knows N, t_{I-1}, and t_I before it decodes, it knows that the sequence must contain between $\lceil Nt_{I-1} \rceil$ and $\lfloor Nt_I \rfloor$ zeros. In the current implementation, the decoder does not explicitly exploit this information. We could update P as we encode/decode, taking into account the number of zeros that must remain in the sequence, "like counting cards in Vegas," comments Sam Dolinar.

(6) We might also get a slight improvement by combining the overhead of blocks, thus not requiring that M, the number of ρ_i, be a power of two.

(7) We could require that $\rho_1 = 0$, so that if a block consisted of the all-zeros sequence, we could encode the entire sequence simply by using the m overhead bits to identify $P = \rho_1$.

(8) It might be convenient to keep track of the encoder output sequence length during the encoder operation, and send the data unencoded if the length exceeds N. This corresponds to forcing one of the ρ_i to be equal to 1/2. Once this happens, we might assume that all remaining bit positions are sufficiently random so that we are better off sending them unencoded and saving the overhead. This reflects the conventional wisdom that for many real-world sources, quantized samples are often compressible only in the most-significant bits.

V. Conclusion

The bit-wise arithmetic encoding technique provides a simple method for data compression. The independent treatment of the codeword bits provides its main assets: The technique is simple, it can be used in progressive transmission or as a means of alleviating buffer overflow problems, and it has low overhead that increases linearly in the number of quantizer bits rather than exponentially. For the Gaussian and Laplacian sources, the rate is quite close to the entropy. The independent treatment of bits can also be its greatest liability—for sources where the codeword bits are highly correlated, the redundancy can be substantial. As with any data compression method, the usefulness of this technique ultimately depends on the source to be compressed.

Acknowledgments

The author is grateful to Sam Dolinar, Kar-Ming Cheung, Fabrizio Pollara, and Laura Ekroot for their helpful comments.

References

[1] Y. S. Abu-Mostafa and R. J. McEliece, "Maximal Codeword Lengths in Huffman Codes," *The Telecommunications and Data Acquisition Progress Report 42-110, vol. April-June 1992*, Jet Propulsion Laboratory, Pasadena, California, pp. 188–193, August 15, 1992.

[2] T. Berger, *Rate-Distortion Theory: A Mathematical Basis for Data Compression*, Englewood Cliffs, New Jersey: Prentice Hall, 1971.

[3] K. Cheung, P. Smyth, and H. Wang, "A High-Speed Distortionless Predictive Image Compression Scheme," *The Telecommunications and Data Acquisition Progress Report 42-102, vol. April-June 1990*, Jet Propulsion Laboratory, Pasadena, California, pp. 73–90, August 15, 1990.

[4] R. G. Gallager, "Variations on a Theme by Huffman," *IEEE Transactions on Information Theory*, vol. IT-24, no. 6, pp. 668–674, November 1978.

[5] R. J. McEliece and T. H. Palmatier, "Estimating the Size of Huffman Code Preambles," *The Telecommunications and Data Acquisition Progress Report 42-114, vol. April-June 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 90–95, August 15, 1993.

[6] J. J. Rissanen, "Generalized Kraft Inequality and Arithmetic Coding," *IBM Journal of Research and Development*, vol. 20, no. 3, pp. 198–203, May 1976.

[7] J. J. Rissanen and G. G. Langdon, Jr., "Universal Modeling and Coding," *IEEE Transaction on Information Theory*, vol. IT-27, no. 1, pp. 12–23, January 1981.

[8] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic Coding for Data Compression," *Communications of the ACM*, vol. 30, no. 6, pp. 520–540, June 1987.

Table 1. The optimal relationship between M and Δ.

M	$\Delta^*(M)$
2^0	1.0
2^1	0.32193
2^2	0.093506
2^3	0.025407
2^4	0.0066389
2^5	0.0016980
2^6	0.00045745
2^7	0.00010800
2^8	0.000027077

Table 2. The optimal relationship between block length N and overhead bits m.

N	$m^*(N)$
1	0
[2, 4]	1
[5, 14]	2
[15, 53]	3
[54, 202]	4
[203, 806]	5
[807, 2861]	6
[2862, 12358]	7

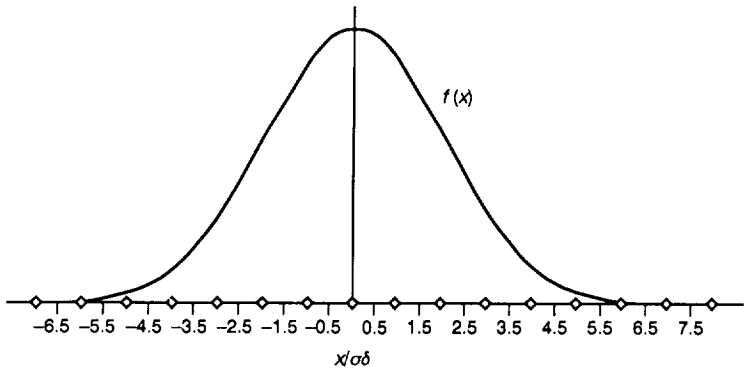


Fig. 1. Example of a pfd and reconstruction points for a four-bit uniform quantizer.

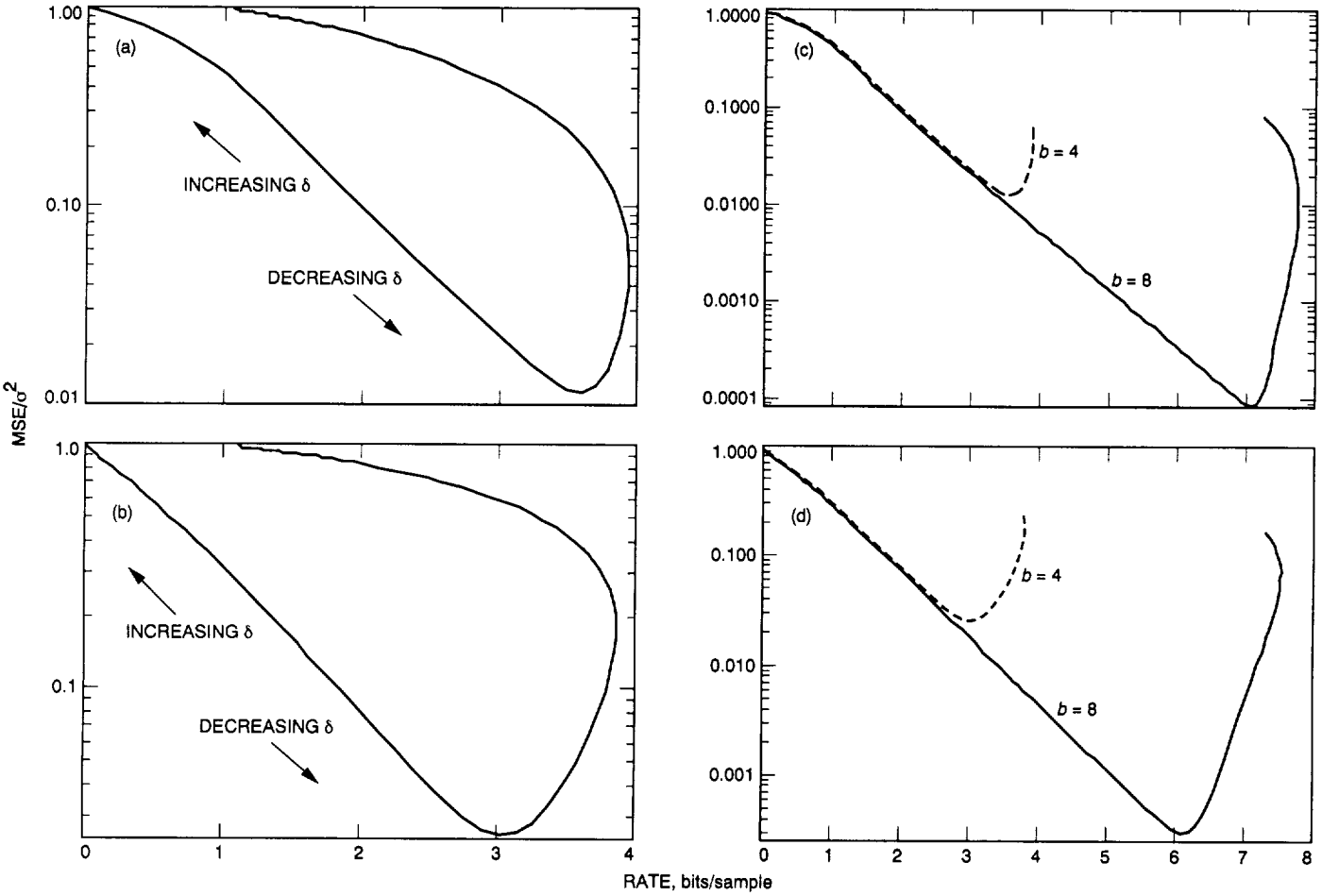


Fig. 2. Rate distortion for the uniform quantizer over a wide range of δ: (a) Gaussian source, $b = 4$; (b) Laplacian source, $b = 4$; (c) Gaussian source, $b = 4$ and $b = 8$; and (d) Laplacian source, $b = 4$ and $b = 8$.

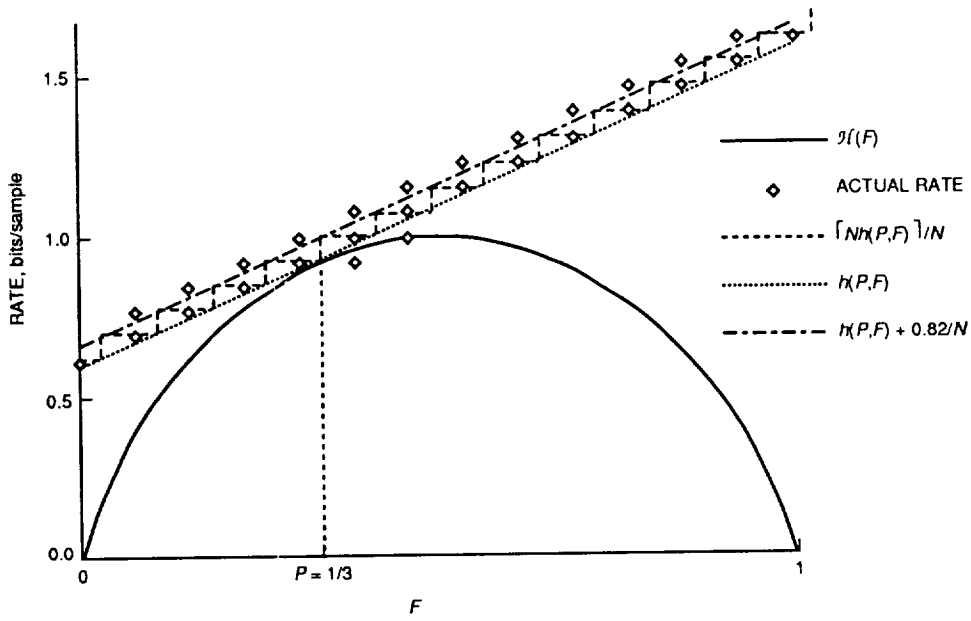


Fig. 3. Binary entropy, predicted and actual performance of the 9-bit binary arithmetic encoder as a function of F for all possible sequences of length $N = 13$ when $P = 1/3$.

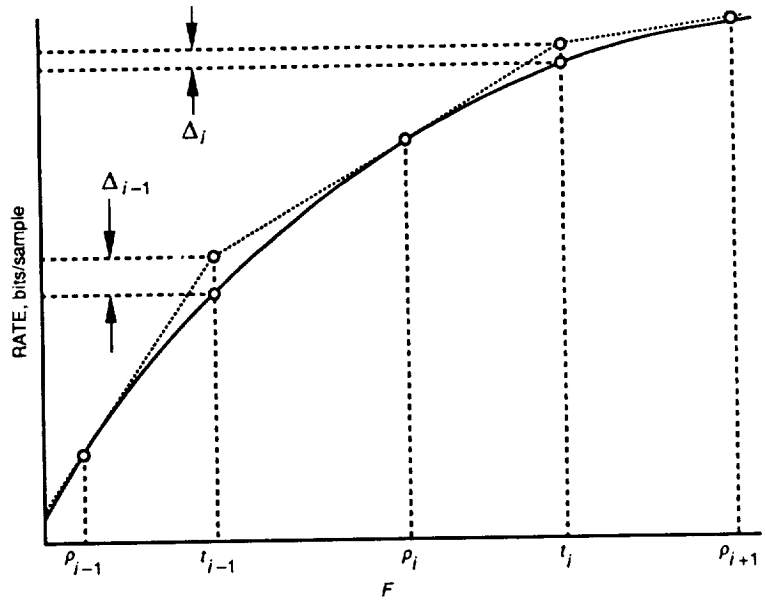


Fig. 4. Binary entropy and line-segment approximation.

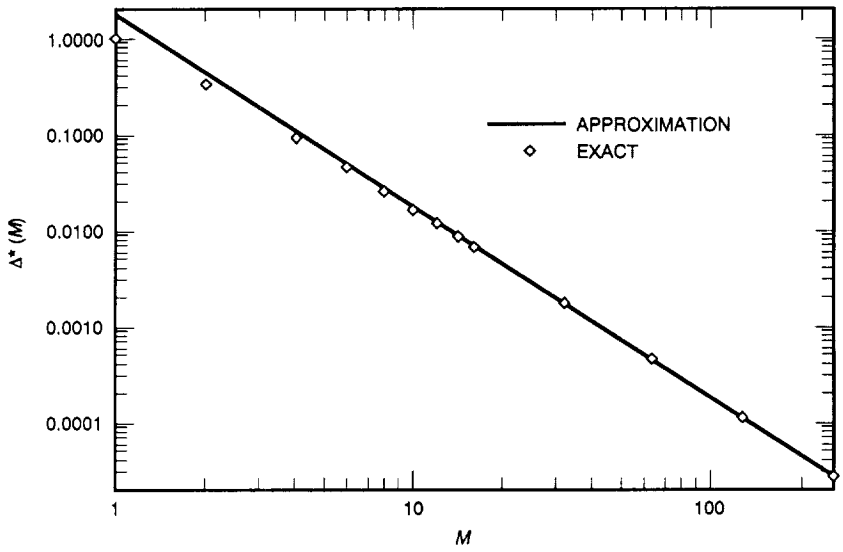


Fig. 5. $\Delta^*(M)$ and approximation $(\pi^2/8 \ln 2)\, M^{-2}$.

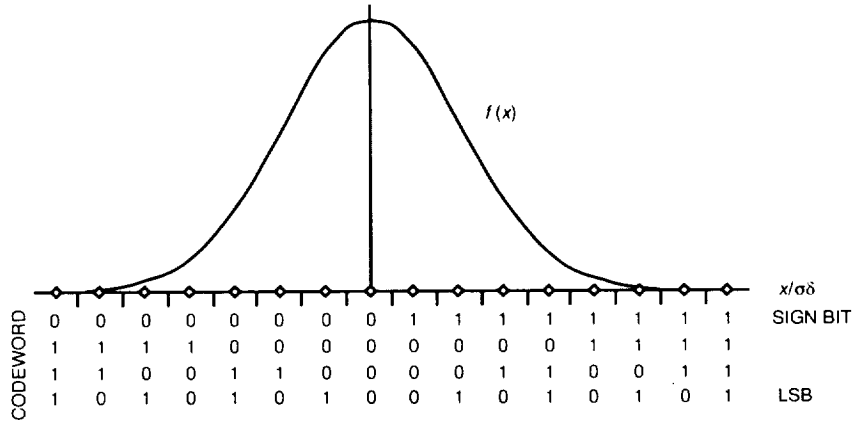


Fig. 6. Codeword assignment for the four-bit quantizer.

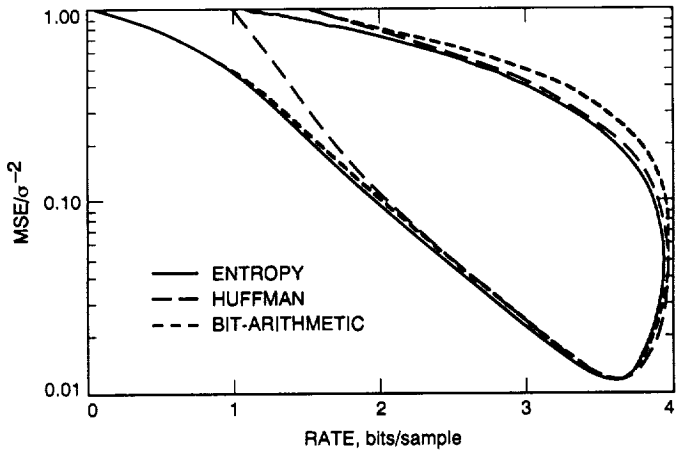


Fig. 7. Bit-wise arithmetic coding, Huffman coding, and entropy for Gaussian source, $b = 4$, without overhead.

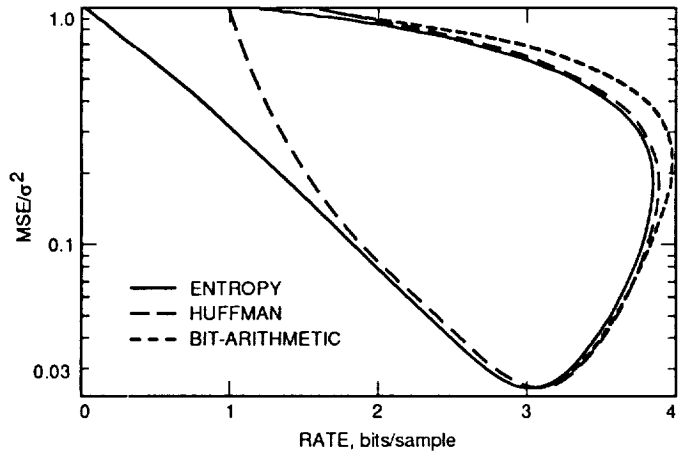


Fig. 8. Bit-wise arithmetic coding, Huffman coding, and entropy for Laplacian source, $b = 4$, without overhead.

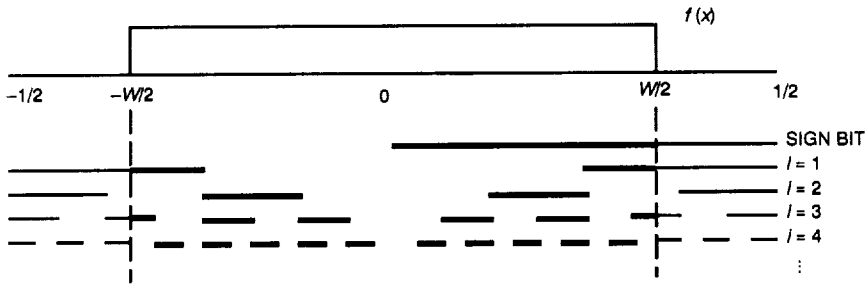


Fig. 9. Uniform distribution and codeword assignment in the limit $b \to \infty$.

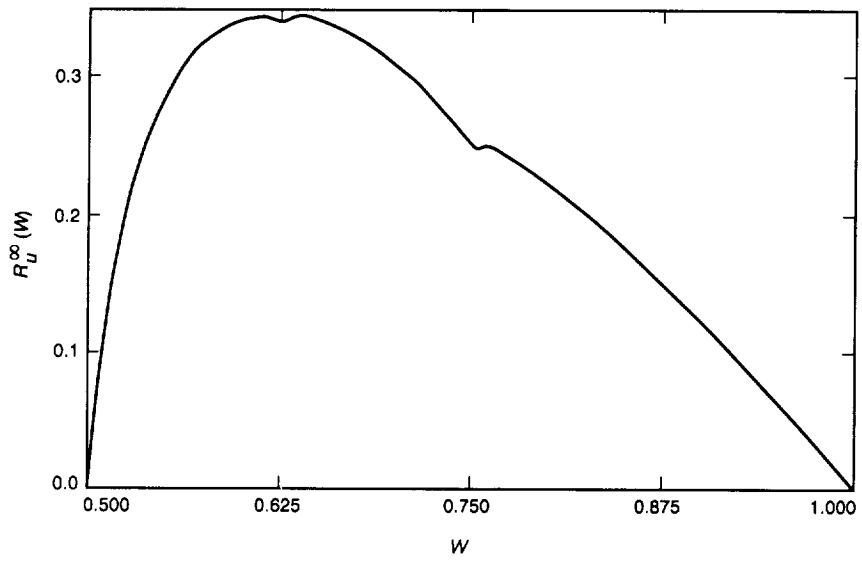


Fig. 10. The redundancy function for a uniform distribution, $R_u^\infty(W)$.

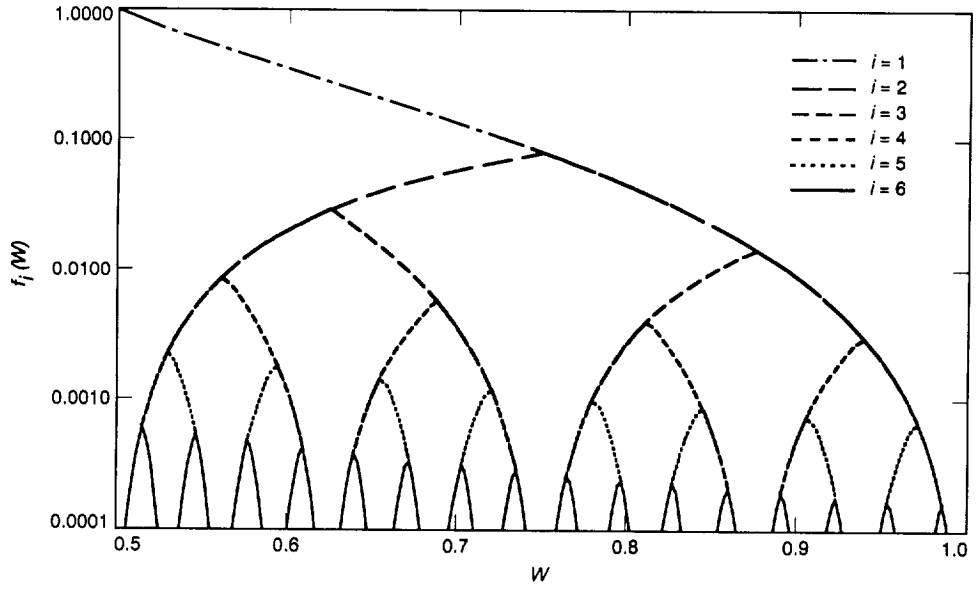


Fig. 11. Components $f_i(W)$ of the redundancy function for a uniform distribution.

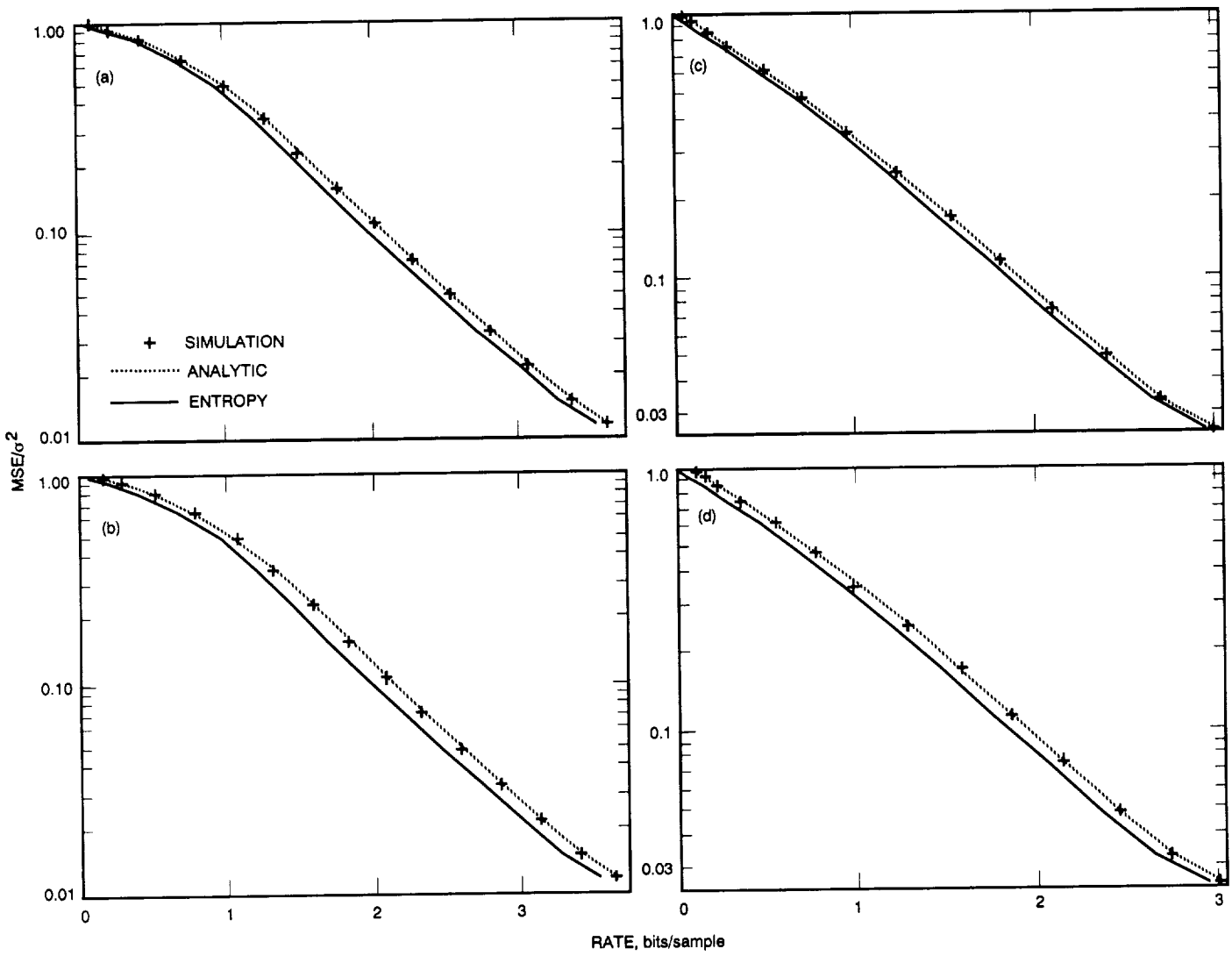


Fig. 12. Bit-wise arithmetic coding performance, including overhead: (a) Gaussian source, $b = 4$, $N = 512$; (b) Gaussian source, $b = 4$, $N = 256$; (c) Laplacian source, $b = 4$, $N = 512$; and (d) Laplacian source, $b = 4$, $N = 256$.

TDA Progress Report 42-117

May 15, 1994

Spacecraft–Spacecraft Radio-Metric Tracking: Signal Acquisition Requirements and Application to Mars Approach Navigation

R. D. Kahn
Tracking Systems and Applications Section

S. Thurman
Navigation Systems Section

C. Edwards
TDA Technology Development

Doppler and ranging measurements between spacecraft can be obtained only when the ratio of the total received signal power to noise power density (P_t/N_0) at the receiving spacecraft is sufficiently large that reliable signal detection can be achieved within a reasonable time period. In this article, the requirement on P_t/N_0 for reliable carrier signal detection is calculated as a function of various system parameters, including characteristics of the spacecraft computing hardware and a priori uncertainty in spacecraft–spacecraft relative velocity and acceleration. Also calculated is the P_t/N_0 requirement for reliable detection of a ranging signal, consisting of a carrier with pseudonoise (PN) phase modulation. Once the P_t/N_0 requirement is determined, then for a given set of assumed spacecraft telecommunication characteristics (transmitted signal power, antenna gains, and receiver noise temperatures) it is possible to calculate the maximum range at which a carrier signal or ranging signal may be acquired. For example, if a Mars lander and a spacecraft approaching Mars are each equipped with 1-m-diam antennas, the transmitted power is 5 W, and the receiver noise temperatures are 350 K, then S-band carrier signal acquisition can be achieved at ranges exceeding 10 million km. An error covariance analysis illustrates the utility of in situ Doppler and ranging measurements for Mars approach navigation. Covariance analysis results indicate that navigation accuracies of a few km can be achieved with either data type. The analysis also illustrates dependency of the achievable accuracy on the approach trajectory velocity.

I. Introduction

In recent years, there has been increased interest in pursuing an intensive program of lunar and Mars exploration. Conducting an expansive program of lunar and Mars exploration with multiple spacecraft will present new technical and operational challenges. For example, future missions to Mars may rely on aerobraking to achieve orbit insertion. The aerobraking technique reduces a spacecraft's propellent requirement, but imposes more stringent navigation constraints than have been needed for previous interplanetary missions. Simultaneous deployment of a large number of spacecraft at the Moon and/or Mars could place unprecedented demands on already taxed ground-based tracking resources. The need for greater navigational accuracy and the need to relieve the burden on ground-based antennas could both be addressed by making use of in situ radio-metric tracking measurements. Doppler and ranging measurements between spacecraft provide navigation information complementary to Earth-based techniques and could potentially decrease reliance on ground-based tracking.

This article investigates requirements for successful acquisition of in situ radio-metric data types and discusses potential measurement precision. An error covariance analysis has also been performed to investigate the navigation accuracy that can be achieved using Doppler and ranging measurements between a spacecraft approaching Mars and a Mars-orbiting satellite used as a navigation aid. The analysis was designed to build upon the results of similar studies conducted previously [1-4]. The error covariance analysis points out some of the most important trajectory and tracking system parameters that influence the navigation performance that can be obtained. While the mission scenario studied herein involves Mars, in situ radio tracking is also applicable to lunar spacecraft navigation, or to any other mission in which a radio beacon or another spacecraft is available near the target body and can be used as a navigation aid.

II. Carrier Signal Acquisition

To perform in situ Doppler measurements, one spacecraft tracks the phase of a carrier signal transmitted from a second spacecraft. If the measurement is one-way, the signal originates at the second spacecraft; in a two-way measurement, the signal is uplinked by the first spacecraft, coherently transponded by the second spacecraft, and then phase tracked by the first spacecraft. The Doppler shift on the received signal provides a direct measure of the relative line-of-sight velocity of the two spacecraft. While one-way measurements are more easily implemented, they are inferior to two-way measurements because an unknown frequency offset between the two spacecraft oscillators can cause a significant error in the inferred relative velocity of the spacecraft.

A. Detection Strategy

A simple signal detection strategy is depicted in Fig. 1. The incoming signal is mixed with a locally generated model; the resulting baseband signal is integrated for T sec; and N consecutive T-sec integrations are Fourier transformed. The resulting Fourier coefficients reflect signal or noise amplitude in N bins of width $1/(NT)$ Hz, centered at the model frequency (Fig. 2). In general, the a priori uncertainty of the signal frequency, $\Delta\nu$, will be larger than the frequency range, $1/T$ Hz, that can be searched with a single Fourier transform, so a number of different model frequencies will have to be tried before the signal is detected.

To determine whether the signal is present in a given frequency span (the range of frequencies probed by a single Fourier transform), an amplitude detection threshold is selected. It is desirable to have the threshold be sufficiently low so that the signal will almost always be detected when it is present, and sufficiently high so that noise will not often be mistaken for signal. If the amplitude in some Fourier resolution bin exceeds the threshold, then the same frequency model is used to process the next NT sec of data. If a signal is detected in the same bin in two consecutive Fourier transforms, then it is assumed that the true signal has been found, and phase-lock-loop tracking commences.

Note that in any given Fourier transform, it is possible that the noise amplitude in at least one Fourier resolution bin will exceed the amplitude detection threshold. If one requires 99-percent confidence of detecting the true signal twice consecutively when it is present, and 1-percent probability of mistaking noise for signal, one obtains a relation between N and $(P_t/N_0)T$ that is approximated by: $(P_t/N_0)T = 29N^{-0.92}$. The Appendix discusses how this equation is obtained.

If the signal is equally likely to lie anywhere within the search space, $\Delta\nu$, then the mean time to signal acquisition is $\overline{T}_{ACQ} \approx \Delta\nu NT^2/2$ [5]. The acquisition time is proportional to NT because that is the time required to accumulate data for each Fourier transform, and is proportional to $T\Delta\nu$ because the frequency search space contains $T\Delta\nu$ spans of width $(1/T)$ Hz.

Taken together, the equations $(P_t/N_0)T = 29N^{-0.92}$ and $\overline{T}_{ACQ} \approx \Delta\nu NT^2/2$ imply that for a given value of

either N, T, or NT, there is a functional relationship between $\overline{T}_{ACQ}/\Delta\nu$ and the minimum P_t/N_0 required for reliable signal acquisition (99-percent confidence of signal detection and 1-percent probability of false alarm). The contours in Fig. 3 illustrate these relationships. As an example, if $N = 1000$ and $T = 0.01$ sec, it is required that $(P_t/N_0) \geq 7$ dB-Hz. The corresponding value of $\overline{T}_{ACQ}/\Delta\nu$ is about 0.05 sec/Hz; e.g., for a frequency uncertainty of 1 kHz, the expected search duration is 50 sec.

It is important to realize that the quantities N and T are not free parameters. Constraints due to both spacecraft dynamics and hardware (and perhaps software) impose restrictions on the allowable values of N, T, and their product, NT.

1. Constraints on the Selection of N and T. Hardware on board the spacecraft dictates a maximum value of N and a minimum value for T. At present, Fourier transform chips that can operate on hundreds of points are commonplace. If N is constrained by hardware to be at most 1024, then the data dump interval, T, is greater than 10^{-4} sec over most of the range of P_t/N_0 in Fig. 3. A dump rate, $1/T$, of 10,000 Hz is well within the capabilities of existing ground-based hardware and may be feasible for space-qualified systems by the late 1990's.

2. Constraints on Data Collection Interval, NT. Constraints on the product NT result from unmodelled beacon accelerations and from oscillator drift, both of which cause the frequency, f, of the received signal to change in time. Signal detection can be severely impeded if the signal moves through many Fourier resolution bins over the time, NT, during which the data for a single Fourier transform are acquired. If it is required that the change in frequency due to an unmodelled acceleration, δa, over time NT be smaller than $1/(2NT)$ Hz (half the size of one Fourier resolution bin), then the following relation results:

$$NT \leq \sqrt{\frac{c}{2f\delta a}} \tag{1}$$

A change in frequency due to oscillator instability has an effect similar to an unmodelled acceleration. Let $\sigma_{y_i}(NT)$ be the Allan standard deviation of spacecraft i's oscillator. Then imposing a constraint analogous to the unmodelled acceleration constraint above yields

$$NT < \frac{1}{2f\sqrt{\sigma_{y_1}^2(NT) + \sigma_{y_2}^2(NT)}} \tag{2}$$

Contours corresponding to $NT = 10^n$ for $n = -2$ to 2 are plotted in Fig. 3.

3. Range of Frequencies to be Searched, $\Delta\nu$. The range of frequencies to be searched, $\Delta\nu$, is a function of the uncertainty in spacecraft–spacecraft relative velocity. Also, for one-way transmission (and for the uplink of a two-way measurement), $\Delta\nu$ depends on an unknown frequency offset in the transmitting spacecraft's reference oscillator.

If the spacecraft–spacecraft velocity uncertainty is δv, the corresponding uncertainty in carrier frequency is $\Delta\nu = f(\delta v/c)$. The day-to-day drift in carrier frequency due to oscillator instability depends on the quality of the spacecraft oscillator. If the oscillator is not temperature-stabilized, a 2.3-GHz (S-band) carrier frequency may vary by thousands of Hz. A highly stable oscillator has day-to-day variations much smaller than 1 Hz at S-band.

4. Example: Calculation of P_t/N_0 Requirement for Carrier Acquisition. The above information may be combined in order to calculate the P_t/N_0 requirement for reliable carrier acquisition.

As an example, suppose one spacecraft is fixed on the surface of Mars, while a second spacecraft is on approach to Mars. The spacecraft are equipped to transmit and receive a 2.3-GHz signal. It is desired that a carrier signal be acquired within 5 min. Neither spacecraft has a priori knowledge of the spacecraft–spacecraft relative velocity and acceleration.

Since the spacecraft have no a priori knowledge of their relative velocity, the range of frequencies to be searched, $\Delta\nu$, is the full spacecraft–spacecraft relative velocity. A typical Mars approach velocity is 4 km/sec, corresponding to a frequency search range of approximately 31 kHz at S-band. Since 5-min acquisition is desired, it is required that $\overline{T}_{ACQ}/\Delta\nu \lesssim 0.01$ sec/Hz.

The relative acceleration of the lander and approach spacecraft is principally due to Mars rotation, which imparts up to 0.017 m/sec^2 of acceleration to the lander. This unmodelled acceleration imposes the constraint $NT \leq 2$ sec. Equation (2) indicates that the upper bound on NT arising from oscillator stability is also 2 sec if the 2-sec Allan standard deviation of the spacecraft oscillators is smaller than 8×10^{-11} sec/sec.

The parameter constraints are summarized in Table 1. The constraints $\overline{T}_{ACQ}/\Delta\nu \lesssim 0.01$ sec/Hz and $NT \leq 2$ sec are represented by the unshaded region in Fig. 4. The minimum P_t/N_0 satisfying both constraints is approximately 13 dB-Hz.

Once the P_t/N_0 requirement is known, then if the spacecraft telecommunication characteristics are specified, one can determine the maximum range at which the carrier can be acquired. Figure 5 illustrates P_t/N_0 as a function of range between the Mars lander and approach spacecraft, for several different assumptions about the spacecraft telecommunications system. The tapering of the curves at smaller ranges is due to the increasing contribution of Mars to the system noise temperature.

III. Doppler Measurement Precision

Once a carrier signal has been detected, tracking the signal's phase is straightforward. Phase tracking precision depends on a variety of factors, including P_t/N_0 and oscillator stability. However, if the carrier can be tracked at all, then the point-to-point phase precision is much smaller than one RF cycle (e.g., 13 cm at S-band). Phase tracking precision does not significantly affect the accuracy with which the Mars approach spacecraft's state can be determined by fitting to an arc of beacon-spacecraft Doppler data. Determination of the approach spacecraft's state should be limited by other effects, such as unmodelled spacecraft accelerations.

IV. Ranging Signal Acquisition

A. Signal Structure

The ranging measurement is performed as follows: A spacecraft transmits a signal to a second spacecraft that locks onto the signal and coherently transponds it back to the first spacecraft. Two-way ranging offers the tremendous advantage of eliminating the problem of unknown spacecraft clock offsets that can significantly degrade the accuracy of a one-way measurement. However, the accuracy of the two-way measurement is still affected by the first spacecraft's clock instability over the round-trip light time. For example, if the spacecraft oscillator has an Allan variance of 10^{-11} sec/sec over time scales of minutes and the two-way light time is 100 sec (30 million km), then clock instability introduces a 30-cm error.

The ranging signal is assumed to be of the following form:

$$s(t) = A \cos(\omega_0 t + \beta PN(p, f_c, t - t_n) + \phi_0)$$

where

$A =$ the signal amplitude

$\omega_0 =$ the carrier frequency

$\beta =$ the modulation index, $0 < \beta < \pi/2$

$PN =$ a pseudonoise sequence with period p and chip rate f_c

$p =$ the period of the PN sequence

$f_c =$ the chip rate of the PN sequence, i.e., $1/f_c$ is the chip duration, τ_c

$t_n =$ the epoch of the PN sequence

$\phi_0 =$ the phase of the carrier signal at $t = 0$

Pseudonoise (PN) sequences [6] are a special class of sequences of 1's and -1's that can be generated by linear feedback shift registers. Individual elements of such a sequence will be referred to as "symbols." PN sequences are periodic; a given sequence consists of repetitions of a symbol pattern of length p symbols. For every integer n, there exists at least one PN sequence of period $2^n - 1$. If $r(j)$ is a PN sequence with period p, then the autocorrelation of $r(j)$, $R_p(i) = (1/p) \int_1^p r(j)r(j + 1)dj$ is equal to 1 for $i =$ any integer multiple of p and is equal to $-1/p$ for all other i (Fig. 6). The chip rate, f_c, of a PN sequence is the rate at which symbol transitions occur.

During a two-way ranging measurement, the spacecraft originating the PN signal simultaneously receives a PN code that had been transmitted one round-trip light time earlier. The incoming PN code is offset from the current state of the uplink code by an amount equal to the round-trip light time divided by the chip duration, τ_c. This code offset can be determined by cross-correlating the received signal with the locally generated sequence. To ensure that such a two-way ranging measurement is unambiguous, it is required that the period of the PN sequence, $p\tau_c$, be larger than the a priori uncertainty in the spacecraft–spacecraft range. The relationship between the chip rate, PN period, and range ambiguity is plotted in Fig. 7. The ranging precision is proportional to the product of the chip duration and the precision with which the PN sequences can be aligned. If chip alignment precision of 1 percent can be achieved in the tracking loop, then a chip rate of 1 MHz corresponds to a ranging measurement precision of 3 m.

B. Ranging Signal Detection

The incoming signal is

$$s(t) = A \cos(\omega_0 t + \beta PN(p, f_c, t - t_n) + \phi_0)$$

and the locally generated model for the carrier is

$$\tilde{s}(t) = \exp\left[i(\tilde{\omega}_0 t + \beta PN(p, \tilde{f}_c, t - \tilde{t}_n))\right]$$

If the carrier is fully suppressed ($\beta = \pi/2$), then multiplication of the signal and model, followed by low-pass filtering, yields the following:

$$C(t) \equiv s(t)\,(\tilde{s}(t)) = \frac{A}{2} \exp\left[i\left((\tilde{\omega}_0 - \omega_0)t - \phi_0\right)\right]$$

$$\times \left[PN(p, f_c, t - t_n) PN(p, \tilde{f}_c, t - \tilde{t}_n)\right]$$

To detect the signal, it is necessary that the model parameters, $\tilde{\omega}_0, \tilde{f}_c$, and $\tilde{t}_n$, be adequately close to the actual signal parameters, ω_0, f_c, and t_n. However, it is possible to design the spacecraft so that the PN clock and the transmitted carrier frequency are both referenced to the same frequency standard. In this case, a search must be conducted only for ω_0 and t_n; the code frequency, f_c, is a deterministic function of the carrier frequency, ω_0.

The product of actual and model signals is integrated over T-sec intervals, and a set of N consecutive integrations is Fourier transformed. It is important that the time period NT be long enough that the product $PN(p, f_c, t - t_n)\left[PN(p, \tilde{f}_c, t - \tilde{t}_n)\right]$ approximates the PN autocorrelation function. The signal can be detected via the Fourier transform only when (1) the model frequency lies within $1/2T$ Hz of the actual frequency and (2) the difference between the model PN epoch and the true PN epoch is a fraction of a chip.

Since a given Fourier transform probes a region of frequency space of width $1/T$ Hz, centered at the model frequency, a total of $T\Delta\nu$ Fourier transforms are required to search a frequency range of $\Delta\nu$ Hz. Note, however, that a T-sec average of $\{signal \times model\}$ suffers an amplitude loss equal to[1] $\mathrm{sinc}\,(T(\tilde{\omega}_0 - \omega_0/2))$; if the signal and model differ by $1/2T$ Hz, the corresponding loss in amplitude is about 40 percent. It may be advantageous, therefore, to space the frequency models by an amount $1/k_f T (k_f > 1)$ to reduce the distance between the signal and the best model frequency. Reducing the spacing between frequency models increases the number of models needed to cover the search space.

[1] S. A. Stephens, "An Analysis of FFT Tone Acquisition," JPL Interoffice Memorandum 335.1-92-14 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 14, 1992.

Even if the model frequency is perfect, a fully suppressed carrier cannot be detected unless the model PN epoch is also close to truth (within a fraction of a chip). If the model PN epoch is stepped in single-chip increments, then the best PN code alignment achieved may be off by as much as 1/2 chip, resulting in a factor of two amplitude loss (see Fig. 6). In order to decrease the potential loss of amplitude due to fractional PN-code misalignment, one may conduct the search using smaller step sizes, at the cost of requiring additional models to cover the entire search space. If the step size is $1/k_c$ chip, then up to $k_c \times p$ offsets may need to be tested before the signal is detected.

The signal search space is illustrated in Fig. 8 for the case $k_f = 1$. The search space consists of $N_f \times N_s$ cells, where N_f is the number of model frequencies and N_s is the number of trial symbol offsets. Each Fourier transform of $\{signal \times model\}$ probes a range of $(1/T)$ Hz, centered at the model frequency. One method of performing a systematic search is to step through all possible model symbol offsets for a given model frequency. If the signal is not detected, the model frequency is then shifted by $1/T$ Hz. Thus, at each model frequency, N_s N-point Fourier transforms are performed. Assuming the Fourier transforms can be computed in real time, $N_s \times (NT)$ sec are spent searching at each model frequency. Unfortunately, it is possible that this search algorithm would miss the signal entirely if the incoming signal's frequency is changing with time (due to unmodelled spacecraft–spacecraft relative acceleration and oscillator drift).

The problem of a signal with a moving frequency may be addressed by taking the following two steps: (1) set k_f equal to 2, i.e., space the model frequencies by $1/2T$ Hz; each Fourier transform of $\{signal \times model\}$ will still probe a region in frequency space of width $1/T$ Hz, but the region probed by adjacent frequency models will now overlap by 50 percent; and (2) require that the change in signal frequency over a time $N_s \times (NT)$ be smaller than $(1/2T)$ Hz. Given an upper bound on the spacecraft–spacecraft relative acceleration and oscillator drift, this requirement places a constraint on the allowable values of the parameters N_s, N, and T (the impact of this and other constraints is discussed later). The number of Fourier transforms required to search the entire two-dimensional space with this strategy is $2T\Delta\nu N_s$.

The signal detection procedure is the same as the carrier detection method discussed earlier. Model signals are sequentially tested by mixing with the incoming signal and then Fourier transforming N points acquired over NT sec. If a Fourier frequency bin has amplitude exceeding the detection threshold, then the same model is tested again. If

a second detection occurs, it is assumed that the signal has been found and phase- and delay-lock-loop tracking is commenced. The mean time to signal acquisition is $\overline{T}_{ACQ} \approx NT^2\Delta\nu N_s$. The mean time to acquisition could be significantly reduced if the spacecraft hardware is capable of processing multiple model signals in parallel.

If the signal power is known, the amplitude threshold can be selected so that the probability of signal detection when a signal is present, P_{detect}, is high (e.g., 99 percent), while the probability of a false alarm, P_{false}, is low (e.g., 1 percent). For a given probability of detection and probability of false alarm, there is a functional relationship between the quantities $(P_t/N_0)T$ and N. For $P_{detect} = 0.99, P_{false} = 0.01$, and $k_c = 2$ (symbol search proceeding in half-chip steps), the relationship between N and $(P_t/N_0)T$ is well-approximated by the following equation (see Appendix):

$$\frac{P_t}{N_0}T = 43N^{-0.91}$$

The contours in Fig. 9 are derived by combining the equations $\overline{T}_{ACQ} = NT^2\Delta\nu N_s$ and $(P_t/N_0)T = 43N^{-0.91}$. As in the case of pure carrier detection, there are a number of constraints that restrict the allowable values of N, T, and NT.

1. Constraint on NT Arising From Unmodelled Spacecraft Acceleration. The requirement that the change in signal frequency over a time $N_s \times (NT)$ be smaller than $(1/2T)$ Hz is expressed as follows:

$$\frac{d(f_0)}{dt} < \frac{1/2T}{N_s NT} \Rightarrow NT^2 < \frac{c}{2f_0 N_s \delta a}$$

Here f_0 is the carrier frequency, c is the speed of light, and δa is the unmodelled spacecraft acceleration. As in the case of simple carrier detection, it is also required that the time NT be small enough so that unmodelled spacecraft accelerations do not smear the signal over multiple Fourier frequency bins. Together, these two constraints are given by

$$NT < \sqrt{\frac{c}{2f_0\delta a}}\left(MIN\left(1, \sqrt{\frac{N}{N_s}}\right)\right) \qquad (3)$$

2. Constraint on NT Arising From Unmodelled Spacecraft Velocity. It is required that the unmodelled spacecraft–spacecraft relative velocity be small enough so that the range does not change by more than 1/2 chip ($\tau_c/2$ sec) over the NT sec during which data are collected for a Fourier transform.

$$NT < \frac{c\tau_c}{2\,\delta v} \qquad (4)$$

where δv is the unmodelled spacecraft–spacecraft relative velocity.

3. Oscillator Stability Requirements. The constraints on oscillator stability are analogous to the constraints on NT arising from unmodelled spacecraft acceleration and velocity.

In analogy with the acceleration constraint, it is required that the signal change by less than $(1/2T)$ Hz over $N_s \times (NT)$ sec, and also that the carrier frequency not move through many Fourier frequency bins over the NT sec during which data are acquired for the Fourier transform (Δ *frequency* $< 1/(2NT)$ over NT sec). Letting $\sigma_{y_i}(\tau)$ denote the τ-sec Allan standard deviation of spacecraft i's oscillator,

$$\sqrt{\sigma_{y_1}^2(N_sNT) + \sigma_{y_2}^2(N_sNT)} < \frac{1}{2\,Tf_0}$$
$$\sqrt{\sigma_{y_1}^2(NT) + \sigma_{y_2}^2(NT)} < \frac{1}{2\,f_0 NT} \qquad (5)$$

In analogy with the velocity constraint, it is required that oscillator drift not introduce a delay of more than $\tau_c/2$ sec over the NT sec during which data are acquired for a Fourier transform. This constraint is looser than that imposed by the above equations for chip rates smaller than the carrier frequency (the usual case).

4. Example: Calculation of P_t/N_0 Required for Ranging Signal Acquisition. Suppose that ranging measurements are to be performed between a spacecraft approaching Mars and a beacon spacecraft on the planet's surface. The carrier frequency is S-band, and the chipping rate is 1 MHz. The length of the PN code is 2047 symbols, corresponding to a two-way range ambiguity of 600 km. The ambiguity can be resolved via spacecraft position determinations resulting from routine Earth-based Doppler tracking of the two spacecraft.

The spacecraft are provided with a crude estimate of their relative velocity: 0.5 km/sec, equivalent to 4000 Hz at S-band. The spacecraft–spacecraft relative acceleration is not modelled. Until the approach spacecraft is

within the sphere of influence of Mars ($\lesssim$1 million km), the total relative acceleration of the spacecraft and beacon is 0.017 m/sec^2 due to Mars rotation. Finally, it is assumed that the spacecraft hardware is capable of performing 1024-point Fourier transforms.

The parameter constraints corresponding to a signal acquisition time of 5 min are summarized in Table 2. All four constraints are satisfied in the unshaded region of Fig. 10 bounded by the contours $N = 1024$ and $\overline{T}_{ACQ}/2N_s\Delta\nu = 10^{-5}$ sec/Hz. The minimum P_t/N_0 satisfying the constraints is approximately 28 dB-Hz. Equation (5) may now be applied to determine the oscillator stability requirements:

$$\sigma_y(409 \text{ sec}) < 2 \times 10^{-6} \text{ sec/sec}$$

$$\sigma_y(0.1 \text{ sec}) < 2 \times 10^{-9} \text{ sec/sec}$$

These Allan variances are easily met by non-temperature-stabilized crystal oscillators.

Figure 5 indicates that if both the lander and approach spacecraft are equipped with 1-m-diam antennas and mediocre receivers ($T_{rec} = 350$ K), a 5-W S-band signal is sufficient to enable a 28-dB-Hz signal power-to-noise density ratio at spacecraft separations of millions of kilometers.

V. Ranging Measurement Precision

To within a factor of order unity, the precision with which PN codes can be aligned with a delay lock loop is $1/\sqrt{2P_t/N_0T_{dwell}}$, where T_{dwell} is the integration time. The resulting ranging measurement precision is

$$\epsilon_\rho \text{ (meters)} \approx \frac{c\tau_c}{\sqrt{2P_t/N_0T_{dwell}}}$$

The exact expression for the ranging measurement precision depends on details of the delay-lock-loop design. In practice, due to dispersion effects in the processing signal path, the best alignment precision that can be achieved is between 0.01 and 0.001 chips.

The utility of ranging is to eliminate the unknown bias inherent in carrier phase (Doppler) measurements. Once this bias has been determined, range changes can be precisely tracked by measuring the carrier phase. The phase of the carrier can be recovered with precision of a fraction of an RF wavelength, even when the transmitted signal has a modulation index of $\pi/2$ (fully suppressed carrier).

VI. Error Covariance Analysis for Mars Approach Navigation

In this section, the navigational utility of in situ radio tracking is illustrated through some examples drawn from a hypothetical future Mars mission, in which a spacecraft approaching Mars acquires Doppler and ranging measurements from a Mars-orbiting communications relay satellite. Error covariance computations for a sequential estimation scheme were performed using approximate models of the approach spacecraft and relay satellite trajectories, and the Doppler and ranging measurement errors. This analysis was not designed to be an exhaustive treatment of the approach navigation problem, but merely to illustrate some of the more significant aspects of the use of in situ Doppler and ranging measurements in this type of application. It should be remembered that although the particular application discussed below refers to Mars, in situ radio tracking is also applicable to missions to the Moon and other target bodies where radio beacons or additional spacecraft may be used as navigation aids.

The error analysis was performed using the model parameters summarized in Table 3. The approach spacecraft was assumed to move on a hyperbolic path with respect to Mars; except for the final hours prior to arrival at periapsis, this assumption leads to a trajectory that is essentially rectilinear and of constant velocity. To illustrate the effect of different approach velocities on navigation performance, two different values of asymptotic approach velocity were considered, representative of the minimum and maximum possible values for low-energy ballistic Earth-to-Mars transfer trajectories. The radius of closest approach chosen for this study (3417 km) yields a periapsis altitude of about 20 km, which is representative of trajectories leading to a direct entry and landing, or the initiation of an aerobraking maneuver for orbit insertion. The Mars-orbiting relay satellite was assumed to be in a circular orbit with a 12-hr period. The spacecraft acceleration-process noise level used in this analysis is intended to account for the small, nongravitational forces (e.g., solar radiation pressure mismodelling, gas leaks from valves, and pressurized tanks) that act on an interplanetary spacecraft.

The navigational utility of both one-way integrated Doppler measurements and two-way ranging measurements has been investigated. Integrated Doppler data are effectively a measurement of the accumulative change in the spacecraft-to-spacecraft range over the contact period. As shown by the parameters given in Table 3, integrated Doppler provides a highly precise measurement of range, but with a large constant bias. The frequency drift uncertainty assumed for the Doppler measurements is represen-

tative of oscillator stability on the order of 10^{-12} sec/sec. Two-way ranging measurements are much less precise than Doppler measurements, but provide an accurate measure of the spacecraft–spacecraft separation (Table 3). As described earlier, acquisition of a ranging signal requires a more extensive search than simple carrier signal acquisition. For fixed assumptions about the spacecraft onboard computing resources and the desired mean time to signal acquisition, the requirement on P_t/N_0 is higher to support ranging acquisition. As a result, the acquisition range for ranging data is considerably smaller than for Doppler data.

The results of the error covariance analysis are summarized in Figs. 11 and 12. These figures show the uncertainties in the approach spacecraft encounter coordinates as a function of time prior to arrival, expressed in an aiming plane (B-plane) coordinate system.[2] Figure 11 shows the semimajor axis of the aiming plane dispersion ellipse and the linearized time-of-flight uncertainty (position uncertainty in the S unit direction divided by τ_∞) for an approach velocity of 3 km/sec, while Fig. 12 shows the same two quantities for an approach velocity of 6 km/sec. (Mars Observer's approach velocity was about 2.4 km/sec; MESUR Pathfinder's approach velocity will be about 5.5 km/sec.) Note that, in practice, delay-lock-loop tracking of a PN-modulated carrier signal enables recovery of the carrier phase, thus enabling simultaneous acquisition of ranging and integrated Doppler measurements. However, Figs. 11 and 12 show the navigational uncertainties resulting from the use of either data type independently, in order to illustrate differences in the abilities of the two data types to determine various components of the approach trajectory.

Both Figs. 11 and 12 illustrate a clear difference between Doppler and ranging performance. Since ranging data provide a direct measurement of the spacecraft–spacecraft separation, the time of flight (which in this case is essentially the spacecraft range) can be very accurately determined immediately upon acquisition. In contrast, Doppler data can only measure this component of the flight path indirectly and need to be collected for up to several days before sufficient information is obtained. This occurs because Doppler data depend upon the relative acceleration of the approach spacecraft and the relay satellite (which is inversely dependent upon their relative range) to determine the time of flight. Additional experimentation found that the approach-spacecraft acceleration process noise was the principal factor limiting the integrated Doppler's ability to determine the time of flight.

Figure 12 shows that when the approach velocity is relatively high, the smaller acquisition distance associated with two-way ranging results in the aiming plane dispersion ellipse of the approach spacecraft remaining relatively large until about 15 hr prior to arrival. If a final maneuver is needed to correct the spacecraft's aim point, this implies that the maneuver may have to be performed very near the arrival point, with the ensuing risk that there may not be sufficient time remaining to redetermine the trajectory and correct any maneuver execution errors that might have occurred. Thus, if large approach velocities must be accommodated, early acquisition of Doppler followed by joint acquisition of Doppler and ranging data would be desirable.

VII. Summary

The P_t/N_0 required for acquisition of carrier or ranging signals between two spacecraft is a function of a number of parameters, including a priori uncertainty in the spacecraft–spacecraft relative state, and the capabilities of the spacecraft processing hardware and software (e.g., the maximum number of points that can be Fourier transformed in real time). Once these parameters are specified for a given set of assumptions about the spacecraft telecommunications systems, one may calculate the maximum range at which a signal can be acquired. As an example, it has been shown that S-band radio-metric measurements between a Mars lander and Mars approach spacecraft equipped with 1-m-diam antennas are feasible over distances of millions of kilometers. The analysis presented here can easily be applied to a wide variety of other mission scenarios.

To illustrate the utility of in situ radio-metric measurements between spacecraft, a covariance analysis was performed for the case of Doppler and ranging measurements between a Mars orbiter and a spacecraft on Mars approach. The results obtained are consistent with previous studies, and suggest that approach navigation accuracies of a few kilometers at Mars may be obtained with either in situ Doppler or ranging measurements.

In the near-term, in situ ranging and Doppler data between spacecraft would likely be relayed to Earth for navigation processing. Ultimately, onboard computing resources could be used to perform the navigation updates, enabling a near-real-time in situ navigation capability.

[2] The aiming plane, or B-plane, coordinate system is defined by three unit vectors, $\boldsymbol{S}$, $\boldsymbol{T}$, and $\boldsymbol{R}$. $\boldsymbol{S}$ is parallel to the incoming asymptote of the approach hyperbola; $\boldsymbol{T}$ is parallel to the Martian equatorial plane; and $\boldsymbol{R}$ completes an orthogonal triad with $\boldsymbol{S}$ and $\boldsymbol{T}$. The aim point for a planetary encounter is defined by the miss vector, $\boldsymbol{B}$, which lies in the $\boldsymbol{T}$–$\boldsymbol{R}$ plane, and specifies where the point of closest approach would be if the target planet had no mass and did not deflect the spacecraft's flight path.

References

[1] A. S. Konopliv and L. J. Wood, "High-Accuracy Mars Approach Navigation with Radio-Metric and Optical Data," paper AIAA-90-2907, presented at the AIAA/AAS Astrodynamics Conference, Portland, Oregon, August 20–22, 1990.

[2] S. W. Shepperd, D. P. Fuhry, and T. J. Brand, "Onboard Preaerocapture Navigation Performance at Mars," paper AAS 91-119, presented at the AAS/AIAA Spaceflight Mechanics Meeting, Houston, Texas, February 11–13, 1991.

[3] S. W. Thurman and J. A. Estefan, "Mars Approach Navigation Using Doppler and Range Measurements to Surface Beacons and Orbiting Spacecraft," paper AAS 91-118, presented at the AAS/AIAA Spaceflight Mechanics Meeting, Houston, Texas, February 11–13, 1991.

[4] S. W. Thurman and S. E. Matousek, "Trajectory and Navigation System Design for Robotic and Piloted Missions to Mars," *The Telecommunications and Data Acquisition Progress Report 42-107, vol. July–September 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 113–131, November 15, 1991.

[5] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications*, vol. III, Rockville, Maryland: Computer Science Press, 1989.

[6] R. Dixon, *Spread Spectrum Systems*, New York: John Wiley and Sons, 1984.

Table 1. Parameter constraints for 5-min carrier signal acquisition between a Mars lander and a spacecraft on Mars approach.

Constraint	Reason
$\overline{T}_{ACQ}/\Delta\nu \lesssim 10^{-2}$ sec/Hz	5-min acquisition, $\Delta\nu = 31$ kHz
$NT < 2$ sec	Eq. (1) (unmodelled acceleration)

Table 2. Parameter constraints for 5-min ranging signal acquisition between a Mars lander and a spacecraft on Mars approach.

Constraint	Reason
$\overline{T}_{ACQ}/2N_s\Delta\nu \lesssim 10^{-5}$ sec/Hz	5-min acquisition, $\Delta\nu = 4000$ Hz, and $N_s = 4094$
$N < 1024$	Capability of Fourier transform chip on spacecraft
$NT < 0.03\sqrt{N}$ sec	Eq. (3) (unmodelled acceleration)
$NT < 0.3$ sec	Eq. (4) (unmodelled velocity)

Table 3. Mars approach navigation error analysis parameters.

Trajectory parameter	Value
Asymptotic approach velocity, $v\infty$	3 or 6 km/sec
Radius of closest approach	3417 km
Inclination of approach trajectory to Martian equator	+15 deg
Mars relay satellite orbital period	2 hr
Mars relay satellite inclination	0 deg
Spacecraft acceleration process noise (approach spacecraft and relay satellite)	1.4×10^{-5} m/sec/$\sqrt{\text{hr}}$
A priori state uncertainty (1σ)	**Value**
Approach spacecraft epoch position	100 km
Approach spacecraft epoch velocity	1 m/sec
Relay satellite epoch position	3 km
Relay satellite epoch velocity	0.1 m/sec
Doppler measurement parameter	**Value**
Maximum acquisition range	2×10^6 km
Additive random measurement noise	4 mm (1σ)
Accumulative random measurement noise	0.58 m^2/hr
Frequency drift stability	0.3 mm/sec (1σ, a priori)
Carrier phase bias uncertainty	100 km (1σ, a priori)
Ranging measurement parameter	**Value**
Maximum acquisition range	8×10^5 km
Additive random measurement noise	5 m (1σ)
Range bias uncertainty	15 m (1σ)

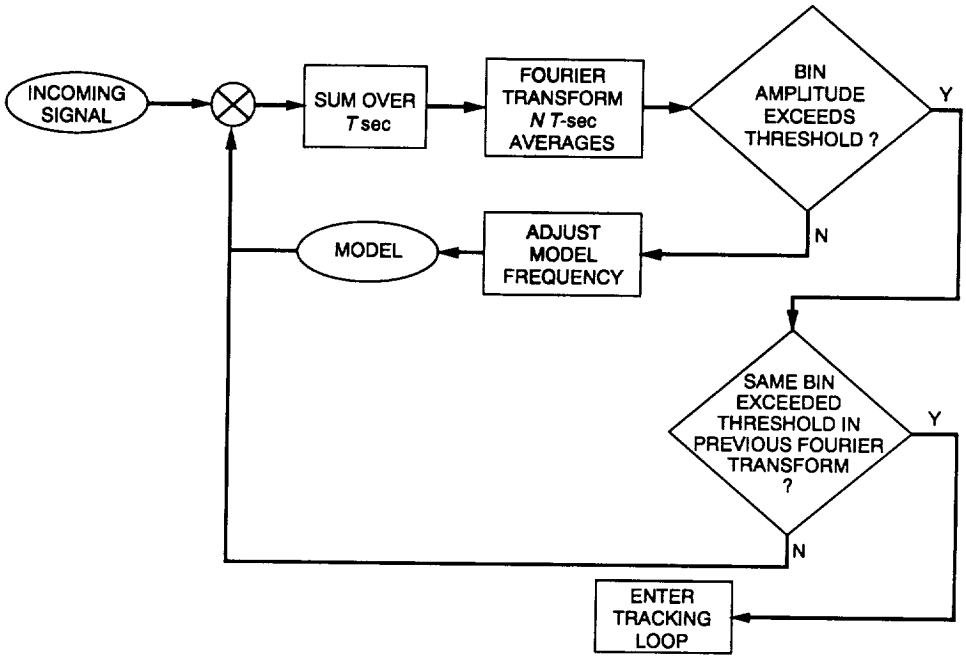


Fig. 1. A simple carrier signal detection strategy.

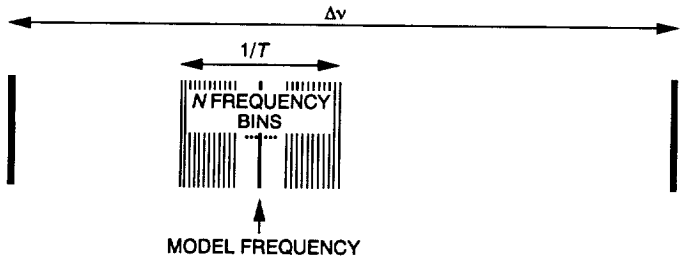


Fig. 2. The relationship between frequency search range, $\Delta\nu$ Hz, and frequency span, $1/T$ Hz, probed by a single Fourier transform.

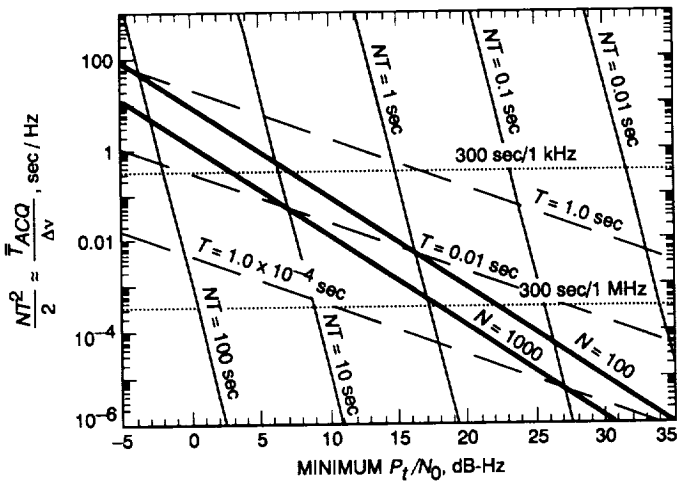


Fig. 3. Relationships between $T_{ACQ}/\Delta\nu$, N, T, and the minimum P_t/N_0 required for carrier signal acquisition with P_{false} = 0.01 while P_{detect} = 0.99.

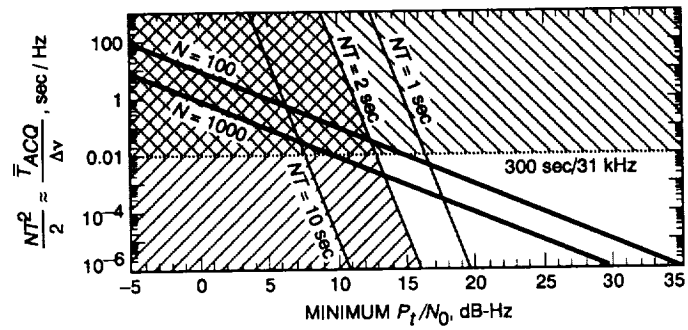


Fig. 4. Given the constraints $NT \leq 2$ sec and $T_{ACQ}/\Delta\nu \leq 0.01$ sec/Hz, the minimum P_t/N_0 needed for reliable carrier detection is 13 dB-Hz.

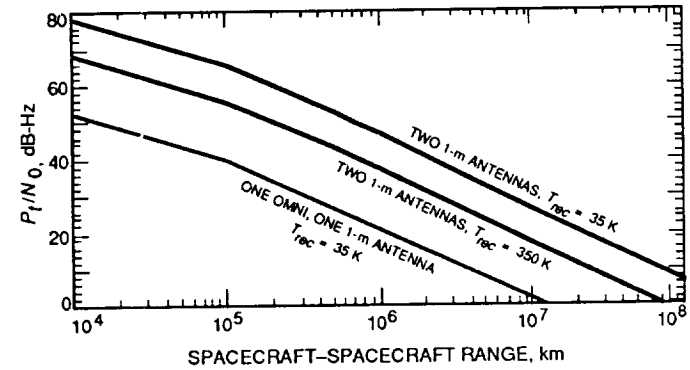


Fig. 5. P_t/N_0 as a function of Mars lander–approach spacecraft range. Transmitted power is 5 W. T_{rec} is the receiver noise temperature.

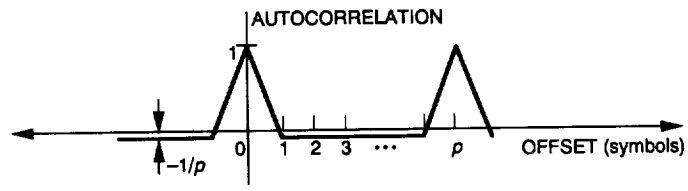


Fig. 6. Autocorrelation of a PN sequence of period p.

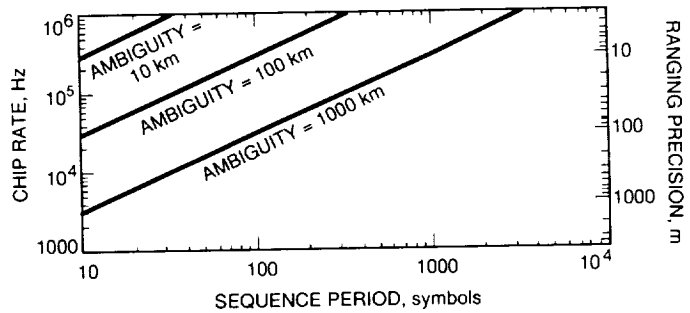


Fig. 7. Relationship between chip rate, length of PN sequence, range ambiguity, and ranging measurement precision (assumed tracking error is 1 percent of chip length).

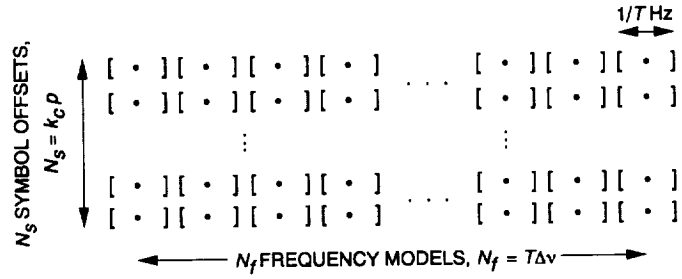


Fig. 8. Search space for carrier signal with PN phase modulation. Each bracket pair represents the range of frequencies probed by a Fourier transform of N T-sec *signal* × *model* points.

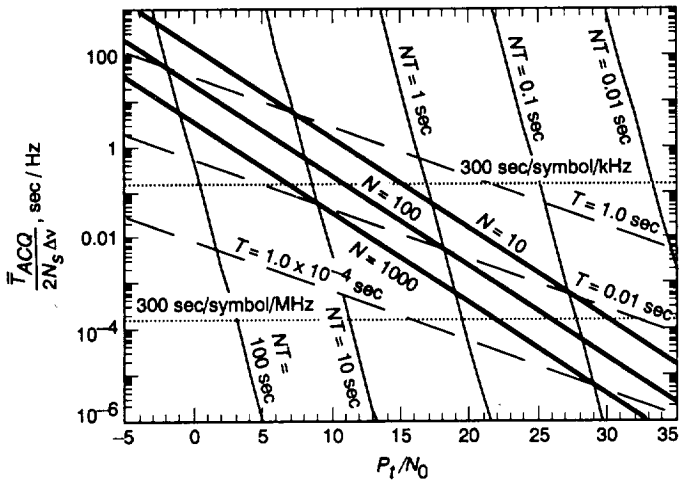


Fig. 9. Relationships between $T_{ACQ}/(2N_s\Delta\nu)$, N, T, and the P_t/N_0 required for ranging signal acquisition with P_{false} = 0.01 while P_{detect} = 0.99.

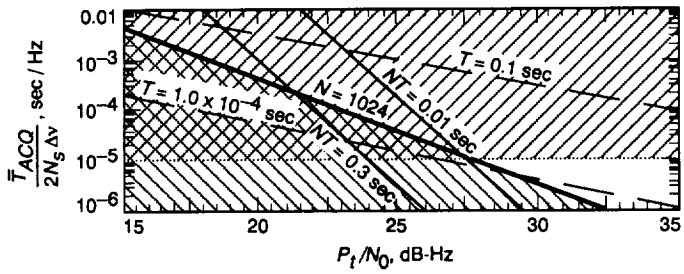


Fig. 10. Given the constraints $T_{ACQ}/(2N_s\Delta\nu) < 10^{-5}$ and N < 1024, the minimum P_t/N_0 enabling reliable ranging signal detection is 29 db–Hz.

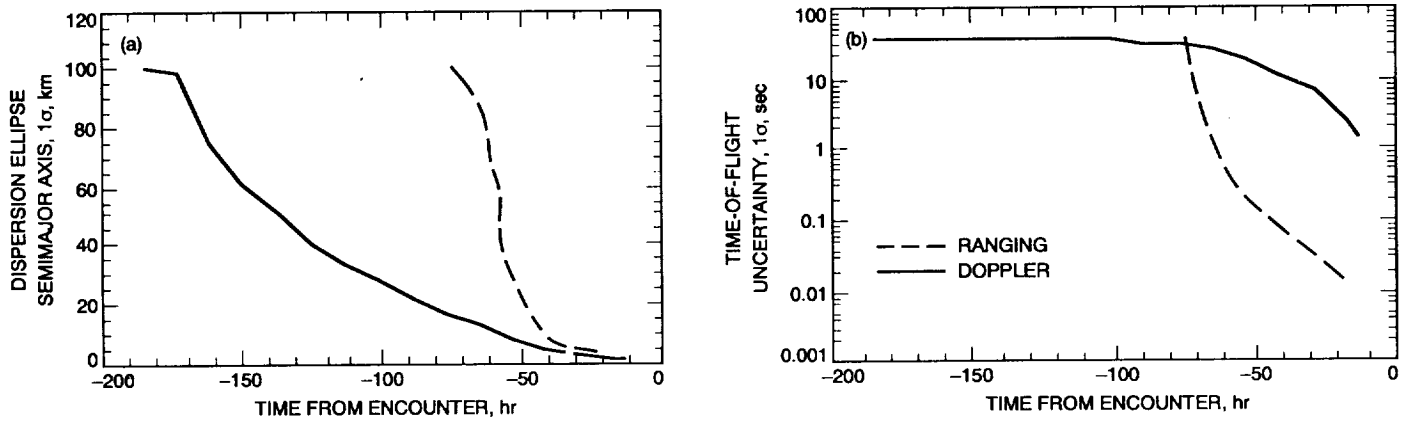


Fig. 11. Mars approach navigation dispersions (v_∞ = 3 km/sec): (a) aim-point uncertainty and (b) linearized time-of-flight uncertainty.

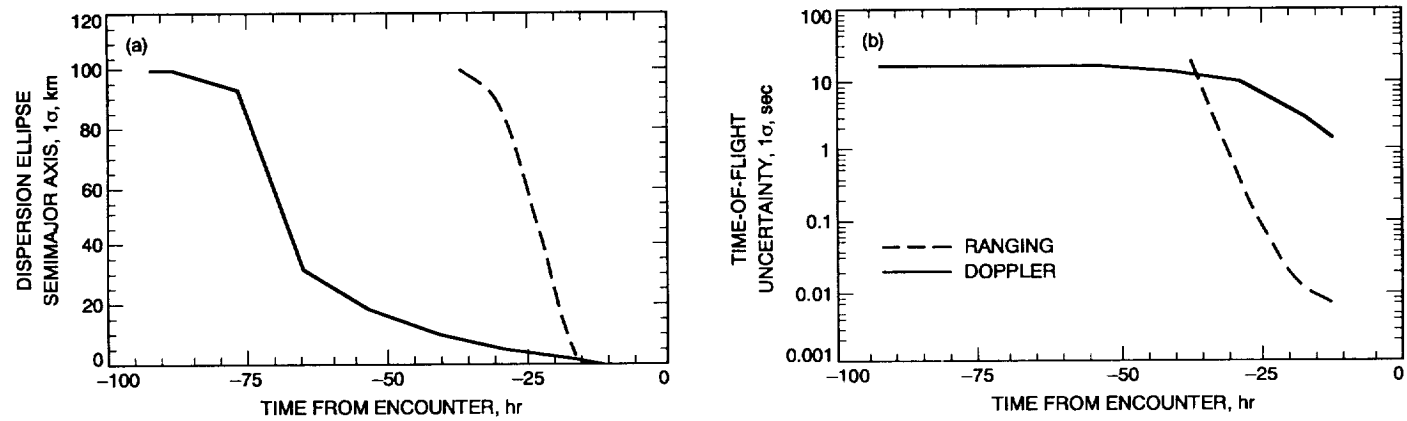


Fig. 12. Mars approach navigation dispersions (v_∞ = 6 km/sec): (a) aim-point uncertainty and (b) linearized time-of-flight uncertainty.

C-3

Appendix

Calculation of Relations between N and $(P_t/N_0)T$

As described in the article, the carrier detection strategy involves looking for Fourier resolution bins whose amplitude exceeds a threshold level. The detection threshold parameter, ξ, is selected such that the probability that the amplitude of the signal plus noise exceeds $\xi \times \{expected\ signal\ amplitude\}$ is 0.995. The probability of two consecutive signal detections when the signal is present, which shall be referred to as P_{detect}, is then 0.99. Calculation of the detection threshold must take into account signal amplitude losses resulting from two effects: (1) If the signal lies within the $1/T$ Hz range probed by the Fourier transform, it may differ from the model by up to $1/T$ Hz. Thus, each T-sec average of the *signal* × *model* phasor suffers amplitude loss because the phasor is not completely "stopped." (2) The signal will in general lie somewhere between Fourier resolution bins, so that the amplitude of the Fourier bin closest to the signal can be smaller than the true signal strength. The first effect can be ameliorated by discarding the Fourier resolution bins that lie toward the edges of the $1/T$ Hz frequency span, at the cost of increasing the number of Fourier transforms required to cover the entire search space of $\Delta\nu$ Hz. The fraction of Fourier resolution bins retained is referred to as χ. The second effect can be reduced by "oversampling," i.e., padding the data points with additional zeros, to decrease the frequency bin size of the Fourier transform. The degree of oversampling is described by a quantity, κ, which is equal to $N/(N + the\ number\ of\ padded\ zeros)$. Note that decreasing κ allows finer sampling of the Fourier transform at the cost of requiring increased computation (i.e., Fourier transforming a larger number of data points).

A reasonable selection of χ and κ, from the standpoint of minimizing acquisition time is $\chi = 1$ and $\kappa = 1/2$. For this choice of χ and κ, it can be shown that the detection threshold parameter, ξ, is given by the following formula:[1]

$$0.995 = 2 \int_{-1/2}^{1/2} du \int_{-1/4}^{1/4} dw (SNR_v)^2 e^{-0.5(\alpha SNR_v)^2} \times \int_{\xi}^{\infty} dr\ r\ e^{-0.5r^2 SNR_v^2} I_0(\alpha r SNR_v^2)$$

where $\alpha = (\sin(\pi u)/\pi u)(\sin(\pi w)/\pi w)$, I_0 is the 0th order modified Bessel function, and SNR_v is the voltage signal-to-noise ratio over NT sec, $\sqrt{2P_t/N_0 NT}$.

The probability that the noise amplitude in a given frequency bin exceeds ξSNR_v is $\exp(-0.5(\xi SNR_v)^2)$. Thus, the probability, P_{false}, that the noise amplitude exceeds ξSNR_v in at least one of N frequency bins is

$$P_{false} = 1 - (1 - \exp(-0.5(\xi SNR_v)^2))^N$$

The first equation allows numerical calculation of ξ as a function of SNR_v. Now if it is required that $P_{false} = 0.01$, then the second equation allows calculation of N as a function of SNR_v. But since $SNR_v = \sqrt{2P_t/N_0 NT}$, N may then be calculated as a function of $(P_t/N_0)T$. Over the range $6 < SNR_v < 12$, $(P_t/N_0)T \approx 29N^{-0.92}$.

In the case of detection of a PN-modulated carrier, the calculation is analogous, but SNR_v must be scaled by $(1 - 1/(2k_c))$, where $1/k_c$ chips is the model symbol offset step size. This factor accounts for the SNR_v loss during detection, due to the fact that the best alignment of the model and signal sequences may be off by as much as $1/2k_c$ chips. If the search is conducted in half-chip increments ($k_c = 2$) and $P_{detect} = 0.99$ while $P_{false} = 0.01$, then the relationship between $(P_t/N_0)T$ and N is $(P_t/N_0)T \approx 43N^{-0.91}$

[1] S. A. Stephens, "An Analysis of FFT Tone Acquisition," JPL Interoffice Memorandum 335.1-92-14 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 14, 1992

N94-35558

TDA Progress Report 42-117 May 15, 1994

Performance Evaluation of Digital Phase-Locked Loops for Advanced Deep Space Transponders

T. M. Nguyen and S. M. Hinedi
Communications Systems Research Section

H.-G. Yeh and C. Kyriacou
Spacecraft Telecommunications Equipment Section

The performances of the digital phase-locked loops (DPLL's) for the advanced deep-space transponders (ADT's) are investigated. DPLL's considered in this article are derived from the analog phase-locked loop, which is currently employed by the NASA standard deep space transponder, using S-domain to Z-domain mapping techniques. Three mappings are used to develop digital approximations of the standard deep space analog phase-locked loop, namely the bilinear transformation (BT), impulse invariant transformation (IIT), and step invariant transformation (SIT) techniques. The performance in terms of the closed loop phase and magnitude responses, carrier tracking jitter, and response of the loop to the phase offset (the difference between the incoming phase and reference phase) is evaluated for each digital approximation. Theoretical results of the carrier tracking jitter for command-on and command-off cases are then validated by computer simulation. Both theoretical and computer simulation results show that at high sampling frequency, the DPLL's approximated by all three transformations have the same tracking jitter. However, at low sampling frequency, the digital approximation using BT outperforms the others. The minimum sampling frequency for adequate tracking performance is determined for each digital approximation of the analog loop. In addition, computer simulation shows that the DPLL developed by BT provides faster response to the phase offset than IIT and SIT.

I. Introduction

In recent years, the topic of the digital phase-locked loop (DPLL) has been studied in great detail and well documented in the literature [1–11]. An excellent survey of the work accomplished during 1960–1980 is provided in [1]. The analysis, design, and performance of the DPLL are dealt with in [4–6]. Optimum DPLL and digital approximation of the analog loop filter are discussed in [7,8]. Currently, most of the work on the DPLL concentrates in these areas, and very little of it focuses on the optimum digital approximation of the analog phase-locked loop (APLL) [8,9]. Aguirre et al. deal only with the design of an optimum loop filter using the impulse invariant transform (IIT) method, minimization method, estimation-prediction technique, and classical control the-

ory approach. The digital loop filters derived by these methods were compared in [8] in terms of stability, gain margin, steady state, and transient performance. On the other hand, [9] focused on the design of the DPLL based on the APLL. Boman [9] considered four different transformations, namely bilinear transformation (BT), IIT, step invariant transformation (SIT), and rotational transformation (RT). The output phase responses of the approximated digital loops using these transformations were evaluated at low sampling rates in the absence of noise and compared. It was found in [9] that for a simple second-order APLL, the phase response of the digital approximation of the APLL using the IIT method exhibits less overshoot and ringing than the others.

The present work is an extension of [8,9,14] to include many other aspects in determining an optimum transformation technique to develop a good digital approximation of a given APLL. The digital approximation is developed by mapping the continuous time S-domain to the discrete time Z-domain. The mapping is accomplished using BT, IIT, SIT, and RT. Because the RT technique is identical to the IIT technique, only three techniques, namely BT, IIT, and SIT, are considered in this article. For each of these mapping techniques, the phase and magnitude responses of the closed-loop transfer function, the response of the loop to the phase offset, the minimum sampling frequency for adequate tracking performance, and the carrier tracking jitter will be evaluated.

The article is divided into five remaining sections. Section II introduces the current command signal format that is received by the deep space transponder along with a simplified model of the APLL for tracking the carrier. Equivalent DPLL's are also described in this section. Detailed recursive implementations of the DPLL's using BT, IIT and SIT are described in Section III. Included in Section III are the plots of the phase and magnitude responses of the closed-loop transfer functions for each digital approximation. Section IV derives the carrier tracking phase jitter for both analog and digital loops with command-on and command-off. Section V presents the computer simulation results to verify the theoretical results obtained in Section IV and to determine the transient response of the digital loops to the initial phase offset. Furthermore, computer simulation results for determining the minimum sampling frequency for each approximation are also presented in Section V. Section VI presents the key conclusions of the article.

II. System Modeling

The mathematical model for the command signal to the spacecraft transponder, $S(t)$, is defined as

$$S(t) = \sqrt{2P}\sin\left((\omega_C + \omega_d)t + \Theta(t) + \phi\right) \tag{1}$$

where P denotes the total received power; $\omega_C = 2\pi f_C$ is the angular carrier frequency; ω_d is the Doppler angular frequency offset; $\Theta(t)$ characterizes the phase modulation, and ϕ characterizes the phase offset. The phase modulation employed by the deep space transponder is $\Theta(t) = md(t)\sin(\omega_{SC}t) + m_R R(t)$, where m is the command modulation index; $d(t)$ denotes the command nonreturn-to-zero (NRZ) data; $\omega_{SC} = 2\pi f_{SC}$ is the command angular subcarrier frequency; m_R is the ranging modulation index, and $R(t)$ denotes the ranging signal.

Without loss of generality, we can set $\omega_C = 2\pi F_{IF}$, and $\omega_d = 0$, and $\phi = 0$, and expand Eq. (1) to get

$$\begin{aligned} S(t) &= \sqrt{2P}\left[\cos(\Theta(t))\sin(2\pi F_{IF}t)\right. \\ &\quad \left. + \sin(\Theta(t))\cos(2\pi F_{IF}t)\right] \end{aligned} \tag{2}$$

Ignoring the higher-order-harmonic component, it can be shown that the first term in Eq. (2) represents the carrier component, and the second is the command signal component [16]. Presently, the carrier component is tracked by an APLL. Illustrated in Fig. 1 is a simplified block diagram of the analog carrier tracking loop which is currently employed by the NASA standard deep space transponder. The APLL depicted in Fig. 1 is a type I, second-order loop with the following characteristics:

$$AK = \text{loop gain} = 2.4 \times 10^7 \tag{3}$$

$$B(S) = \frac{1}{(1 + \tau_{RC}S)};\quad \tau_{RC} = 1.6 \times 10^{-5}\text{sec} \tag{4}$$

$$F(S) = \frac{1 + \tau_2 S}{1 + \tau_1 S};\quad \tau_1 = 4707 \text{ sec};\ \tau_2 = 0.0442 \text{ sec} \tag{5}$$

$$V(S) = \frac{1}{(1 + \tau_V S)};\ \tau_V = 1.0 \times 10^{-6} \text{ sec} \tag{6}$$

$$K(S) = \frac{1}{S} \tag{7}$$

Note that $B(S)$ is the typical lowpass filter (LPF); $F(S)$ is the loop filter; $V(S)$ is the roll-off filter of the voltage-controlled oscillator (VCO); and $K(S)$ is the VCO integrator. Let $G(S)$ be the transfer function, excluding the ideal integrator $K(S)$, of the analog loop defined as follows:

$$G(S) = B(S)F(S)V(S) \tag{8}$$

Based on the APLL described in Fig. 1, the equivalent digital counterparts are shown in Figs. 2 and 3. Figure 2 shows the first configuration, the so-called configuration I, for the digital approximation of the analog loop. Configuration I is developed using direct transformation of each functional block in the analog loop—i.e., $B(S)$, $F(S)$, $V(S)$, and $K(S)$—into the Z-domain. In Fig. 2, M stands for the sample rate reduction factor, and NCO stands for numerically controlled oscillator. On the other hand, configuration II, shown in Fig. 3, is developed by transforming $G(S)$ and $K(S)$ into the Z-domain. Notice that the digital approximations of the APLL illustrated in Figs. 2 and 3 have the sum-and-dump circuit to reduce the sample rate by a factor of M before digital filtering. The sample rate is reduced to a rate such that the implementation of the digital filter is feasible using current digital signal processors. In the following section, the recursive implementations of the LPF, VCO roll-off filter, loop filter, VCO integrator, and the transfer function $G(S)$ will be described.

III. Recursive Implementations of $B(S)$, $F(S)$, $V(S)$, $G(S)$, and $K(S)$

To obtain the digital approximation of the analog carrier PLL described in Figs. 2 and 3, each functional block in the analog loop—i.e., $B(S)$, $F(S)$, $V(S)$, and $K(S)$—can be mapped directly into the Z-domain using BT or the composite function $G(S)$, which uses IIT or SIT. As mentioned earlier, this section will deal only with BT/IIT/SIT, and BT and IIT/SIT correspond to configurations I and II, respectively. Notice that when using the BT technique, one does not map the composite function $G(S)$ because of the mathematical complexity associated with this technique. Moreover, when using the IIT and SIT techniques, one does not map each functional block in the analog loop because one wants to preserve the impulse and step responses of the loop, respectively, at the sampling points.

A. Bilinear Transformation Method

Given a proper sampling frequency, this method preserves the phase characteristics in the narrow passband when mapping the analog PLL into the digital domain. The mapping from analog (S-domain) to discrete domain (Z-domain) can be achieved by direct substitution of the following equation into the analog transfer function [12–14].

$$S = \frac{2(Z-1)}{T_S(Z+1)} \tag{9}$$

where T_S denotes the sampling period, and $F_S = 1/T_S$ denotes the sampling frequency. To obtain the digital approximation of the analog filters using bilinear transformation, one substitutes Eq. (9) into Eqs. (4), (5), (6), and (7) to get $B(Z)$, loop filter $F(Z)$, $V(Z)$, and $K(Z)$. The results are

$$B(Z) = \frac{(1+Z^{-1})}{(A_{00}Z^{-1} + A_{11})} \tag{10}$$

$$F(Z) = \frac{(A_0 Z - B_0)}{(A_1 Z - B_1)} \tag{11}$$

$$K(Z) = \frac{T_S(Z+1)}{2(Z-1)} \tag{12}$$

where

$$A_{00} = 1 - C_0; \quad A_{11} = 1 + C_0 \tag{13}$$

$$\left.\begin{aligned} A_0 &= 1 + a_0 \\ A_1 &= 1 + b_0 \\ B_0 &= a_0 - 1 \\ B_1 &= b_0 - 1 > 0 \end{aligned}\right\} \tag{14}$$

and

$$C_0 = \frac{2\tau_{RC}}{T_S}; \quad a_0 = \frac{2\tau_2}{T_S}; \quad b_0 = \frac{2\tau_1}{T_S} \tag{15}$$

Note that the Z-domain representation for $V(S)$ is exactly the same as in Eq. (10) except that C_0, A_{00}, and A_{11} are replaced by, respectively,

$$C_{01} = \frac{2\tau_V}{T_S};\ A_{01} = 1 - C_{01};\ \ A_{10} = 1 + C_{01} \qquad (16)$$

The digital closed-loop transfer function, $H(Z)$, for this case is given by

$$H(Z) = \frac{AK\ [B(Z)F(Z)V(Z)K(Z)]}{1 + AK\ [B(Z)F(Z)V(Z)K(Z)]} \qquad (17)$$

Plots of the analog and digital closed-loop phase and magnitude responses are shown in Figs. 4(a) and 4(b). These figures show that for sampling frequencies below 80 kHz, distortions in phase and magnitude can occur for the digital approximation loop. In addition, the figures show that for sampling frequencies greater than or equal to 80 kHz, the response of the digital loop approaches that of the analog counterpart. Hence, to achieve the same response as the analog loop, the minimum sampling frequency for this case is 80 kHz. Later on, the minimum sampling frequency to achieve acceptable tracking performances will be investigated by computer simulation. Figures 5(a), 5(b), and 5(c) show the recursive implementations of the loop filter $F(Z)$, the integrator $K(Z)$, and the LPF $B(Z)$, respectively.

B. Impulse Invariant Transformation Method

This mapping technique preserves the impulse response at the sampling points. Let $g(t)$ be the impulse response of $G(S)$, i.e., $g(t) = L^{-1}\{G(S)\}$, where $L^{-1}\{.\}$ denotes the inverse Laplace transform of $\{.\}$. Thus, the digital approximation of the analog transfer function $G(S)$ is given by [12-14]

$$G_D(Z) = T_S\ [z\{g(t)|t = nT_s\}] \qquad (18)$$

where $z\{.\}$ is the z-transform of $\{.\}$. Note that the analog transfer function $G(S)$ considered in this article is defined as in Eq. (8). Similarly, one can get the equivalent digital approximation for the integrator $K(S)$. It is found to be

$$K(Z) = \frac{ZT_S}{(Z-1)} \qquad (19)$$

The digital approximation for the analog transfer function $G(S)$, which is given in Eq. (8), is obtained by finding the inverse Laplace transform of $G(S)$ and then substituting the resultant into Eq. (18). Evaluating Eq. (18), one has

$$G_D(Z) =$$

$$T_S\left[\frac{\alpha_0}{1 - Z^{-1}e^{-aT_S}} + \frac{\alpha_1}{1 - Z^{-1}e^{-bT_S}} + \frac{\alpha_2}{1 - Z^{-1}e^{-cT_S}}\right] \qquad (20)$$

where

$$\alpha_0 = \frac{\tau_1 - \tau_2}{(\tau_1 - \tau_{RC})(\tau_1 - \tau_V)} \qquad (21)$$

$$\alpha_1 = \frac{\tau_{RC} - \tau_2}{(\tau_{RC} - \tau_1)(\tau_{RC} - \tau_V)} \qquad (22)$$

$$\alpha_2 = \frac{\tau_V - \tau_2}{(\tau_2 - \tau_1)(\tau_V - \tau_{RC})} \qquad (23)$$

and

$$a = \frac{1}{\tau_1};\ \ b = \frac{1}{\tau_{RC}};\ \ c = \frac{1}{\tau_V} \qquad (24)$$

The digital closed-loop transfer function for this case is given by

$$H(Z) = \frac{AZ[G_D(Z)K(Z)]}{1 + AK[G_D(Z)K(Z)]} \qquad (25)$$

From Eq. (25), the plots of the phase and magnitude responses can be obtained for the digital approximation loop. Figures 6(a) and 6(b) illustrate the closed-loop phase and magnitude responses for both analog and digital loops. The figures show that the response of the digital loop approximated using impulse invariant transformation is the same as that of the analog loop when the sampling frequency is higher than or equal to 80 kHz. When the sampling frequency is less than 80 kHz, the digital loop can encounter serious distortion in both phase and amplitude responses. The recursive implementations $G_D(Z)$ and $K(Z)$ using impulse invariant transformation are shown in Figs. 7(a) and 7(b).

C. Step Invariant Transformation Method

This method preserves the step response at the sampling points when mapping S-domain to Z-domain. The

relationship between the analog and digital transfer function is [12–14]

$$G_D(Z) = \frac{Z-1}{Z} z\left\{\left[L^{-1}\left[\frac{G(S)}{S}\right]\right]_{t=nT_S}\right\} \tag{26}$$

where $z\{.\}$ and $G(S)$ are defined as they are above. The digital approximation $K(Z)$ for $K(S)$ using step invariant transformation can be obtained in a similar manner. The results are

$$K(Z) = \frac{T_S}{(Z-1)} \tag{27}$$

$$\left.\begin{aligned} G_D(Z) &= \beta_0 + \beta_1\left[\frac{1-Z^{-1}}{1-Z^{-1}e^{-aT_S}}\right] \\ &+ \beta_2\left[\frac{1-Z^{-1}}{1-Z^{-1}e^{-bT_S}}\right] \\ &+ \beta_3\left[\frac{1-Z^{-1}}{1-Z^{-1}e^{-cT_S}}\right] \end{aligned}\right\} \tag{28}$$

where

$$\beta_0 = \frac{\alpha_0}{a} + \frac{\alpha_1}{b} + \frac{\alpha_2}{c} \tag{29}$$

$$\beta_1 = -\frac{\alpha_0}{a};\quad \beta_2 = -\frac{\alpha_1}{b};\quad \beta_3 = -\frac{\alpha_2}{c} \tag{30}$$

The parameters α_0, α_1, α_2, a, b, and c are defined in Eqs. (21) through (24). Again, Eq. (25) can be used to evaluate the closed-loop transfer function for this case. The plots of the closed-loop transfer functions for both analog and digital loops are shown in Figs. 8(a) and 8(b). The figures show that the magnitude response approaches the analog response when the sampling frequency is higher than or equal to 100 kHz. However, the phase response suffers serious distortion when the sampling frequency is less than 1 MHz. Thus, in order to achieve the same response as the analog loop, the digital approximation loop using step invariant transformation must be sampled at least at 1 MHz, i.e., this method requires 10 times higher sampling frequency than the previous methods. Table 1 summarizes the results in finding the minimum sampling frequency, F_S, that is required for the digital loop to achieve the same phase and amplitude responses as the analog loop. The recursive implementations of $G_D(Z)$ and $K(Z)$ using step invariant transformation are shown in Figs. 9(a) and 9(b).

IV. Carrier Tracking Performances of the Approximated Digital Loops

The tracking performance of the APLL for high loop signal-to-noise ratio (LSNR) is well known [15,16]. For LSNR > 5 dB, the variance of the tracking phase error is approximated by

$$\sigma_\phi^2 = \frac{N_0 B_L}{P_C} \tag{31}$$

where N_0 is the one-sided thermal noise spectral density, B_L denotes one-sided tracking loop noise bandwidth, and P_C is the carrier power. Note that LSNR = 5 dB is the loop threshold point where the nonlinear theory and linear theory depart severely (by about 1 dB or more in terms of tracking variance). The mathematical expressions for the analog loop bandwidth and carrier power are given by [15,16]

$$2B_L = \frac{1}{2\pi}\int_{-\infty}^{\infty} |H(j\omega)|^2 d\omega \tag{32}$$

where $H(j\omega)$ is the analog closed-loop transfer function, which is identical to Eq. (17) with Z replaced by $s = j\omega$. Using the loop gain, the LPF, the loop filter, the roll-off filter of the VCO, and the VCO integrator given in Eqs. (3) to (7), respectively, the one-sided tracking loop noise bandwidth is calculated using Eq. (32). The result is $B_L = 62$ Hz.

For the digital loops, the one-sided loop noise bandwidth B_{DL} is given by

$$B_{DL} = \frac{1}{4\pi j T_S H^2(1)} \oint_{|Z|\equiv 1} H(Z)H(Z^{-1})\frac{dZ}{Z} \tag{33}$$

where $j = \sqrt{-1}$ and $H(Z)$ is the closed-loop digital transfer function which is given by Eqs. (17) and (25) for BT and IIT/SIT, respectively. The digital loop noise bandwidth for IIT can be calculated by substituting the digital transfer function $G_D(Z)$, shown in Eq. (20), into Eq. (25) and then substituting the resultant into Eq. (33). For SIT, Eq. (26) is used instead of Eq. (20) for the digital transfer function $G_D(Z)$.

In this article, Eq. (32) will be evaluated numerically using an analytical computer program for the three transformation methods under investigation. The numerical results are plotted in Fig. 10. Figure 10 shows a plot B_L/F_S

(or B_LT_S) versus B_{DL}/F_S (or $B_{DL}T_S$) for BT, IIT, and SIT. This figure shows that, for $B_LT_S \leq 0.01$, the tracking loop noise bandwidth of the digital approximation of the analog loop using BT is almost identical to that of the analog loop. On the other hand, the digital loop noise bandwidth obtained by using IIT/SIT departs from the analog loop bandwidth when $B_LT_S \geq 0.001$. Notice that SIT provides the worst digital approximation, and BT is the best among the three transformations. Table 2 gives a brief summary of the numerical results shown in Fig. 10.

Figure 10 shows that, for $B_LT_S \leq 0.01$ (corresponding to $F_S \leq 6.2$ kHz), the digital tracking loop bandwidth approximated by BT is the same as the analog loop. Moreover, the loop bandwidths of the digital loops approximated by IIT and SIT are worse than those of the analog counterpart for $B_LT_S \geq 0.001$ (corresponding to $F_S = 62$ kHz). This implies that in order to achieve the same tracking phase error as the analog loop, the digital loop approximated by BT requires lower sampling frequency than the IIT and SIT loops. For the analog loop with characteristics specified in Section II, it is found that the minimum digital-loop sampling frequency that is required for the digital loop to have the same tracking loop bandwidths as the analog is 6.2 kHz, and this is only achievable through BT. It has been shown in Table 1 that the minimum sampling frequency required to achieve the same phase and amplitude responses as the analog is 80 kHz for both BT and IIT, and 1 MHz for SIT. Hence, what will be the minimum sampling frequency that one would select for optimum performance? The answer to this question will be deferred until Section V.

It should be mentioned that Eq. (33) can also be evaluated analytically by expressing $H(Z)$ in the following form:

$$H(Z) = \frac{b_0Z^4 + b_1Z^3 + b_2Z^2 + b_3Z^1 + b_4}{a_0Z^4 + a_1Z^3 + a_2Z^2 + a_3Z^1 + a_4} \tag{34}$$

and then from Table III in [17], Eq. (35) becomes

$$B_{DL} = \left[\frac{1}{2T_SH^2(1)}\right]\left[\frac{a_0b_0Q_0 - a_0B_1Q_1 + a_0B_2Q_2 - a_0B_3Q_3 + B_4Q_4}{a_0\{(a_0^2 - a_4^2)Q_0 - (a_0a_1 - a_3a_4)Q_1 + (a_0a_2 - a_2a_4)Q_2 - (a_0a_3 - a_1a_4)Q_3\}}\right] \tag{35}$$

where

$$B_0 = b_0^2 + b_1^2 + b_2^2 + b_3^2 + b_4^2; \quad B_1 = 2(b_0b_1 + b_1b_2 + b_2b_3 + b_3b_4) \tag{36}$$

$$B_2 = 2(b_0b_2 + b_1b_3 + b_2b_4); \; B_3 = 2(b_0b_3 + b_1b_4); \; B_4 = 2b_0b_4 \tag{37}$$

$$Q_0 = a_0e_1e_4 - a_0a_3e_2 + a_4(a_1e_2 - e_3e_4); \; Q_1 = a_0a_1e_4 - a_0a_2a_3 + a_4(a_1a_2 - a_3e_4) \tag{38}$$

$$Q_2 = a_0a_1e_2 - a_0a_2e_1 + a_4(a_2e_3 - a_3e_2); \; Q_3 = a_1(a_1e_2 - e_3e_4) - a_2(a_1e_1 - a_3e_3) + a_3(e_1e_4 - a_3e_2) \tag{39}$$

$$Q_4 = a_0\left[e_2(a_1a_4 - a_0a_3) + e_5(a_0^2 - a_4^2)\right] + (e_2^2 - e_5^2)[a_1(a_1 - a_3) + (a_0 - a_4)(e_4 - a_2)] \tag{40}$$

$$e_1 = a_0 + a_2; \; e_2 = a_1 + a_3; \; e_3 = a_2 + a_4 \tag{41}$$

$$e_4 = a_0 + a_4; \; e_5 = a_0 + a_2 + a_4 \tag{42}$$

As an example, for BT, one gets

$$a_0 = 2F_SA_{10}A_{11} + AKA_0; \; a_1 = 2F_S(A_{10}A_{00}A_1 + A_{01}A_{11}A_1 - A_{10}A_{11}B_1 - A_{11}A_{10}A_1) + AK(3A_0 - B_0) \tag{43}$$

$$a_2 = 2F_S A_1(A_{00}A_{01} - A_{01}A_{11}) + 2F_S B_1(A_{10}A_{11} - A_{10}A_{00} - A_{01}A_{11}) + 3AK(A_0 - B_0) \quad (44)$$

$$a_3 = 2F_S B_1(A_{00}A_{10} - A_{01}A_{00} + A_{01}A_{11}) - 2F_S A_1 A_{01}A_{00} + AK(A_0 - 3B_0); \quad a_4 = 2F_S B_1 A_{10}A_{00} - AKB_0 \quad (45)$$

$$b_0 = AKA_0;\ b_1 = AK(3A_0 - B_0);\ b_2 = 3AK(A_0 - 3B_0) \quad (46)$$

$$b_3 = AK(A_0 - 3B_0);\ b_4 = -AKB_0 \quad (47)$$

Having determined the corresponding digital loop noise bandwidth, one can evaluate the variance of the tracking phase error for the digital approximations of the analog loop using the following formula from Eq. (31):

$$\sigma_\phi^2 = \frac{NB_{DL}}{P_C} \quad (48)$$

where N is the total one-sided noise spectral density.

When the command and ranging are turned off, i.e., $m = m_R = 0$, all power is allocated to the carrier and there is no interference from the command and ranging to the carrier tracking loop, and hence $N = N_0$. When the command (or ranging) is turned on, there exists some interference between the carrier and the command (or ranging). Since the ranging tones will be placed farther away from the carrier and the power allocated to the ranging is always smaller than the power allocated to the command, the effects of ranging to the carrier tracking loop are negligible and will not be considered here. However, the effects of the command to the carrier tracking may not be neglected because of the increase in the command data rate. Recently, the international Consultative Committee for Space Data Systems (CCSDS) has considered increasing the maximum command data rate from 2 kbits/sec to 4 kbits/sec and the possibility of using a 32-kHz subcarrier frequency for both 2 kbits/sec and 4 kbits/sec.

To determine the effect of interference of the command on the carrier tracking loop, a model of the command data must be provided. Here it is assumed that the command data symbols are equally likely to be +1's and −1's and that successive symbols are uncorrelated. This assumption leads to the power spectral density (PDS) of a unit power sinusoidal wave subcarrier phase-reversal-keyed by the command data stream. See Eq. (1) for the signal; the PSD is given by [18]

$$S_{CD}(f, T_S, f_{SC}) =$$

$$\sum_{n=2,n=\text{even}}^{\infty} J_n^2(m)\left[\delta(f - nf_{SC}) + \delta(f + nf_{SC})\right]$$

$$+ \sum_{k=1,k=\text{odd}}^{\infty} J_k^2(m)\left[S_D(f - kf_{SC}) + S_D(f + kf_{SC})\right] \quad (49)$$

where

$$S_D(f) = T_S\left[\frac{\sin(\pi fT)}{(\pi fT)}\right]^2 \quad (50)$$

where T is the command symbol period. Note that the PSD shown above is evaluated at the carrier frequency. Hence, when the command is on, the total noise spectral density, N, seen by the carrier tracking loop can be approximated using the following relationship:

$$N = N_0\left[1 + \frac{P_C}{N_0}S_{CD}(B_{DL}, T, f_{SC})\right] \quad (51)$$

where the definition of P_C, taken from Eq. (1) and [16,18], is

$$P_C = (P)J_0^2(m) \quad (52)$$

where $J_0(.)$ is the zero-order Bessel function. Figure 11 shows the theoretical results obtained for the variance of the carrier tracking phase jitter, σ^2, as a function of the received signal-to-noise spectral density ratio (SNR), which is P/N_0, for both analog and digital loops when the command is on. The results were plotted for the modulation index $m = 70$ deg, command subcarrier frequency $f_{SC} =$

32 kHz, command data rate R_S = 2 kbits/sec, and sampling frequency F_S = 1 MHz. As expected, for high sampling frequency, the tracking jitter of the digital loop using BT/IIT/SIT approaches that of the analog loop. Figure 12 presents the numerical results for the digital loop using BT with both command-on and command-off. The theoretical results shown in this figure will be verified by computer simulations discussed in Section V. Note that the relationship between the total received SNR, P/N_0, and the carrier tracking loop signal-to-thermal noise spectral density ratio, P_C/N_0, can be evaluated from Eq. (52), and the results are plotted in Fig. 13.

V. Computer Simulation Results

The digital PPL's shown in Figs. 2 and 3 have been implemented using the Signal Processing Workstation (SPW) of Comdisco, Inc. Simulations have been run to verify the carrier tracking jitter obtained in Section IV and to determine the time responses of the digital loops due to the phase offset between the incoming and the NCO reference phase. The update rate of the loop has been set to be the same as the sampling rate. This is done in the simulation by setting the parameter M = 1 (see Figs. 2 and 3). In addition, the simulations have been performed to determine the minimum achievable sampling frequency for each digital approximation.

A. Measurements of the Tracking Jitter and Time Response of the Digital Loops

Computer simulations for the digital loops approximated by BT (see Fig. 2) and IIT/SIT (see Fig. 3) have been run for both command-on (with modulation index set at 70 deg, command data rate of 2 kbits/sec, and subcarrier frequency of 32 kHz) and command-off at 1-MHz sampling frequency. The simulations were run for 2.5 million iterations, and the variance of the carrier phase jitter was measured for four different noise seeds. Table 3 presents the average results of four noise seeds. For the sake of comparison, the results are also plotted in Fig. 12. On the other hand, Fig. 14 shows the simulation results of the digital loops approximated by IIT and SIT at 1 MHz sampling frequency. The simulation results, for all cases at 1-MHz sampling frequency, are in good agreement with the theoretical results.

Computer simulation has been performed to determine the time responses of the digital approximations of the loops using the three transformation techniques described in Section III. A phase offset of $\pi/9$ rad between the incoming phase of the signal and the reference NCO has been injected into the loop with a 1-MHz sampling rate, and the settling time, t_S, of each loop to the phase offset was measured. Here, the settling time is defined as the time it takes the loop to catch up with the phase offset, or the time it takes the loop to stabilize in the presence of the phase offset. The results are summarized in Table 4 for the command-off and noise-free cases.

B. Minimum Achievable Sampling Frequency

As shown in Fig. 10 and Table 2, for the analog PLL characteristics specified in Section II, the minimum sampling frequencies required for the digital loop to achieve the same analog tracking loop bandwidth are 6.2 kHz and 62 kHz for BT and for IIT/SIT, respectively. On the other hand, it has been shown in Section III that the minimum sampling frequencies required for the digital loop to have the same closed-loop phase and amplitude responses are 80 kHz and 1 MHz for BT/IIT and SIT, respectively. Based on these results, one is tempted to select the smallest sampling frequency so that the requirements on the speed of the digital signal processor and power consumption can be minimized. However, the selected sampling frequencies (based on these criteria) may not be able to provide the required tracking performance. Computer simulation will be used as an additional tool to assist in deciding the minimum achievable sampling frequency. Here, the minimum achievable sampling frequency, denoted as F_{Sm}, is defined as the frequency that satisfies the tracking performance requirement. Table 5 summarizes the simulation results for a 200-kHz carrier frequency, 32-kHz subcarrier frequency, data rate of 2 kbits/sec, modulation index of 70 deg, and P/N_0 of 35 dB-Hz.

The phase jitters shown in Table 5 are then compared with the analog phase jitter of 0.045 rad^2 for this particular case—Eq. (50) with B_{DL} replaced by B_L = 62 Hz. It is observed that the variance of the tracking phase error, σ_ϕ^2, of the digital loop approximated by BT is as good as that of the analog loop when the sampling frequency is about 24.8 kHz. Moreover, the tracking phase errors of the digital loops approximated by IIT and SIT are close to those of the analog loop when the sampling frequencies are 240 kHz and 1000 kHz, respectively. The results for the minimum achievable sampling frequency (or optimum sampling frequency) for the three transformations, shown in Table 5, are then compared to the results obtained in Sections III and IV. Recall that Section III determines the minimum sampling frequency, denoted as F_{Smb}, that is required for the digital loops to achieve the same amplitude and phase responses as the analog loop, and that Section III calculates the minimum sampling frequency, denoted as F_{Smb}, for the digital loops to have the same tracking loop

bandwidth as the analog loop. Table 6 summarizes the final results regarding the optimum sampling frequency that is required for each digital approximation method. Table 6 shows that the minimum achievable sampling frequency for both BT and IIT is about four times F_{Smb} and, for SIT, the minimum achievable sampling frequency is the same as F_{Smr}.

VI. Conclusion

Digital approximations of the current analog deep space carrier tracking loop have been investigated in detail. The performance of each approximation was determined for the closed-loop phase and magnitude responses, carrier tracking jitter, response of the loop to the phase offset, and minimum achievable sampling frequency. The numerical results show that BT appears to give the best performance at a low sampling rate, i.e., about 100 B_L, as compared to the other transformations. The best performance at a low sampling frequency is evident from the closed-loop phase and magnitude response curves, the carrier tracking loop bandwidth curves, and the computer simulation results for the tracking phase error. However, at a high sampling frequency (higher than or equal to 1 MHz, for the case considered in this article), the performance of the DPLL approximated by all three transformations approaches that of the analog loop. It was found that in order to achieve the same tracking phase error as the analog loop, the minimum sampling frequencies required for BT/IIT and SIT are 4 F_{Smb} and F_{Smr}, respectively. Here, F_{Smb} and F_{Smr} denote the minimum sampling frequencies for the digital loops to have the same tracking loop bandwidth and phase/magnitude responses, respectively, as the analog loop. In addition, using the particular analog loop considered in this article, the simulation results show that the response to the incoming phase offset of $\pi/9$ rad of the digital loop approximated by BT is faster than that of the loops approximated by IIT and SIT by about 20 and 30 msec, respectively.

As pointed out in [9], in the absence of noise, the digital loop approximated by the IIT method exhibits less overshoot and ringing in the output response than the others. However, this may not be the key criterion in the selection of the optimum transformation method for approximating the analog loop. This article has shown that, for applications that require low sampling frequency, the BT method appears to give the best performance in terms of the tracking phase error and response to the initial phase offset. Therefore, when the key requirements, such as low sampling rate, low tracking phase error, and fast response to the initial phase offset, for approximating the analog loop are desired, then the BT method is recommended. Furthermore, the performance evaluation approach presented in this article can easily be extended to (1) find the minimum achievable sampling frequency required to approximate any analog loop, and (2) determine the tracking phase error of the digital approximation of the analog loop.

Acknowledgments

The authors thank S. Million and B. Shah for their assistance in computer simulation, R. Sadr and S. Kayalar for their useful comments and suggestions and A. Kermode for his constant support.

References

[1] W. C. Lindsey and C. M. Chie, "A Survey of Digital Phase-Locked Loops," *Proceedings of the IEEE*, vol. 69, no. 4, pp. 410–431, April 1981.

[2] C. M. Chie, "Analysis of Digital Phase-Locked Loops," Ph.D. dissertation, University of Southern California, January 1977.

[3] J. B. Thomas, *An Analysis of Digital Phase-Locked Loops*, JPL Publication 89-2, Jet Propulsion Laboratory, Pasadena, California, February 1989.

[4] R. Sadr and W. J. Hurd, "Digital Carrier Demodulation for the DSN Advanced Receiver," *The Telecommunications and Data Acquisition Progress Report 42-93, vol. January-March 1988*, Jet Propulsion Laboratory, Pasadena, California, pp. 45–63, May 15, 1988.

[5] R. Sfeir, S. Aguirre, and W. J. Hurd,"Coherent Digital Demodulation of a Residual Carrier Signal Using IF Sampling," *The Telecommunications and Data Acquisition Progress Report 42-78, vol. April-June 1984*, Jet Propulsion Laboratory, Pasadena, California, pp. 135–142, August 15, 1984.

[6] S. Aguirre and W. J. Hurd, "Design and Performance of Sampled Data Loops for Subcarrier and Carrier Tracking," *The Telecommunications and Data Acquisition Progress Report 42-79, vol. July-September 1984*, Jet Propulsion Laboratory, Pasadena, California, pp. 81–95, November 15, 1984.

[7] R. Kumar and W. J. Hurd, "A Class of Optimum Digital Phase-Locked Loops for the DSN Advanced Receiver," *The Telecommunications and Data Acquisition Progress Report 42-83, vol. July-September 1985*, Jet Propulsion Laboratory, Pasadena, California, pp. 63–80, November 15, 1985.

[8] S. Aguirre, W. J. Hurd, R. Kumar, and J. Statman, "A Comparison of Methods for DPLL Loop Filter Design," *The Telecommunications and Data Acquisition Progress Report 42-87, vol. July-September 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 114–124, November 15, 1986.

[9] D. J. Boman, "Performance of a Low Sampling Rate Digital Phase-Locked Loop," master's thesis, Arizona State University, May 1987.

[10] C. R. Cahn and D. K. Leimer, "Digital Phase Sampling for Microcomputer Implementation of Carrier Acquisition and Coherent Tracking," *IEEE Transactions on Communications*, vol. COM-28, no. 8, pp. 1190–1196, August 1980.

[11] R. Cali and G. Ferrari,"Algorithms for Computing Phase and AGC in Digital PLL Receivers," *RF Design*, pp. 33–40, October 1992.

[12] L. R. Rabiner and B. Gold, *Theory and Applications of Digital Signal Processing*, Englewood Cliffs, New Jersey: Prentice Hall, 1975.

[13] W. D. Stanley, *Digital Signal Processing*, Reston, Virginia: Reston Publishing Company, 1975.

[14] K. Ogata, *Discrete Signal Processing*, Englewood Cliffs, New Jersey: Prentice Hall, 1987.

[15] J. Yuen, ed., *Deep Space Telecommunications Systems Engineering*, New York: Plenum Press, Chapters 3 and 5, 1983.

[16] J. K. Holmes, *Coherent Spread Spectrum Systems*, New York: John Wiley & Sons, Chapters 3 and 4, 1982.

[17] E. I. Jury, *Theory and Application of the z-Transform Method*, Malabar, Florida: Robert E. Krieger Publishing Company, 1964.

[18] T. M. Nguyen,"Closed Form Expressions for Computing the Occupied Bandwidth of PCM/PSK/PM Signals," *IEEE/EMC-91 Proceedings on Electromagnetic Compatibility*, Cherry Hill, New Jersey, pp. 414–415, August 1991.

Table 1. Minimum sampling frequency, F_S, required for the digital approximation to achieve the same phase and amplitude responses as the analog loop.

Transformation method	Minimum F_S required, kHz
Bilinear	80
Impulse invariant	80
Step invariant	1000

Table 2. Loop noise bandwidth of the digital approximations for F_S = 6.2 kHz and 62 kHz.

Transformation method	Analog loop noise bandwidth, B_L, Hz	Digital loop noise bandwidth, B_{DL}, at F_S = 6.2 kHz, Hz	Digital loop noise bandwidth, B_{DL}, at F_S = 62 kHz, Hz
Bilinear	62	62	62
Impulse invariant	62	76.88	62
Step invariant	62	114.08	62

Table 3. Simulation results for command-on (with m = 70 deg, f_{SC} = 32 kHz, and R_S = 2 kbits/sec), and command-off at 1-MHz sampling frequency.

P/N_0, dB-Hz	Variance of the carrier phase jitter, rad^2					
	Command-off			Command-on		
	BT	IIT	SIT	BT	IIT	SIT
30	0.064000	0.062250	0.063800	0.151000	0.150750	0.150250
35	0.019670	0.019175	0.019400	0.045100	0.044725	0.045250
40	0.006165	0.006013	0.006040	0.014000	0.013875	0.013875
45	0.001945	0.001895	0.001900	0.004398	0.004360	0.004355
50	0.000613	0.000598	0.000598	0.001390	0.001377	0.001378
55	0.000194	0.000189	0.000189	0.000441	0.000435	0.000445
60	6.145×10^{-5}	5.973×10^{-5}	5.960×10^{-5}	0.000139	0.000140	0.000140
65	6.955×10^{-5}	1.818×10^{-5}	1.890×10^{-5}	4.66×10^{-5}	4.623×10^{-5}	4.623×10^{-5}

Table 4. Settling time, t_S, for the phase offset of $\pi/9$ rad.

Transformation method	Settling time, t_S, sec
Bilinear	0.12
Impulse invariant	0.14
Step invariant	0.15

Table 5. Tracking phase jitter as a function of sampling frequency.

F_S, kHz	σ_ϕ^2, rad^2		
	BT	IIT	SIT
16.5	0.0725	Out of lock	Out of lock
18.5	0.0489	Out of lock	Out of lock
24.8	0.0476	Out of lock	Out of lock
240	0.0453	0.0458	Out of lock
1000	0.0451	0.0447	0.0453

Table 6. Minimum achievable sampling frequency for each transformation method.

Transformation method	Minimum sampling frequency required to achieve the same phase/ amplitude responses, F_{Smr}, kHz	Minimum sampling frequency required to achieve the same analog loop bandwidth, F_{Smb}, kHz	Minimum achievable sampling frequency for a specified tracking jitter, F_{Sm}, kHz	Remarks
BT	80	6.20	24.8	$F_{Sm} = 4\ F_{Smb}$
IIT	80	62.0	240	$F_{Sm} = 3.9\ F_{Smb}$
SIT	1000	62.0	1000	$F_{Sm} = F_{Smr}$

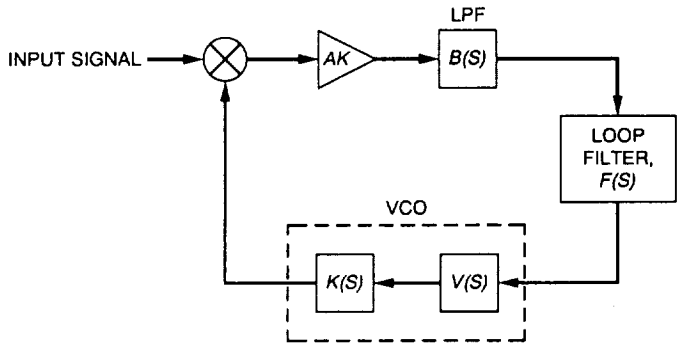


Fig. 1. Simplified block diagram of the analog PLL.

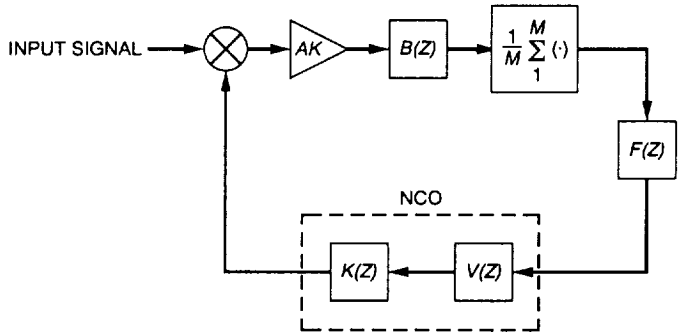


Fig. 2. Digital approximation of the analog PLL—configuration I.

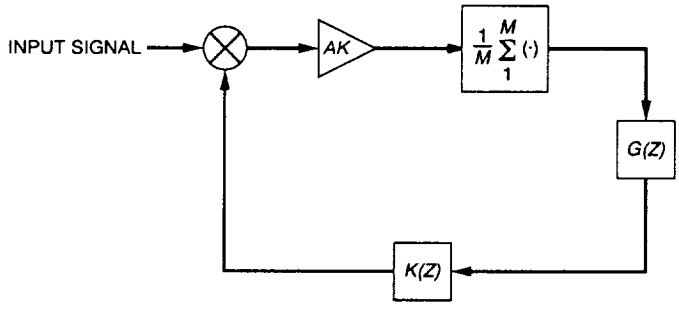


Fig. 3. Digital approximation of the analog PLL—configuration II.

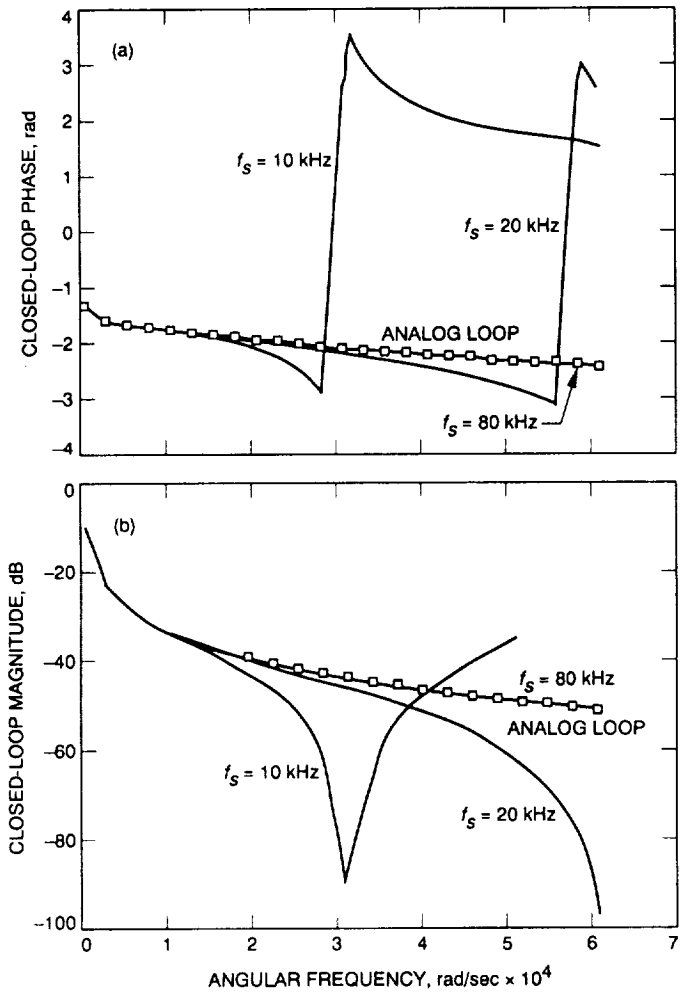


Fig. 4. Closed-loop response of digital approximation using the bilinear transformation method: (a) closed-loop phase characteristics and (b) closed-loop magnitude response.

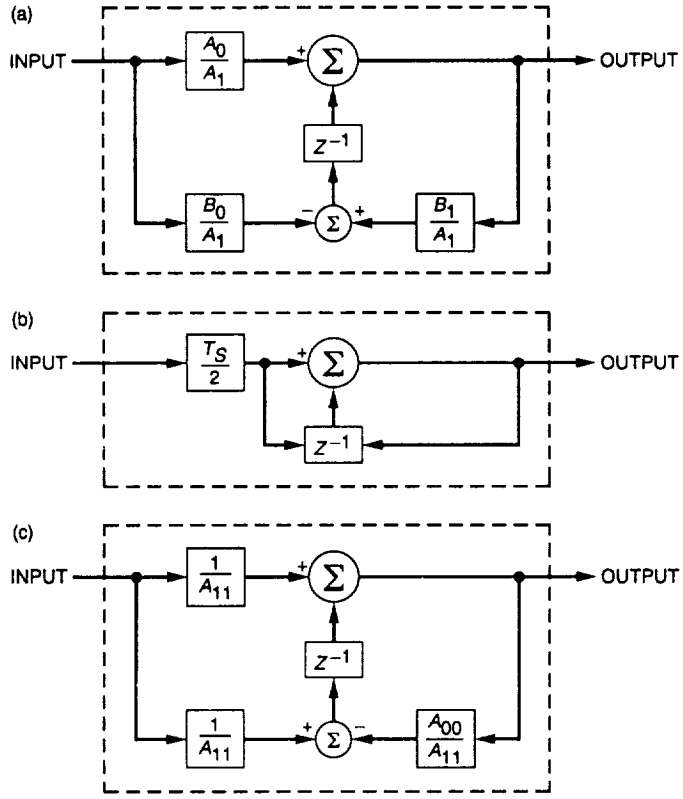


Fig. 5. Recursive implementation of (a) the loop filter *F(Z)* (b) the integrator *K(Z)* and (c) the lowpass filter *B(Z)* using bilinear transformation.

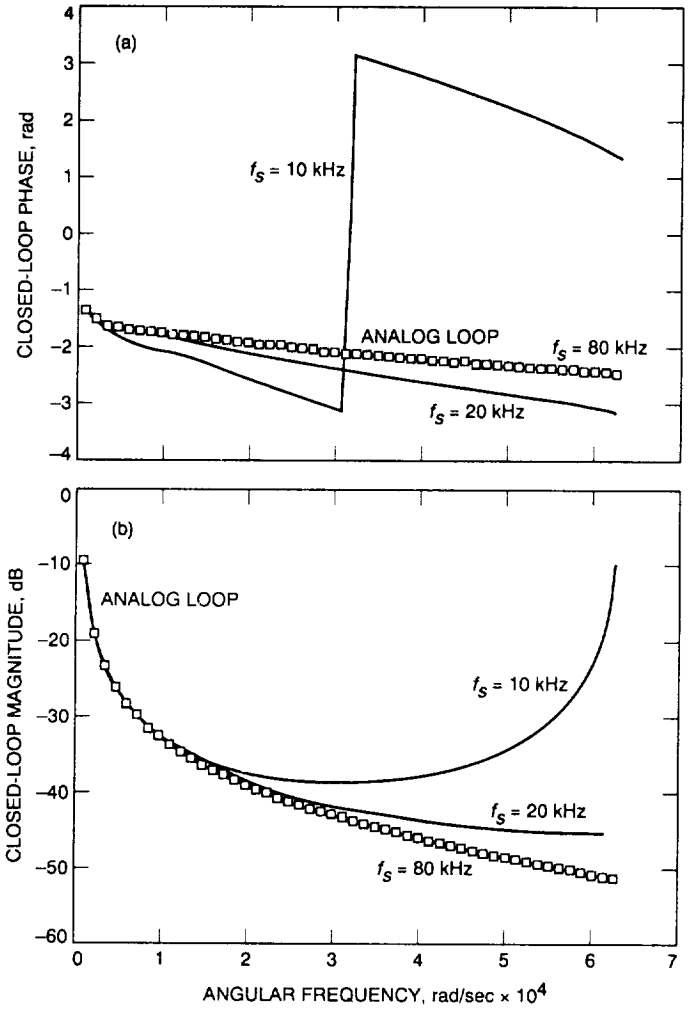


Fig. 6. Closed-loop response of digital approximation using the impulse invariant transformation method: (a) closed-loop phase characteristics and (b) closed-loop magnitude response.

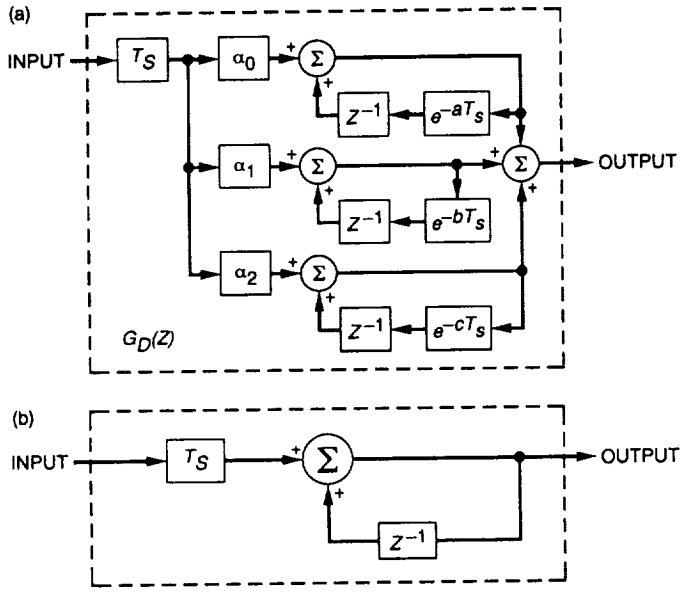


Fig. 7. Recursive implementation of (a) the open-loop transfer function $G_D(Z)$ and (b) the integrator $K(Z)$ using impulse invariant transformation.

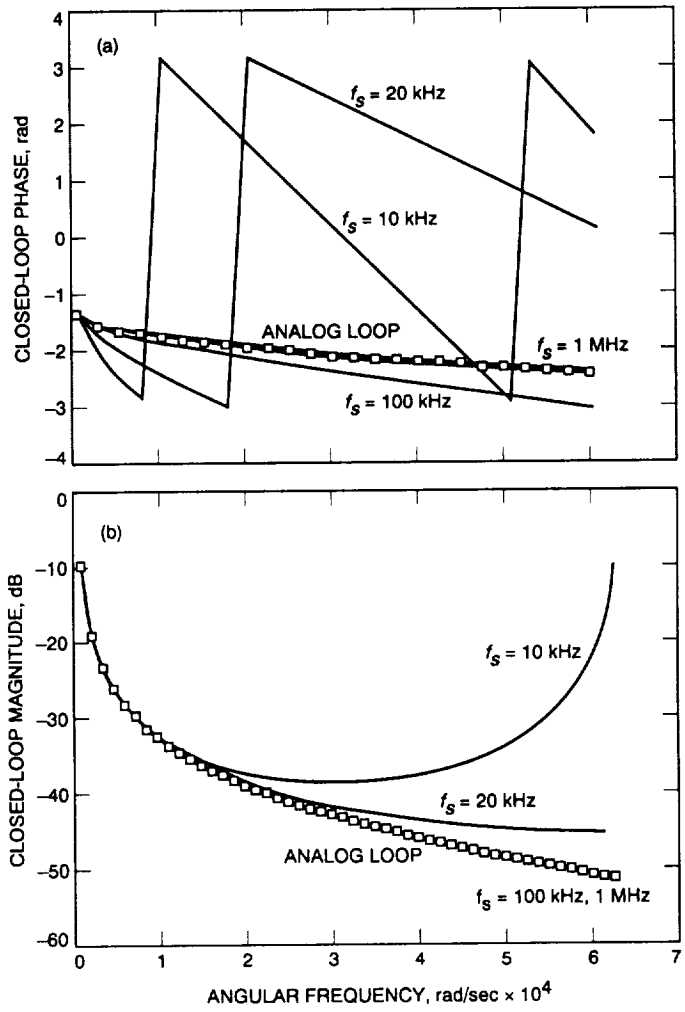


Fig. 8. Closed-loop response of digital approximation using the step invariant transformation method: (a) closed-loop phase characteristics and (b) closed-loop magnitude response.

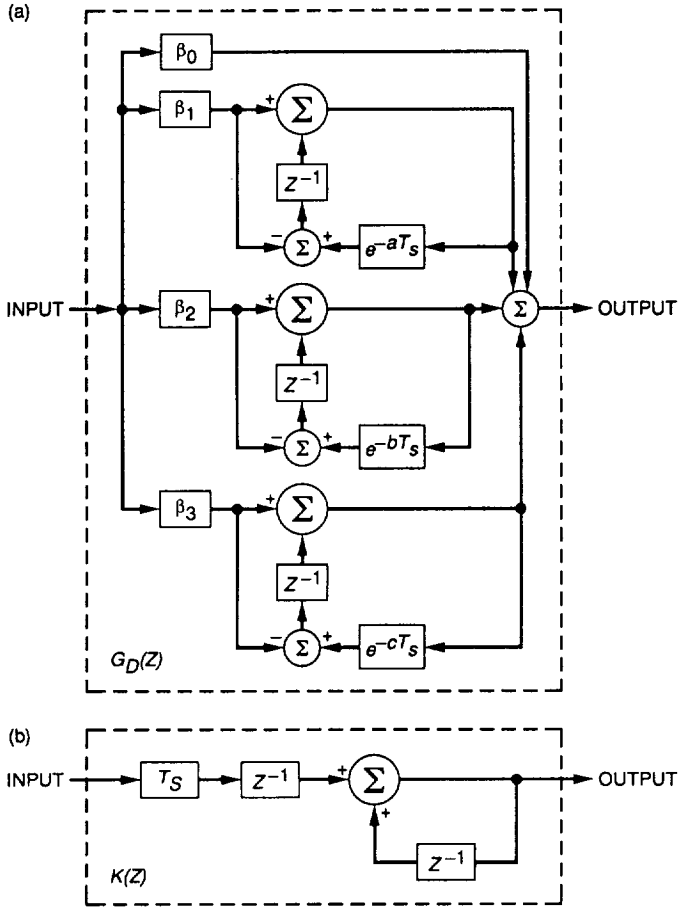


Fig. 9. Recursive implementation of (a) the open-loop transfer function $G_D(Z)$ and (b) the integrator $K(Z)$ using step invariant transformation.

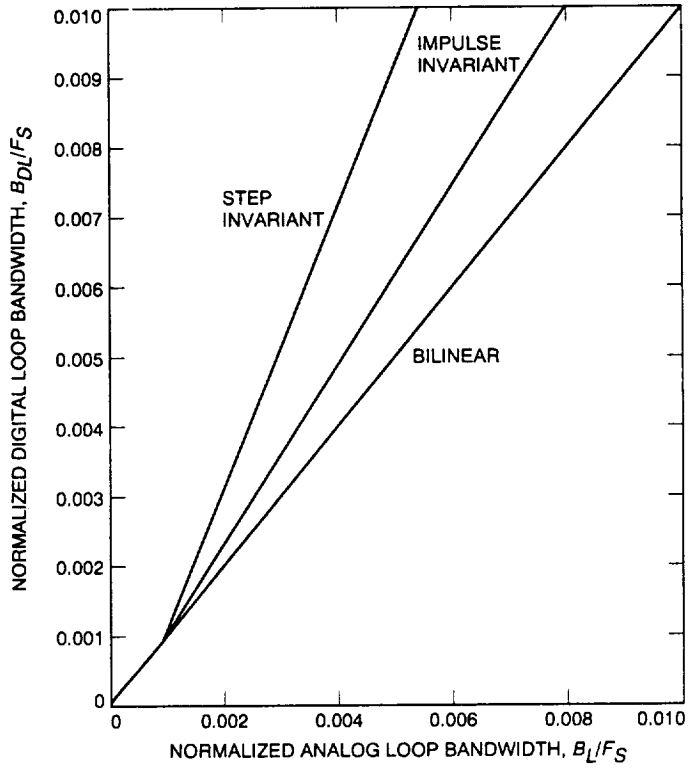


Fig. 10. One-sided digital loop bandwidth for three transformation methods.

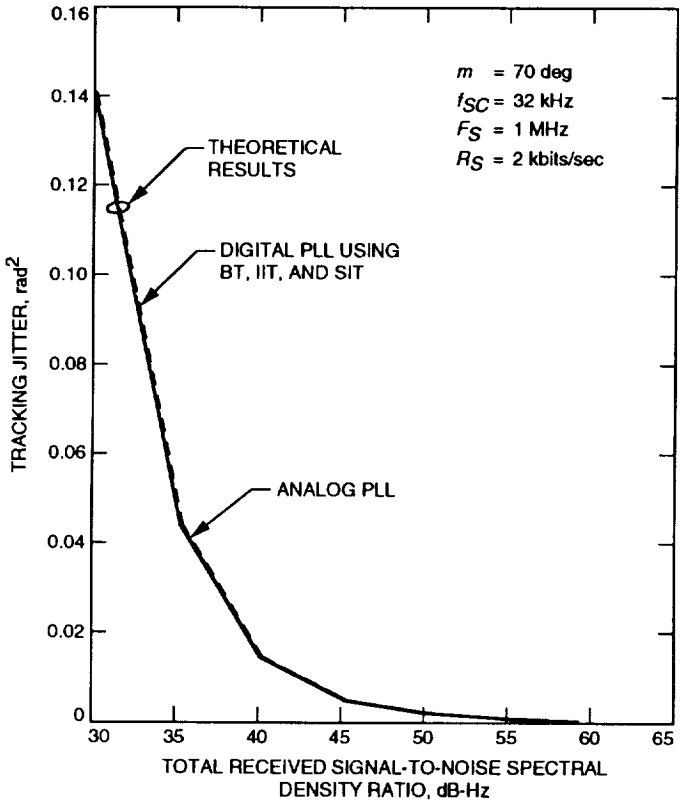


Fig. 11. Theoretical comparison of tracking jitter for analog and digital loops for command-on.

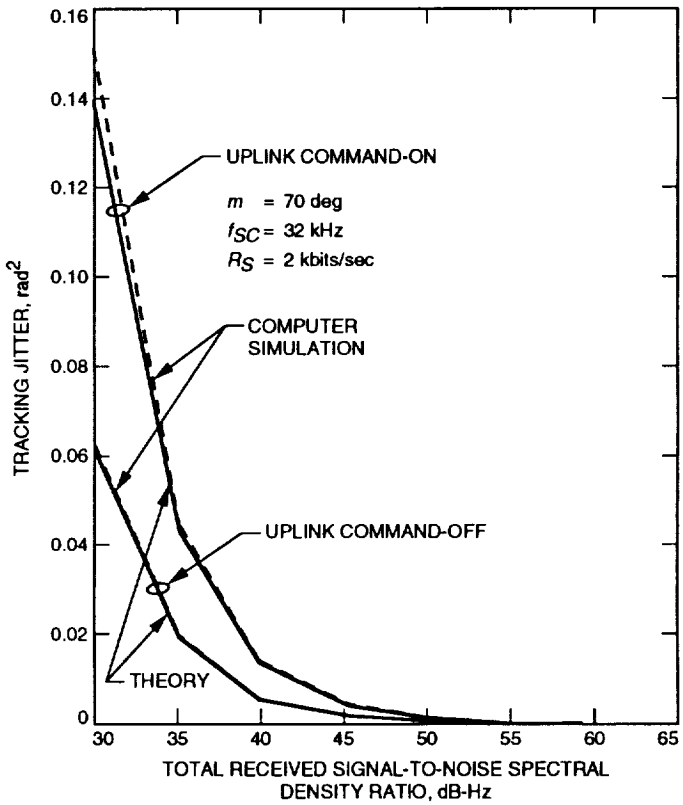


Fig. 12. Tracking phase jitter for the digital loop using bilinear transformation.

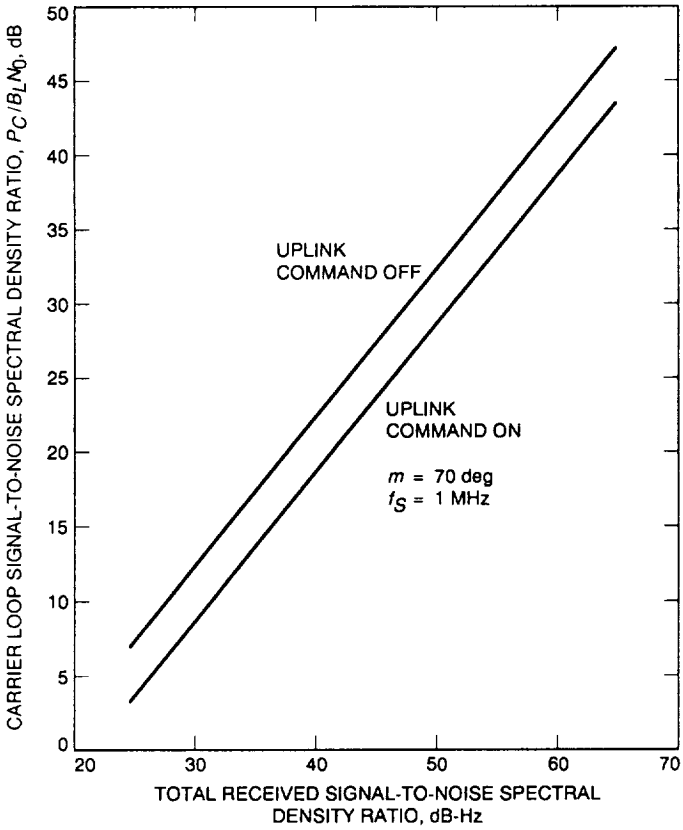


Fig. 13. Carrier loop received signal-to-noise spectral density ratio versus total received signal-to-noise spectral density ratio.

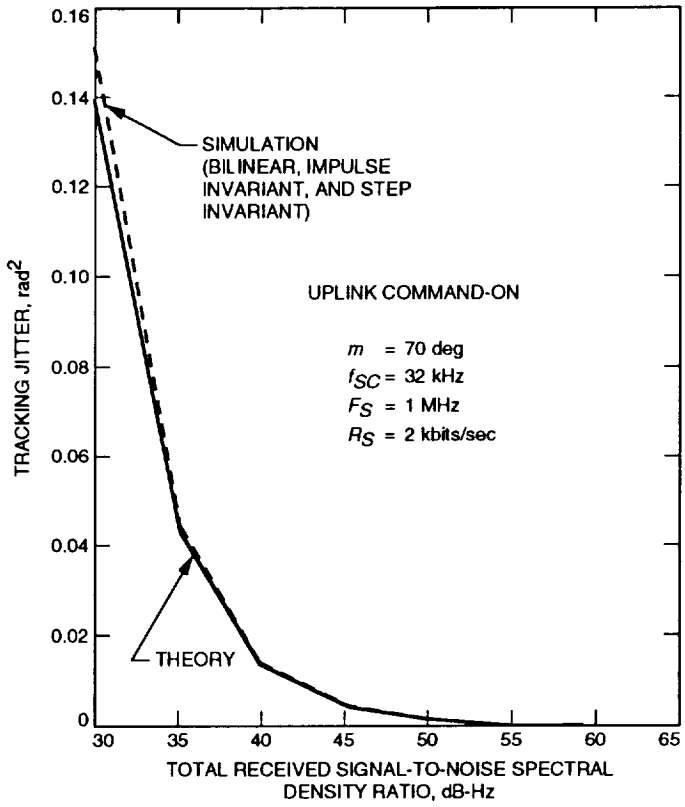


Fig. 14. Comparison of theoretical and simulated tracking phase jitter.

N94-35559

TDA Progress Report 42-117

May 15, 1994

The Development and Application of Composite Complexity Models and a Relative Complexity Metric in a Software Maintenance Environment

J. M. Hops
Radio Frequency and Microwave Subsystems Section

J. S. Sherif
Software Product Assurance Section
and
California State University, Fullerton

A great deal of effort is now being devoted to the study, analysis, prediction, and minimization of software maintenance expected cost, long before software is delivered to users or customers. It has been estimated that, on the average, the effort spent on software maintenance is as costly as the effort spent on all other software costs. Software design methods should be the starting point to aid in alleviating the problems of software maintenance complexity and high costs. Two aspects of maintenance deserve attention: (1) protocols for locating and rectifying defects, and for ensuring that no new defects are introduced in the development phase of the software process, and (2) protocols for modification, enhancement, and upgrading. This article focuses primarily on the second aspect, the development of protocols to help increase the quality and reduce the costs associated with modifications, enhancements, and upgrades of existing software. This study developed parsimonious models and a relative complexity metric for complexity measurement of software that were used to rank the modules in the system relative to one another. Some success was achieved in using the models and the relative metric to identify maintenance-prone modules.

I. Introduction

A. Project Objectives

The primary objective of this study was to determine whether software metrics could help guide our efforts in the development and maintenance of the real-time embedded systems that we develop for NASA's Deep Space Network (DSN). Generally, the systems that are developed control receivers, transmitters, exciters, and signal paths through the communication hardware. The most common programming language in our systems is PL/M for Intel 8080, 8086, and 80286 microprocessors; and the systems range in size from 20,000 to 100,000 non-commented lines of code (NCLOC). Approximately 65 percent of the fund-

ing received in our environment is dedicated to extending the life span of the previously developed systems; of this, 15 percent is spent on finding and fixing defects, while 85 percent is spent on adding automation features, adding capabilities, and increasing capacity.

Our efforts have been successful in that the life spans of our systems now range from 4 to 8 years and are increasing. As support for new spacecraft becomes necessary, these older systems are being used in new ways, thereby increasing the importance of high-quality, defect-free, and cost-effective enhancements to the software. Protocols and guidance for locating and rectifying defects in the software-sustaining environment were deemed critical, especially with the added complications that the maintainers of the systems are not the original developers and that there is little or no confidence in the software documentation.

Specifically, we were looking for ways to identify which modules should be reengineered and which modules would need extra development and test time in order to maintain. The problems we face in our environment are quite common in the industry. Software maintenance cost is about two to four times the original development cost [3,13,10,21]. Charette [5] emphasizes the fact that 60 to 80 percent of the total software costs are related to maintenance. This will likely remain so for the indefinite future [7,11,24].

Figure 1 shows the initial cost breakdown in developing a new project (unfortunately with maintenance costs hidden), and Fig. 2 shows the costs of software during its life cycle, as discussed by Zelkowitz [34]. Software maintenance is not what people think it is: Software maintenance actually encompasses fixing software errors in addition to software enhancements and adding new functions to existing systems, system conversion, training and supporting users, and improving system performance [31–33]. Error correction, which is often perceived as the substance of maintenance, is only a small part of the software maintenance effort [8,4]. Table 1 shows the distribution of the average time spent on various maintenance tasks for 4 years, as reported by Lientz and Swanson [19]. Note that functional enhancement constitutes the major portion of the time spent on software maintenance. Charette [5] discusses another reason why the cost of software is so high and cites some statistics as reported by the Comptroller General [6] and as shown in Table 2. It is reported that only 2 percent of the software contracted for could work on delivery; 3 percent could work after some rework; 45 percent was delivered, but was never successfully put to use; 20 percent was used, but was either extensively reworked or abandoned; and 30 percent was paid for, but was never delivered.

For the study described in this article, we took the following steps:

(1) Determined what the literature suggests.

(2) Developed a course of action to be tried on one of our operational systems that would be representative of all the others.

(3) Performed the steps and analyzed the results.

The process and results of each of these steps are described below.

B. Suggestions from the Literature and Course of Action

One of the earlier studies encountered pertaining to our objectives was undertaken by Shen, Yu, Thebaut, and Paulsen [27]. They assessed the potential usefulness of product and process metrics in identifying components of the system that were most likely to contain errors. Their goal was to establish an empirical basis for the use of objective criteria in developing strategies for the allocation of testing effort in the software-maintenance environment. It was found that the number of unique operands, as defined by Halstead [14], was the best predictor of problem reports on modules that were reported after the initial delivery. Additionally, simple metrics related to the number of unique operands, such as the cyclomatic complexity (defined by McCabe [20]), also performed well. Shen et al. concluded that these metrics are useful in finding error-prone modules at an early stage [27].

In 1987, Kafura and Reddy [17] published the results of their study on using software complexity metrics during the software maintenance phase of a system. They related seven separate metrics to the experience of maintenance activities on medium-sized systems. Two of the results reported were that the overall complexity of a system grows with time and that the individual complexity scores of the software modules agree well with the expert opinions of the programmers. Their conclusion was that metrics could form the control element in a formal maintenance method.

Harrison and Cook [15,16] discuss the decision, frequently encountered by software maintenance personnel, of whether to make an isolated change in a module or to totally redesign and rewrite the module anew. They developed an objective decision rule to identify modules

that should be rewritten rather than modified. This decision rule is whether the total change in the Halstead software science volume metric exceeds a threshold value. This threshold value seems to be subjective since it depends upon the decision maker's risk-taking propensity and experience and since it must be tuned for a particular environment.

Lennselius, Wohlin, and Vrana [18] discuss the possibility of using complexity metrics to identify error-prone, and thus maintenance-prone, modules. They suggest that a module whose complexity lies at least one standard deviation above the acceptable mean of complexity of the project may be considered to be a maintenance-prone module. The authors, however, emphasize that metrics cannot replace the decision-making process of software managers.

Rodriguez and Tsai [23] use discriminant analysis to develop a methodology to evaluate software metrics. They suggest that when classifying units of software as either complex or normal, more attention is usually paid to the complex group to either redesign it or test it more thoroughly. Their methodology is based on the assumption of normal distribution and homogeneity of variances of the two groups. The authors consider 13 metrics depicting Halstead's software science metrics, McCabe complexity metrics, and NCLOC metrics. They conclude that these metrics are correlated.

Stalhane [29] discusses how to estimate the number of defects in a software unit from various software metrics and how to estimate the reliability of the same software. The author also concludes that complexity increases as the size of code increases. Stalhane asserts that misunderstanding the specifications will increase with the specification complexity and that complexity may be transferred to the code and thus lead to maintenance-prone complex code and complex modules.

Munson and Khoshgoftaar [21] employ factor analytic techniques to reduce the dimensionality of the complexity problem space to produce a set of reduced metrics. The reduced complexity metrics are subsequently combined into a single relative complexity measure for the purpose of comparing and classifying programs. In particular, the relative complexity metric can be seen to represent the complexity of a particular software module at a particular level of system release. The authors investigate McCabe complexity metrics, Halstead software science metrics, and NCLOC metrics. The comparison of complexity is again of a relative and subjective nature.

Binder and Poore [2] investigate the possibility of including the number of comments in the code as a variable in determining the quality of the code. They assert that comments only contribute to quality when they are needed and meaningful. The authors suggest a software quality measure called the "LB-ratio," which is defined as the ratio of the number of operators to the sum of the number of operands and the number of comments. The authors agree that their experiments with the LB-ratio need additional work and refinement since including the concept of meaningful comments in the formula seems to be problematic and subjective at best.

The following suggestions were deduced from these sources:

(1) An estimate of errors and reliability can be determined from software product metrics [20,27,29].

(2) Software product metrics could be used to find error-prone modules and could form the control element in a formal software maintenance methodology [15–18].

(3) The software product metrics that may be considered include all of Halstead's software science metrics, McCabe's complexity metric [14,23,27], and NCLOC [21].

(4) Factor analysis can be used to identify those software measures that are highly and significantly related to all other measures. This economy of description will facilitate the analysis of software complexity [21].

(5) Comments in the code contribute to the quality of software [2].

We therefore took the following actions:

(1) Determined the Halstead software science, McCabe complexity, NCLOC, and LB-ratio from sequential releases of a representative software system.

(2) Performed factor analysis on the metrics from the software modules to determine the unique dimensions represented by the metrics.

(3) Proposed a model to calculate a relative metric.

(4) Determined if this metric can identify maintenance-prone modules in the software by using the mean-plus-one standard deviation as the relative metric cut-off value.

II. Method, Analysis, and Results

A. Representative System and Metrics Collection

1. Nature of Software. We analyzed the source program in the very long baseline interferometry (VLBI) receiver controller (VRC) software system by using factor analysis for 16 software measures. The source program is a real-time embedded system in the receiver–exciter subsystem of NASA's DSN. It serves as a communication interface to VLBI subsystems and configures and monitors the status of the narrow-channel bandwidth VLBI receiver assembly. Three releases of the system software were analyzed: OP-B (222 modules), OP-C (224 modules), and a draft version of OP-D (235 modules). These were used as a representative maintenance project in this study. The source code for these three releases was originally written in PL/M but was later converted to C using the PLC86 conversion program (from Micro-Processor Services).

2. Software Metrics and Measures. Software metrics are quantitative measures of certain characteristics of a development project that can be valuable management and engineering tools. Software metrics can be used to achieve various project-specific results, such as predicting source-code complexity at the design phase; monitoring and controlling software reliability and functionality; predicting cost and schedule; and identifying high-risk modules in a software project [28].

The 16 software measures that were used to analyze the VRC software are given in Table 3. The first eight measures belong to the Halstead software science family of software complexity measures. Halstead [14] uses a series of software science equations to measure the complexity of a program based on the lexical counts of symbols used. Generally, the measurements are made for each module, and the total measurements of the modules constitute the measurement of the program. Halstead's metrics become available only after the coding is done, and therefore can be of use only during the testing and maintenance phases. Although Halstead's metrics are useful in determining the complexity of programs, their weaknesses are that they do not measure control flow complexity and have little predictive value.

Measures 9 and 10, i.e., VG_1 and VG_2, belong to McCabe and were adapted from the mathematical concepts of graph theory. McCabe cyclomatic complexity metric VG_1 is a measure of the maximum number of linearly independent circuits in a program control graph. The primary purpose of this metric is to identify software modules that will be difficult to test or maintain, as explained by McCabe [20]. The value of the McCabe metric is available only after the detailed design is done. Although the McCabe metric is very useful for measuring control flow complexity, its weakness is that it is not sensitive to program size; for example, if programs of different sizes are composed exclusively of sequential statements, then they may have the same cyclomatic number.

Measures 11–15 deal with the size of the program or the number of lines. Although many researchers do not find this measure as appealing, Boehm [3] points out that no other metric has a clear advantage over NCLOC as a metric. It is easy to measure, is conceptually familiar to software developers, and is used in most productivity databases and cost estimation models.

Measure 16, the LB-ratio, is defined by Binder and Poore [2] as the ratio of the number of operators to the sum of the number of operands and the number of comments. It appears to capture the idea of distinguishing between meaningful comments in the code and just comments in general. The weakness of this metric is its reliance on defining the number of meaningful comments, which seems to be more subjective than quantitative.

B. Analysis of Data, Models, and Validation

The 16 software measures of the three releases of the VRC code, OP-B, OP-C, and draft OP-D, were analyzed using factor analysis, correlation, analysis of variance, and regression analysis. Table 4 shows the number of modules and the mean value per module for each of the 16 measures. Figures 3–5 show the correlation matrix of the 16 measures for the three releases. The data show a high degree of correlation. Except for the LB-ratio measure, the remaining 15 measures are highly correlated. It can be seen that the Halstead volume metric (V), the McCabe cyclomatic complexity metric (VG_1), and the NCLOC metric are highly and significantly correlated, while the LB-ratio metric is not. These results agree with those of other researchers, such as Ramamurthy and Melton [22], Gill and Kemerer [12], Samadzadeh and Nandakumar [25], Basili and Hutchins [1], Evangelist [9], and Kafura and Reddy [17].

The factor analysis matrix is shown in Table 5. All measures except the LB-ratio are loaded on factor 1, and thus there is no cross-loading. This is a desired result, since cross-loading on many factors makes the interpretation of the result ambiguous. The analysis of variance of the three sets of releases did not show any significant difference at the level of significance of 0.05. This means that, on the average, the values of, say, the McCabe cyclomatic complexity metric (VG_1) of the three releases are

not significantly different at alpha of 5 percent. The same is also true for the other 15 measures.

Regression analysis had been used to develop models of relationships of the most interrelated measures. These are the Halstead volume metric (V), the McCabe cyclomatic metric (VG_1), and the non-commented lines of code ($NCLOC$) metric, as discussed next.

1. Factor Analysis Discussion. Three releases of software were analyzed by factor analysis to show the existence of meaningful relationships among known software complexity measures. The analysis shows the number of factors where software complexity measures tend to load high or low, and also the percentage of the variability explained by each factor. This research also shows the matrix of correlation summarizing the relationships among the 16 software complexity measures for each release.

Factor analysis of the three releases of software had shown that the first 15 measures of complexity are closely related to some measure of similarity and are consequently all interrelated. However, the 16th complexity measure (LB-ratio) does not seem to be typical of the other 15 measures, and thus it is unlike the rest of the data set. The 3 releases show 2 factors that concisely state the pattern of relationships within the 16 measures. However, measures 1–15 load most strongly on the first factor with explained variability of 90 to 91 percent, while the second factor displays less interesting patterns with loading of 9 to 10 percent. Factor analysis had also shown that three complexity measures, the McCabe cyclomatic complexity metric (VG_1), the Halstead volume metric (V), and ($NCLOC$), are highly and strongly related. Therefore, in order to achieve an economy of description, these three measures are considered to give a strong similarity and representation of all the 15 measures.

The correlation matrix for each release of the software also shows that the first 15 complexity measures are related, while the LB-ratio measure is not related or interrelated to any of the other 15 measures.

Analysis of variance does not show any significant difference between the three releases at the level of significance of 5 percent. This means that as the software evolves through its releases, the interrelationships between the complexity measures seem to be preserved. However, we should note that without normalization to size, adding on to a program will make a more complex program. This seems to agree with the findings of other researchers, as discussed by Valett and McGarry [30], Harrison and Cook [15], and Schneidewind [26].

Since factor analysis techniques showed that the first 15 software measures are closely related to some measure of similarity, and since 3 of these measures, the McCabe cyclomatic complexity metric (VG_1), the Halstead volume metric (V), and the NCLOC metric, are highly and significantly related, they are considered to give a strong similarity and representation of all 15 measures. This economy of description made it appealing to develop a set of parsimonious models for software complexity measurements using data from the three software releases. The five composite models together with their coefficients of determination (R^2) are shown in Table 6.

Statistical analysis, model back testing, and model testing with independent segments of software are used for validation of the composite models and ascertaining their degree of accuracy. The developed models had shown a high degree of accuracy in predicting software complexity, and thus they can serve as a baseline for other software projects in identifying software modules with high complexity (maintenance prone), so that actions can be taken before their release to users.

2. Back Testing of Models. The five composite complexity models shown in Table 6 were checked with actual data from the three releases, OP-B, OP-C, and OP-D. Table 7 and Fig. 6 show the actual average values of the dependent variables (VG_1) and values predicted by the first three models. Table 8 and Fig. 7 show the actual average values of (V) and values predicted by models 4 and 5. It can be seen that the difference in predicting (VG_1) by the first three composite models ranges from 3.2 to 10.6 percent below the actual average value of (VG_1), as calculated by the McCabe cyclomatic complexity metric. Also, the difference in predicting (V) by models 4 and 5 ranges from 1.2 to 1.3 percent above the actual average value of (V), as calculated by Halstead's volume metric.

3. Testing the Five Composite Models by External Check. The five composite complexity models were tested against four independent segments of software with characteristics as shown in Table 9. A sample calculation of actual average values of (VG_1) and values predicted by model 1 for the four segments of software is shown in Table 10. The summary of the actual grand average values of (VG_1) and (V) and their values, as predicted by models 1, 2, and 3 and models 4 and 5, respectively, for the four segments of software, is shown in Tables 11 and 12 and

Figs. 8 and 9. It can be seen that the difference in predicting (VG_1) by the first three composite models ranges from 17.3 percent below to 0.7 percent above the actual average value of (VG_1). Also, the difference in predicting (V) by models 4 and 5 is 9.7 percent above the actual average value of (V) for the four segments of software.

C. Parsimonious Model and Relative Complexity

Since the five complexity models developed in this study show direct relationships between (VG_1) and (V) and also ($NCLOC$), we chose the third model,

$$<VG1> = 0.786 + 0.0013(V) + 0.0976(NCLOC)$$

as a representative model for estimating the value of (VG_1), given the measured values of (V) and ($NCLOC$).

1. Development of the Relative Complexity Metric. We propose to capture the total complexity of a program based on its control flow complexity, the lexical counts of symbols used, and the program size. In essence, a complexity metric that accounts for a program total complexity due to volume and control flow and normalized by the number of lines of code would present a relative complexity metric that is more useful to consider for detecting maintenance-prone programs. The relative complexity metric (RCM) will be derived for each module from the measured value of (V), the estimated value of (VG_1) from model 3, and normalized by the module lines of code. The RCM for a module is

$$(RCM)_i = \left(\frac{<VG_1> + V}{NCLOC}\right)_i$$

2. Analysis of the Three Releases Using the Relative Complexity Metric. The RCM was used to analyze the modules of the three releases, as shown in Table 13. Note that, as reported by Kafura and Reddy [17], the RCM has grown with each release, from a 2799 total in OP-B to a 3470 total in the draft of OP-D.

Using the criterion of the mean relative complexity value plus one standard deviation as a cut-off value for acceptable modules, we can identify those modules that can be considered as outliers, or maintenance-prone modules. Results for the three releases are given in Table 14.

In order to determine whether the modules above the cut-off value were more at risk to be modified for enhancement or fixes than modules below the cut-off value, the transitions between the releases were examined. The results appear in Table 15. Of the 33 modules over the cut-off value of RCM in OP-B, 40 percent were actually modified in order to implement OP-C. Of the 36 modules in OP-C over OP-C's RCM cut-off value, 50 percent were actually modified to implement the draft version of OP-D.

Although the cut-off value seems to evenly divide the modules that were actually modified, the modules over the cut-off value for each release were more likely to be changed than the modules below the cut-off value. The RCM was, therefore, able to identify maintenance-prone modules.

III. Discussion and Conclusion

Given that a metric that measures software complexity should prove to be a useful predictor of software maintenance costs, it is recommended that modules that show a high order of complexity within a release be looked upon as modules with a propensity to become maintenance prone after release and delivery to users. It is imperative that a maintenance-prone module be improved, enhanced, or simplified into two or more modules before final delivery. The composite complexity models and the relative complexity metric developed in this study can be considered as a baseline for comparison with other projects and may serve as a set point for simplifying and reducing complexity of developed software.

Acknowledgments

The authors would like to express their sincere thanks to Dr. Harry Detweiler, Manager, Dr. William J. Hurd, Deputy Manager, and Paul A. Willis, Supervisor, Radio Frequency and Microwave Subsystems Section; Dr. Robert C. Tausworthe, Chief Technologist, Information Systems Division; and Dr. Donald S. Remer, Telecommunications and Data Acquisition Planning, for comments and suggestions that greatly improved this article.

References

[1] V. R. Basili and D. H. Hutchins, "An Empirical Study of a Synthetic Complexity Family," *IEEE Transactions on Software Engineering*, vol. 9, no. 6, pp. 664–672, November 1983.

[2] L. H. Binder and J. H. Poore, "Field Experiments With Local Software Quality Metrics," *Software Practice and Experience*, vol. 20, no. 7, pp. 631–647, July 1990.

[3] B. Boehm, *Software Engineering Economics*, Englewood Cliffs, New Jersey: Prentice Hall, 1981.

[4] B. Boehm and P. Papaccio, "Understanding and Controlling Software Costs," *IEEE Transactions on Software Engineering*, vol. 14, no. 10, pp. 1462–1477, October 1988.

[5] R. N. Charette, *Software Engineering Environment*, New York: McGraw–Hill, Inc., 1986.

[6] Comptroller General, *Contracting for Computer Software Development*, General Accounting Office Report, FGMSD-80-4, GAO, 1979.

[7] B. Curtis, S. Sheppard, P. Milliman, M. Borst, and T. Love, "Measuring the Psychological Complexity of Software Maintenance Tasks With the Halstead and McCabe Metrics," *IEEE Transactions on Software Engineering*, vol. 5, pp. 96–104, March 1979.

[8] S. Dekleva, "Software Maintenance: Any News Besides the Name," *The Software Practitioner*, vol. 3, no. 3, pp. 5–8, March 1993.

[9] W. M. Evangelist, "Software Complexity Metric Sensitivity to Program Structure Rules," *Journal of Systems and Software*, vol. 3, no. 3, pp. 231–243, March 1983.

[10] R. E. Fairley, *Software Engineering Concepts*, New York: McGraw–Hill, Inc., 1985.

[11] V. R. Gibson and J. A. Senn, "System Structure and Software Maintenance Performance," *Communications ACM*, vol. 32, no. 3, pp. 347–358, March 1989.

[12] G. K. Gill and C. F. Kemerer, "Cyclomatic Complexity Density and Software Maintenance Productivity," *IEEE Transactions on Software Engineering*, vol. 17, no. 12, pp. 1284–1288, December 1991.

[13] R. L. Glass, *Software Maintenance Handbook*, Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1981.

[14] M. Halstead, *Elements of Software Science*, New York: Elsevier North Holland, Inc., 1977.

[15] W. Harrison and C. Cook, "A Micro/Macro Measure of Software Complexity," *The Journal of Systems and Software*, vol. 7, no. 2, pp. 213–219, August 1987.

[16] W. Harrison and C. Cook, *Insights on Improving The Maintenance Process Through Software Measurements*, Naval Ocean Systems Center Report TR 90-4, N66001-87-D-0136, 1990.

[17] D. Kafura and G. R. Reddy, "The Use of Software Complexity Metrics in Software Maintenance," *IEEE Transactions on Software Engineering*, vol. 13, no. 13, pp. 335–343, March 1987.

[18] B. Lennselius, C. Wohlin, and C. Vrana, "Software Metrics: Fault Content Estimation and Software Process Control," *Microprocessors and Microsystems*, vol. 11, no. 7, pp. 365–375, September 1987.

[19] B. P. Lientz and E. B. Swanson, *Software Maintenance Management*, Reading, Massachusetts: Addison–Wesley, 1990.

[20] T. J. McCabe, "A Complexity Measure," *IEEE Transactions on Software Engineering*, vol. 2, no. 4, pp. 308–320, December 1976.

[21] J. C. Munson and T. M. Khoshgoftaar, "Application of a Relative Complexity Metric for Software Project Management," *Journal of Systems and Software*, vol. 12, no. 3, pp. 283–291, July 1990.

[22] B. Ramamurthy and A. Melton, "A Synthesis of Software Sciences Measures and the Cyclomatic Number," *IEEE Transactions on Software Engineering*, vol. 14, no. 8, pp. 1116–1121, August 1988.

[23] V. Rodriguez and W. T. Tsai, "Evaluation of Software Metrics Using Discriminant Analysis," *Proceedings of the Eleventh Annual International Computer Software and Applications Conference*, Tokyo, Japan, pp. 245–251, October 1987.

[24] H. D. Rombach, "A Controlled Experiment on the Impact of Software Structure on Maintainability," *IEEE Transactions on Software Engineering*, vol. 13, no. 3, pp. 344–354, March 1987.

[25] M. H. Samadzadeh and K. Nandakumar, "A Study of Software Metrics," *Journal of Systems Software*, vol. 16, no. 3, pp. 229–234, November 1991.

[26] N. F. Schneidewind, "Methodology For Validating Software Metrics," *IEEE Transactions on Software Engineering*, vol. 18, no. 5, pp. 410–422, May 1992.

[27] V. Y. Shen, T. Yu, S. M. Thebaut, and L. R. Paulsen, "Identifying Error-Prone Software—An Empirical Study," *IEEE Transactions on Software Engineering*, vol. 11, no. 4, pp. 317–323, April 1985.

[28] Y. S. Sherif, E. Ng, and J. Steinbacher, "Computer Software Development: Quality Attributes, Measurements and Metrics," *Naval Research Logistics*, vol. 35, no. 1, pp. 425–436, January 1988.

[29] T. Stalhane, *A Discussion of Software Metrics as a Means for Software Reliability Evaluation*, Report PB89-210322, U.S. Department of Commerce, National Technical Information Service, 1988.

[30] J. D. Valett and F. E. McGarry, "A Summary of Software Measurement Experiences in the Software Engineering Laboratory," *The Journal of Systems and Software*, vol. 9, no. 2, pp. 137–148, February 1989.

[31] I. Vessey and R. Weber, "Some Factors Affecting Program Maintenance: An Empirical Study," *Communications ACM*, vol. 26, no. 2, pp. 128–134, February 1983.

[32] S. Wake and S. Henry, "A Model Based on Software Quality Factors Which Predicts Maintainability," *Proceedings of the Conference on Software Maintenance*, Phoenix, Arizona, pp. 382–387, October 24, 1988.

[33] S. S. Yau and J. S. Collofello, "Some Stability Measures for Software Maintenance," *IEEE Transactions on Software Engineering*, vol. 6, no. 6, pp. 545–552, November 1980.

[34] M. V. Zelkowitz, A. C. Shaw, and J. D. Grannon, *Principles of Software Engineering and Design*, Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1979.

Table 1. Percentage of time spent on various maintenance tasks.

Maintenance tasks	Percentage of time spent			
	1977	1985	1987	1990
Enhancements	59	44	41	43
Corrections	22	15	18	16
Supporting users	NA[a]	21	12	12
Reengineering	NA	NA	10	9
Adaptations	6	8	9	8
Documentation	6	NA	5	6
Tuning	4	NA	3	5
Evaluating requests	NA	8	NA	NA
Other	3	4	2	1

[a] Not applicable.

Table 2. Comptroller General statistics on delivered software.

Quality of software delivered	Percentage of software delivered
Could work on delivery	2
Could work after some rework	3
Never successfully put to use	45
Extensively reworked	20
Useless	30
Total	100

Table 3. Software measures used to analyze the VRC software.

Measure number	Measure	Measure definition
1	n_1	Number of unique operators
2	n_2	Number of unique operands
3	N_1	Number of total operators
4	N_2	Number of total operands
5	N	Length $(N_1 + N_2)$
6	$\hat{N}$	Estimated length = $[n_1(\log_2(n_1)) + n_2(\log_2(n_2))]$
7	V	Volume = $(N)\log_2(n) = (N_1 + N_2)\log_2(n_1 + n_2)$
8	E	Effort = $V/[(2/n_1)(n_2/N_2)]$
9	VG_1	McCabe cyclomatic complexity (number of decisions + 1)
10	VG_2	Extended complexity (decisions + ANDs + ORs +1)
11	LOC	Lines of code (includes blank and comment lines)
12	B/C	Number of blank lines + number of comment lines
13	$<;>$	Number of executable semicolons
14	SP	Average maximum lines between variable references
15	$NCLOC$	Non-commented lines of code = $LOC - B/C$
16	LB-ratio	$[N_1/(N_2 + B/C)]$

Table 4. OP-B, OP-C, and OP-D modules and the mean values of the 16 measures.

Measure number	Measure	OP-B (222 modules) mean	OP-C (224 modules) mean	OP-D (235 modules) mean
1	n_1	12	12	13
2	n_2	12	12	15
3	N_1	70	75	87
4	N_2	42	44	52
5	N	113	119	140
6	$\hat{N}$	103	110	126
7	V	704	721	844
8	E	53,781	58,198	61,715
9	VG_1	4	4	5
10	VG_2	5	4	5
11	LOC	73	78	83
12	B/C	43	46	49
13	$<;>$	12	13	15
14	SP	5	5	6
15	$NCLOC$	30	31	34
16	LB-ratio	1	1	1

Table 5. The factor matrix for the 16 measures of OP-C, OP-B, and OP-D.

Measure number	Measure	OP-B		OP-C		OP-D	
		Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
1	n_1	0.78	−0.17	0.79	−0.12	0.78	−0.17
2	n_2	0.94	−0.02	0.94	−0.02	0.93	−0.03
3	N_1	0.97	0.10	0.98	0.83	0.97	0.08
4	N_2	0.97	0.06	0.97	0.04	0.96	−0.05
5	N	0.98	0.09	0.98	0.07	0.97	0.07
6	$\hat{N}$	0.91	−0.01	0.96	−0.00	0.96	−0.01
7	V	0.96	0.14	0.97	0.09	0.96	0.09
8	E	0.89	0.22	0.90	0.15	0.88	0.15
9	VG_1	0.94	0.09	0.95	0.08	0.93	0.10
10	VG_2	0.77	0.12	0.95	0.07	0.93	0.10
11	LOC	0.94	−0.25	0.96	−0.17	0.95	−0.19
12	B/C	0.61	−0.64	0.72	−0.50	0.70	−0.53
13	$<;>$	0.97	0.03	0.97	0.04	0.97	0.06
14	SP	0.70	−0.05	0.60	−0.01	0.72	0.04
15	$NCLOC$	0.98	0.05	0.98	0.05	0.98	0.05
16	LB-ratio	−0.03	0.83	−0.01	0.92	−0.02	0.90
Percentage of explained variability		90	10	91	9	91	9

Table 6. Five composite complexity models and their coefficients of determination.

Model number	Model	Coefficient of determination, percent
1	$<VG_1> = 1.48 + 0.005(V)$	$R^2 = 96$
2	$<VG_1> = 0.510 + 0.136(NCLOC)$	$R^2 = 96$
3	$<VG_1> = 0.786 + 0.0013(V) + 0.0976(NCLOC)$	$R^2 = 96$
4	$<V> = -206 + 29.5(NCLOC)$	$R^2 = 99$
5	$<V> = -210 + 8.7(VG_1) + 28.3(NCLOC)$	$R^2 = 99$

Table 7. Summary of actual average values of (VG_1) and values predicted by models 1, 2, and 3.

Model	Release	(V) value		Delta, $(A) - (P)$	Error percentage, delta/(A)
		Actual, (A)	Predicted, (P)		
1	OP-B	4.45	5.00	−0.55	−12.40
	OP-C	4.53	5.09	−0.56	−12.40
	OP-D	5.30	5.70	−0.40	−7.50
Grand average		4.76	5.26	−0.50	−10.60
2	OP-B	4.45	4.59	−0.14	−3.10
	OP-C	4.53	4.86	−0.33	−7.30
	OP-D	5.30	5.27	−0.03	0.60
Grand average		4.76	4.91	−0.15	−3.10
3	OP-B	4.45	4.62	−0.17	−3.80
	OP-C	4.53	4.84	−0.31	−6.80
	OP-D	5.30	5.30	−0.00	0.00
Grand average		4.76	4.92	−0.16	−3.40

Table 8. Summary of actual average values of (V) and values predicted by models 4 and 5.

Model	Release	(V) value		Delta, $(A) - (P)$	Error percentage, delta/(A)
		Actual, (A)	Predicted, (P)		
4	OP-B	704	679	+25	+3.6
	OP-C	722	738	−16	−2.2
	OP-D	845	826	+19	+2.2
Grand average		757	748	+9	+1.2
5	OP-B	704	678	+26	+3.7
	OP-C	722	735	−13	−1.8
	OP-D	845	826	+19	+2.2
Grand average		757	746	−10	+1.3

Table 9. Characteristics of four independent segments of software.

Segment number	Number of modules	Actual average value		
		VG_1	V	$NCLOC$
1	16	16.4	3343	102
2	16	17.9	4016	139
3	50	8.16	1823	64
4	55	11.10	2212	71

Table 10. Sample calculation of actual average values of (VG_1) and values predicted by model 1 for segments 1–4.

Model	Segment	(V) value		Delta, $(A) - (P)$	Error percentage, delta/(A)
		Actual, (A)	Predicted, (P)		
1	1	16.40	18.19	−1.79	−10.9
	2	17.90	21.56	−3.66	−20.4
	3	8.16	10.59	−2.03	−24.4
	4	11.10	12.54	−1.44	−13.0
Grand average		13.39	15.72	−2.33	−17.3

Table 11. Summary of actual grand average values of (VG_1) and values predicted by models 1, 2, and 3 for segments 1–4.

Model	Segment	(VG_1) grand average value		Delta, $(A) - (P)$	Error percentage, delta/(A)
		Actual, (A)	Predicted, (P)		
1	1–4	13.39	15.57	−2.33	−17.3
2	1–4	13.39	13.31	+0.08	+0.6
3	1–4	13.39	13.48	−0.09	+0.7

Table 12. Summary of actual grand average values of (V) and values predicted by models 4 and 5 for segments 1–4.

Model	Segment	(VG_1) grand average value		Delta, $(A) - (P)$	Error percentage, delta/(A)
		Actual, (A)	Predicted, (P)		
4	1–4	2848	2570	+278	+9.7
5	1–4	2848	2571	+277	+9.7

Table 13. Analysis of three software releases using the relative complexity metric.

Release	Total number of modules	Relative complexity					
		Total	Maximum	Minimum	Median	Mean	Standard deviation
OP-B	222	2799	45	0.4	10.9	12.6	10.0
OP-C	224	2837	45	0.4	10.9	12.7	9.6
OP-D	235	3470	49	0.4	12.2	14.8	11.3

Table 14. Cut-off values of the three software releases.

Release	Total number of modules	(RCM) cut-off value	Number of modules exceeding (RCM) cut-off value	Percentage of modules over (RCM) cut-off value
OP-B	222	22.6	33.0	15.0
OP-C	224	22.3	36.0	16.0
OP-D	235	26.1	35.0	15.0

Table 15. Analysis of transitions between the three software releases.

Transition	Number of modules modified	(RCM) cut-off value	Percentage of modified modules over cut-off value	Percentage of all modules over cut-off value that were actually modified
From OP-B to OP-C	13	22.6	46	40
From OP-C to OP-D	38	22.3	47	50

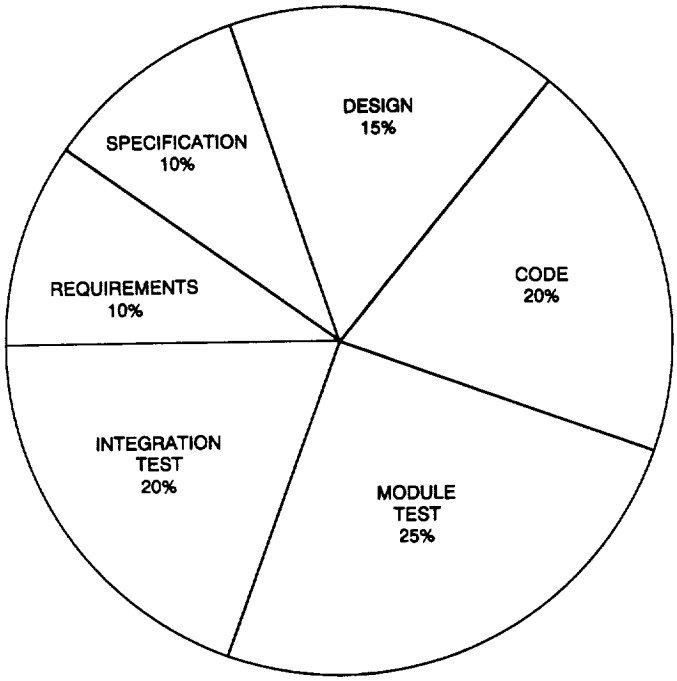


Fig. 1. The initial cost breakdown in developing a new project.

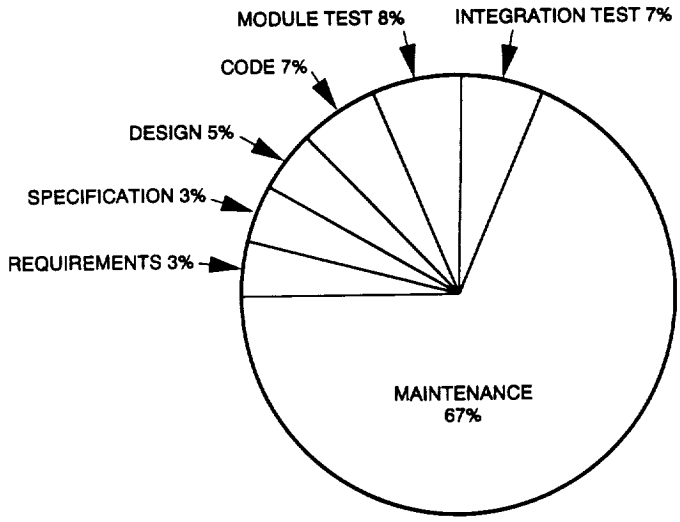


Fig. 2. The cost of software during its life cycle.

	B1	B2	N1	N2	N	NV	V	E	VG1	VG2	LOC	BC	CR	SP	NLOC	LBR
B1	1.00000 0.0	0.77205 0.0001	0.70872 0.0001	0.68055 0.0001	0.70054 0.0001	0.72123 0.0001	0.65058 0.0001	0.54830 0.0001	0.74993 0.0001	0.56794 0.0001	0.75723 0.0001	0.59043 0.0001	0.76131 0.0001	0.71629 0.0001	0.73042 0.0001	-0.03466 0.6075
B2	0.77205 0.0001	1.00000 0.0	0.91155 0.0001	0.93063 0.0001	0.92233 0.0001	0.96182 0.0001	0.91273 0.0001	0.78438 0.0001	0.83249 0.0001	0.73144 0.0001	0.87705 0.0001	0.58032 0.0001	0.93684 0.0001	0.63632 0.0001	0.91476 0.0001	-0.03778 0.5755
N1	0.70872 0.0001	0.91155 0.0001	1.00000 0.0	0.98580 0.0001	0.99801 0.0001	0.90913 0.0001	0.98176 0.0001	0.95128 0.0001	0.93309 0.0001	0.71636 0.0001	0.89288 0.0001	0.51917 0.0001	0.95732 0.0001	0.62622 0.0001	0.97883 0.0001	0.01953 0.7723
N2	0.68055 0.0001	0.93063 0.0001	0.98580 0.0001	1.00000 0.0	0.99437 0.0001	0.93243 0.0001	0.97363 0.0001	0.90950 0.0001	0.89624 0.0001	0.71234 0.0001	0.91111 0.0001	0.56573 0.0001	0.95695 0.0001	0.61197 0.0001	0.97494 0.0001	-0.00079 0.9906
N	0.70054 0.0001	0.92233 0.0001	0.99801 0.0001	0.99437 0.0001	1.00000 0.0	0.92114 0.0001	0.98277 0.0001	0.93915 0.0001	0.92274 0.0001	0.72116 0.0001	0.90250 0.0001	0.53733 0.0001	0.96022 0.0001	0.62291 0.0001	0.98104 0.0001	0.01181 0.8611
NV	0.72123 0.0001	0.96182 0.0001	0.90913 0.0001	0.93243 0.0001	0.92114 0.0001	1.00000 0.0	0.90741 0.0001	0.79853 0.0001	0.76636 0.0001	0.64927 0.0001	0.85885 0.0001	0.57506 0.0001	0.88360 0.0001	0.47706 0.0001	0.89127 0.0001	-0.02969 0.6600
V	0.65058 0.0001	0.91273 0.0001	0.98176 0.0001	0.97363 0.0001	0.98277 0.0001	0.90741 0.0001	1.00000 0.0	0.95253 0.0001	0.91233 0.0001	0.79787 0.0001	0.86609 0.0001	0.47319 0.0001	0.93203 0.0001	0.58001 0.0001	0.96965 0.0001	0.01466 0.8281
E	0.54830 0.0001	0.78438 0.0001	0.95128 0.0001	0.90950 0.0001	0.93915 0.0001	0.79853 0.0001	0.95253 0.0001	1.00000 0.0	0.89454 0.0001	0.69025 0.0001	0.78596 0.0001	0.37656 0.0001	0.85869 0.0001	0.49117 0.0001	0.91505 0.0001	0.04023 0.5510
VG1	0.74993 0.0001	0.83249 0.0001	0.93309 0.0001	0.89624 0.0001	0.92274 0.0001	0.76636 0.0001	0.91233 0.0001	0.89454 0.0001	1.00000 0.0	0.78359 0.0001	0.86089 0.0001	0.48352 0.0001	0.92553 0.0001	0.76376 0.0001	0.95509 0.0001	0.01807 0.7889
VG2	0.56794 0.0001	0.73144 0.0001	0.71636 0.0001	0.71234 0.0001	0.72116 0.0001	0.64927 0.0001	0.79787 0.0001	0.69025 0.0001	0.78359 0.0001	1.00000 0.0	0.67531 0.0001	0.32942 0.0001	0.70748 0.0001	0.58539 0.0001	0.78232 0.0001	-0.02037 0.7628
LOC	0.75723 0.0001	0.87705 0.0001	0.89288 0.0001	0.91111 0.0001	0.90250 0.0001	0.85885 0.0001	0.86609 0.0001	0.78596 0.0001	0.86089 0.0001	0.67531 0.0001	1.00000 0.0	0.83221 0.0001	0.90148 0.0001	0.65696 0.0001	0.92972 0.0001	-0.15599 0.0201
BC	0.59043 0.0001	0.58032 0.0001	0.51917 0.0001	0.56573 0.0001	0.53733 0.0001	0.57506 0.0001	0.47319 0.0001	0.37656 0.0001	0.48352 0.0001	0.32942 0.0001	0.83221 0.0001	1.00000 0.0	0.56689 0.0001	0.43891 0.0001	0.56954 0.0001	-0.33967 0.0001
CR	0.76131 0.0001	0.93684 0.0001	0.95732 0.0001	0.95695 0.0001	0.96022 0.0001	0.88360 0.0001	0.93203 0.0001	0.85869 0.0001	0.92553 0.0001	0.70748 0.0001	0.90148 0.0001	0.56689 0.0001	1.00000 0.0	0.75125 0.0001	0.95988 0.0001	0.01323 0.8445
SP	0.71629 0.0001	0.63632 0.0001	0.62622 0.0001	0.61197 0.0001	0.62291 0.0001	0.47706 0.0001	0.58001 0.0001	0.49117 0.0001	0.76376 0.0001	0.58539 0.0001	0.65696 0.0001	0.43891 0.0001	0.75125 0.0001	1.00000 0.0	0.68239 0.0001	0.01060 0.8752
NLOC	0.73042 0.0001	0.91476 0.0001	0.97883 0.0001	0.97494 0.0001	0.98104 0.0001	0.89127 0.0001	0.96965 0.0001	0.91505 0.0001	0.95509 0.0001	0.78232 0.0001	0.92972 0.0001	0.56954 0.0001	0.95988 0.0001	0.68239 0.0001	1.00000 0.0	-0.00566 0.9332
LBR	-0.03466 0.6075	-0.03778 0.5755	0.01953 0.7723	-0.00079 0.9906	0.01181 0.8611	-0.02969 0.6600	0.01466 0.8281	0.04023 0.5510	0.01807 0.7889	-0.02037 0.7628	-0.15599 0.0201	-0.33967 0.0001	0.01323 0.8445	0.01060 0.8752	-0.00566 0.9332	1.00000 0.0

Fig. 3. Correlation matrix of 16 measures for OP-B.

	B1	B2	N1	N2	N	M	V	E	VG1	VG2	LOC	BC	CT	SP	NCLOC	LBR
B1	1.00000 0.0	0.79503 0.0001	0.70854 0.0001	0.68603 0.0001	0.70381 0.0001	0.85592 0.0001	0.66948 0.0001	0.54823 0.0001	0.75377 0.0001	0.75702 0.0001	0.74387 0.0001	0.62827 0.0001	0.77626 0.0001	0.68827 0.0001	0.72105 0.0001	-0.04119 0.5397
B2	0.79503 0.0001	1.00000 0.0	0.91747 0.0001	0.92764 0.0001	0.92585 0.0001	0.98457 0.0001	0.90915 0.0001	0.78541 0.0001	0.83939 0.0001	0.83961 0.0001	0.89238 0.0001	0.68225 0.0001	0.95180 0.0001	0.56033 0.0001	0.90922 0.0001	-0.01885 0.7790
N1	0.70854 0.0001	0.91747 0.0001	1.00000 0.0	0.98567 0.0001	0.99822 0.0001	0.92925 0.0001	0.99735 0.0001	0.95202 0.0001	0.93489 0.0001	0.93273 0.0001	0.93188 0.0001	0.64823 0.0001	0.95255 0.0001	0.49795 0.0001	0.98920 0.0001	0.02460 0.7143
N2	0.68603 0.0001	0.92764 0.0001	0.98567 0.0001	1.00000 0.0	0.99326 0.0001	0.91839 0.0001	0.98180 0.0001	0.90939 0.0001	0.89666 0.0001	0.89658 0.0001	0.93627 0.0001	0.67809 0.0001	0.95307 0.0001	0.49197 0.0001	0.97728 0.0001	0.00933 0.8896
N	0.70381 0.0001	0.92585 0.0001	0.99822 0.0001	0.99326 0.0001	1.00000 0.0	0.92967 0.0001	0.99510 0.0001	0.93890 0.0001	0.92500 0.0001	0.92352 0.0001	0.93758 0.0001	0.66283 0.0001	0.95701 0.0001	0.50224 0.0001	0.98868 0.0001	0.01943 0.7724
M	0.85592 0.0001	0.98457 0.0001	0.92925 0.0001	0.91839 0.0001	0.92967 0.0001	1.00000 0.0	0.91803 0.0001	0.81228 0.0001	0.88911 0.0001	0.88833 0.0001	0.89454 0.0001	0.66532 0.0001	0.96145 0.0001	0.61427 0.0001	0.92292 0.0001	-0.00182 0.9783
V	0.66948 0.0001	0.90915 0.0001	0.99735 0.0001	0.98180 0.0001	0.99510 0.0001	0.91803 0.0001	1.00000 0.0	0.96468 0.0001	0.92532 0.0001	0.92202 0.0001	0.92287 0.0001	0.63301 0.0001	0.94184 0.0001	0.46109 0.0001	0.98518 0.0001	0.02808 0.6760
E	0.54823 0.0001	0.78541 0.0001	0.95202 0.0001	0.90939 0.0001	0.93890 0.0001	0.81228 0.0001	0.96468 0.0001	1.00000 0.0	0.89747 0.0001	0.89059 0.0001	0.85331 0.0001	0.54267 0.0001	0.84510 0.0001	0.33267 0.0001	0.93731 0.0001	0.04231 0.5287
VG1	0.75377 0.0001	0.83939 0.0001	0.93489 0.0001	0.89666 0.0001	0.92500 0.0001	0.88911 0.0001	0.92532 0.0001	0.89747 0.0001	1.00000 0.0	0.99669 0.0001	0.89309 0.0001	0.60974 0.0001	0.92230 0.0001	0.65413 0.0001	0.95515 0.0001	0.02546 0.7047
VG2	0.75702 0.0001	0.83961 0.0001	0.93273 0.0001	0.89658 0.0001	0.92352 0.0001	0.88833 0.0001	0.92202 0.0001	0.89059 0.0001	0.99669 0.0001	1.00000 0.0	0.89196 0.0001	0.61315 0.0001	0.91863 0.0001	0.66041 0.0001	0.95136 0.0001	0.02426 0.7180
LOC	0.74387 0.0001	0.89238 0.0001	0.93188 0.0001	0.93627 0.0001	0.93758 0.0001	0.89454 0.0001	0.92287 0.0001	0.85331 0.0001	0.89309 0.0001	0.89196 0.0001	1.00000 0.0	0.87132 0.0001	0.91200 0.0001	0.50984 0.0001	0.95278 0.0001	-0.12536 0.0611
BC	0.62827 0.0001	0.68225 0.0001	0.64823 0.0001	0.67809 0.0001	0.66283 0.0001	0.66532 0.0001	0.63301 0.0001	0.54267 0.0001	0.60974 0.0001	0.61315 0.0001	0.87132 0.0001	1.00000 0.0	0.65911 0.0001	0.37603 0.0001	0.68116 0.0001	-0.31940 0.0001
CT	0.77626 0.0001	0.95180 0.0001	0.95255 0.0001	0.95307 0.0001	0.95701 0.0001	0.96145 0.0001	0.94184 0.0001	0.84510 0.0001	0.92230 0.0001	0.91863 0.0001	0.91200 0.0001	0.65911 0.0001	1.00000 0.0	0.64830 0.0001	0.95282 0.0001	0.01914 0.7757
SP	0.68827 0.0001	0.56033 0.0001	0.49795 0.0001	0.49197 0.0001	0.50224 0.0001	0.61427 0.0001	0.46109 0.0001	0.33267 0.0001	0.65413 0.0001	0.66041 0.0001	0.50984 0.0001	0.37603 0.0001	0.64830 0.0001	1.00000 0.0	0.52798 0.0001	0.02198 0.7435
NCLOC	0.72105 0.0001	0.90922 0.0001	0.98920 0.0001	0.97728 0.0001	0.98868 0.0001	0.92292 0.0001	0.98518 0.0001	0.93731 0.0001	0.95515 0.0001	0.95136 0.0001	0.95278 0.0001	0.68116 0.0001	0.95282 0.0001	0.52798 0.0001	1.00000 0.0	0.01062 0.8744
LBR	-0.04119 0.5397	-0.01885 0.7790	0.02460 0.7143	0.00933 0.8896	0.01943 0.7724	-0.00182 0.9783	0.02808 0.6760	0.04231 0.5287	0.02546 0.7047	0.02426 0.7180	-0.12536 0.0611	-0.31940 0.0001	0.01914 0.7757	0.02198 0.7435	0.01062 0.8744	1.00000 0.0

Fig. 4. Correlation matrix of 16 measures for OP-C.

	B1	B2	N1	N2	N	M	V	E	VG1	VG2	LOC	BC	CT	SP	NCLOC	LBR
B1	1.00000 0.0	0.76798 0.0001	0.69921 0.0001	0.67628 0.0001	0.69470 0.0001	0.83404 0.0001	0.65911 0.0001	0.54051 0.0001	0.70251 0.0001	0.70917 0.0001	0.75883 0.0001	0.65341 0.0001	0.73748 0.0001	0.70268 0.0001	0.72216 0.0001	-0.06819 0.2979
B2	0.76798 0.0001	1.00000 0.0	0.91318 0.0001	0.92727 0.0001	0.92209 0.0001	0.98371 0.0001	0.90476 0.0001	0.76771 0.0001	0.81731 0.0001	0.82739 0.0001	0.88841 0.0001	0.67125 0.0001	0.93009 0.0001	0.62854 0.0001	0.90618 0.0001	-0.02990 0.6484
N1	0.69921 0.0001	0.91318 0.0001	1.00000 0.0	0.98554 0.0001	0.99823 0.0001	0.93052 0.0001	0.99630 0.0001	0.94063 0.0001	0.89809 0.0001	0.89640 0.0001	0.91856 0.0001	0.62890 0.0001	0.94779 0.0001	0.62333 0.0001	0.97910 0.0001	0.02353 0.7197
N2	0.67628 0.0001	0.92727 0.0001	0.98554 0.0001	1.00000 0.0	0.99310 0.0001	0.92374 0.0001	0.98136 0.0001	0.89946 0.0001	0.86324 0.0001	0.86590 0.0001	0.92109 0.0001	0.65207 0.0001	0.94822 0.0001	0.61295 0.0001	0.96792 0.0001	0.00753 0.9086
N	0.69470 0.0001	0.92209 0.0001	0.99823 0.0001	0.99310 0.0001	1.00000 0.0	0.93185 0.0001	0.99426 0.0001	0.92832 0.0001	0.88867 0.0001	0.88864 0.0001	0.92391 0.0001	0.64142 0.0001	0.95138 0.0001	0.62209 0.0001	0.97908 0.0001	0.01824 0.7809
M	0.83404 0.0001	0.98371 0.0001	0.93052 0.0001	0.92374 0.0001	0.93185 0.0001	1.00000 0.0	0.92089 0.0001	0.80819 0.0001	0.85805 0.0001	0.86369 0.0001	0.90027 0.0001	0.66906 0.0001	0.93851 0.0001	0.67833 0.0001	0.92550 0.0001	-0.01502 0.8189
V	0.65911 0.0001	0.90476 0.0001	0.99630 0.0001	0.98136 0.0001	0.99426 0.0001	0.92089 0.0001	1.00000 0.0	0.95835 0.0001	0.88574 0.0001	0.88249 0.0001	0.90801 0.0001	0.61560 0.0001	0.93267 0.0001	0.58713 0.0001	0.97180 0.0001	0.02684 0.6823
E	0.54051 0.0001	0.76771 0.0001	0.94063 0.0001	0.89946 0.0001	0.92832 0.0001	0.80819 0.0001	0.95835 0.0001	1.00000 0.0	0.83132 0.0001	0.81862 0.0001	0.82410 0.0001	0.52488 0.0001	0.82351 0.0001	0.48411 0.0001	0.90390 0.0001	0.04260 0.5158
VG1	0.70251 0.0001	0.81731 0.0001	0.89809 0.0001	0.86324 0.0001	0.88867 0.0001	0.85805 0.0001	0.88574 0.0001	0.83132 0.0001	1.00000 0.0	0.99307 0.0001	0.86436 0.0001	0.55703 0.0001	0.92329 0.0001	0.78499 0.0001	0.94385 0.0001	0.02552 0.6971
VG2	0.70917 0.0001	0.82739 0.0001	0.89640 0.0001	0.86590 0.0001	0.88864 0.0001	0.86369 0.0001	0.88249 0.0001	0.81862 0.0001	0.99307 0.0001	1.00000 0.0	0.86403 0.0001	0.56309 0.0001	0.92001 0.0001	0.79418 0.0001	0.93942 0.0001	0.02379 0.7168
LOC	0.75883 0.0001	0.88841 0.0001	0.91856 0.0001	0.92109 0.0001	0.92391 0.0001	0.90027 0.0001	0.90801 0.0001	0.82410 0.0001	0.86436 0.0001	0.86403 0.0001	1.00000 0.0	0.86829 0.0001	0.90087 0.0001	0.64902 0.0001	0.94701 0.0001	-0.13903 0.0332
BC	0.65341 0.0001	0.67125 0.0001	0.62890 0.0001	0.65207 0.0001	0.64142 0.0001	0.66906 0.0001	0.61560 0.0001	0.52488 0.0001	0.55703 0.0001	0.56309 0.0001	0.86829 0.0001	1.00000 0.0	0.61517 0.0001	0.43323 0.0001	0.66295 0.0001	-0.33398 0.0001
CT	0.73748 0.0001	0.93009 0.0001	0.94779 0.0001	0.94822 0.0001	0.95138 0.0001	0.93851 0.0001	0.93267 0.0001	0.82351 0.0001	0.92329 0.0001	0.92001 0.0001	0.90087 0.0001	0.61517 0.0001	1.00000 0.0	0.75191 0.0001	0.96130 0.0001	0.01875 0.7749
SP	0.70268 0.0001	0.62854 0.0001	0.62333 0.0001	0.61295 0.0001	0.62209 0.0001	0.67833 0.0001	0.58713 0.0001	0.48411 0.0001	0.78499 0.0001	0.79418 0.0001	0.64902 0.0001	0.43323 0.0001	0.75191 0.0001	1.00000 0.0	0.69901 0.0001	0.02089 0.7500
NCLOC	0.72216 0.0001	0.90618 0.0001	0.97910 0.0001	0.96792 0.0001	0.97908 0.0001	0.92550 0.0001	0.97180 0.0001	0.90390 0.0001	0.94385 0.0001	0.93942 0.0001	0.94701 0.0001	0.66295 0.0001	0.96130 0.0001	0.69901 0.0001	1.00000 0.0	0.00643 0.9218
LBR	-0.06819 0.2979	-0.02990 0.6484	0.02353 0.7197	0.00753 0.9086	0.01824 0.7809	-0.01502 0.8189	0.02684 0.6823	0.04260 0.5158	0.02552 0.6971	0.02379 0.7168	-0.13903 0.0332	-0.33398 0.0001	0.01875 0.7749	0.02089 0.7500	0.00643 0.9218	1.00000 0.0

Fig. 5. Correlation matrix of 16 measures for OP-D.

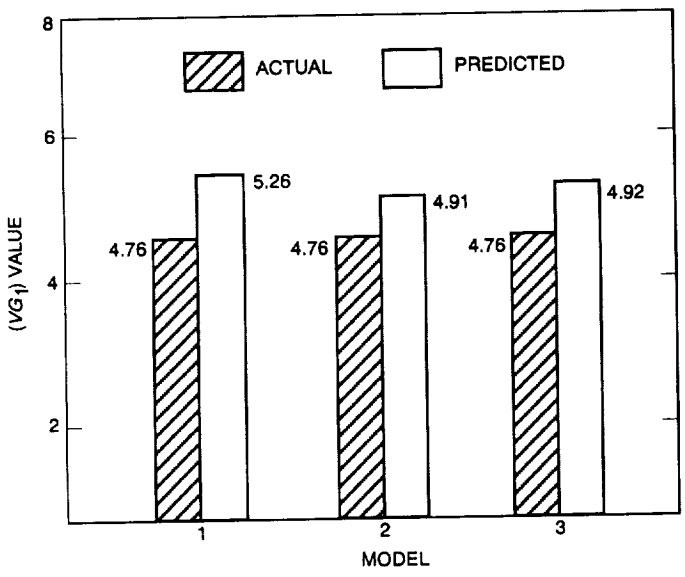


Fig. 6. Actual average values of (VG_1) and values predicted by models 1, 2, and 3.

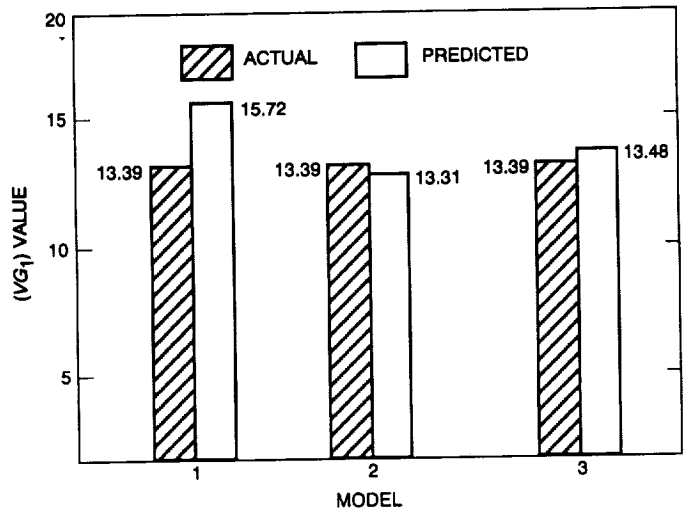


Fig. 8. Actual average values of (VG_1) and values predicted by models 1, 2, and 3 for independent segments of software.

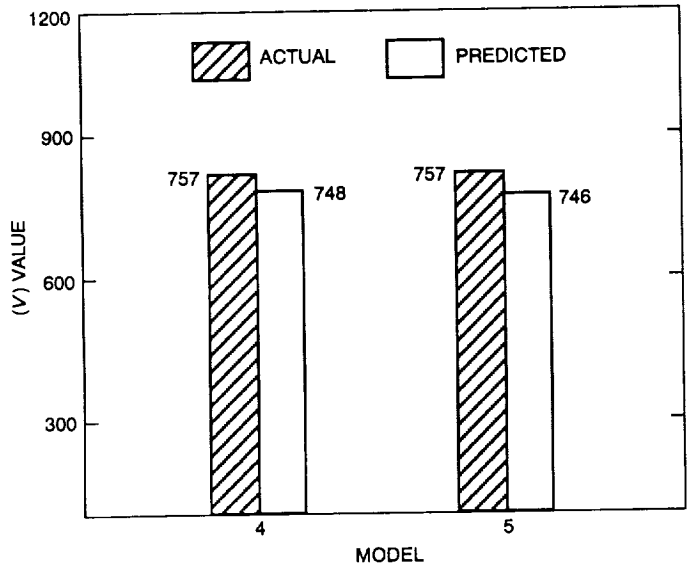


Fig. 7. Actual average values of (V) and values predicted by models 4 and 5.

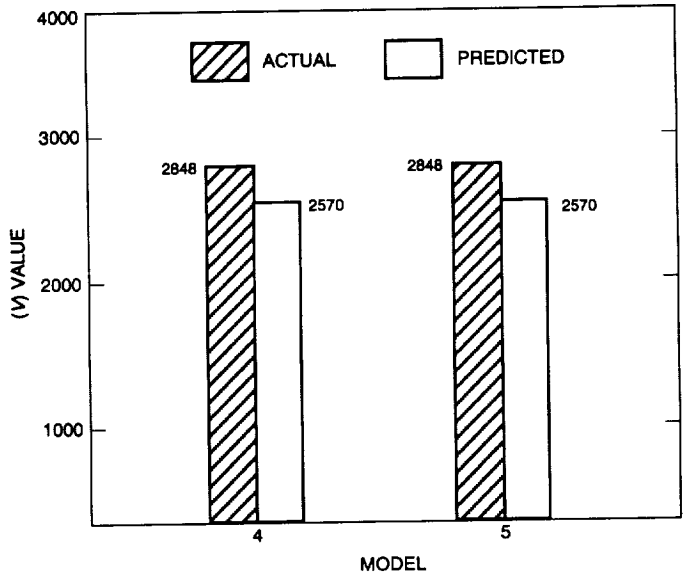


Fig. 9. Actual average values of (V) and values predicted by models 4 and 5 for independent segments of software.

The Real-Time Display of Interferometry Data for Goldstone Radar Astronomy Data Acquisition

S. D. Howard
Communications Systems Research Section

A method for visualizing radar interferometry data in real time has been developed for the Goldstone radar astronomy ranging data-acquisition system. The presentation is similar in appearance to a vector field display or data-based grid. This form was selected to facilitate the recognition of characteristic patterns of local variation in the phase and magnitude of complex elements in a two-dimensional data array. The design emphasized efficiency under the demands of real-time processing and remote monitoring. The interferometry "phase–magnitude" presentation, as it has come to be called, has been used to monitor radar interferometry experiments on three targets, beginning with the asteroid 4179 Toutatis, and continuing with Mars and Mercury.

I. Introduction

The Goldstone radar astronomy ranging system supports ground-based radar imaging. Radar echoes from the surface of a solar system target are mapped by range delay and Doppler frequency. However, delay–Doppler maps of ideal spherical objects are symmetrical about a central axis: the north and south elements are not resolvable without additional information. Radar interferometry, which combines data from more than one receiving station, resolves the north–south ambiguity by adding phase information to the delay and Doppler information. Interferometry products include not only images of surface reflectivity, but also root-mean-square surface roughness and polarization ratios. With a satisfactory signal-to-noise ratio (SNR) and sufficient data, altitude can be estimated as well.

Before November 1990, the interferometry data used for producing unambiguous radar images were derived during data reduction from recorded complex voltage data. In 1990, signal processing software devised for producing a new interferometry data type in real time was added to the data-acquisition system. The Appendix describes the interferometry or "crosspower" data type. Data of this type could be integrated without losing information, thus reducing data transfer and recording rates in the data-acquisition system and making radar imaging derived

from interferometry possible for the first time on high-bandwidth targets like Mars.

The 1990 Mars observations, the first using real-time interferometry signal processing, relied on postacquisition display to confirm that useful interferometry data had been acquired. The real-time display in use at the time could display only limited numbers of delay–Doppler power spectra; there were no means of displaying interferometry data. Even though the first Mars experiments collected useful data, there were significant advantages to being able to monitor data acquisition in real time. For example, if the range delays of echoes returning to each station in the interferometer had not been correctly aligned, the interferometry data would have been unusable. Early indications of incorrect station delay offsets or system malfunctions would have allowed station personnel to take corrective action during the observation.

The redesign of the software that manages the data-acquisition display created an opportunity to address the problem of displaying interferometry data in real time. The new software architecture was designed to be extensible. One of the purposes for the design of an extensible architecture was to provide a basis for experimentation with scientific data visualization. The development of the interferometry presentation was an experiment with data visualization and a demonstration that new display methods could be added to the real-time display software as planned. Research and development (R&D) is often associated with the beginning of the product life cycle. Since the ranging system is a major science instrument late in the life cycle, this experience also became an opportunity to experiment with the concept of R&D late in the product life cycle.

Several frequently occurring terms are often used in a specialized sense in this article. A "method" is an operation or transformation applied to a data type or object. This usage is intended to be close to the object-oriented usage. A "presentation" is an abstraction that describes how information will be transformed into a visual representation. A presentation may be rendered by hand or by computer. The term "display" is also often used to mean the real-time process in the data-acquisition computing system, implemented in software, that manages the display of incoming data.

II. Design Objectives and Constraints

There were several overall objectives for the design of the interferometry presentation. Minimizing the impact of the display on real-time performance was important, and so was retaining the capability for remote monitoring. Economical development and implementation were also primary objectives. Because the display would be introduced into a system in the maintenance phase of the life cycle, it was an objective to adapt the design to prevailing usage, with the intention of minimizing user familiarization time and errors. Both the display software and its platform were to be designed to be maintainable and reliable. Since it was also an objective to make the display available as much as possible, the implementing platform was constrained to low-cost, widely available hardware and software so that the physical display could be quickly replaced, if necessary.

A summary of design objectives related to the performance and function of the system is given in Table 1; design objectives related to the use of the system are listed in Table 2; design objectives derived from an analysis of the project situation are given in Table 3. These objectives shaped the synthesis of a final design. Candidate designs that were variations on presentations used at the time by either data reduction or data acquisition were set aside because they fell outside performance or usability objectives.

III. Design Description

A. Overview

The new interferometry presentation came to be called the phase–magnitude presentation for its emphasis on those dimensions of the data. An example is shown in Fig. 1. The graphic is an adaptation of a data-based grid [1]. Interferometry data are structured as a two-dimensional array of complex elements. The major dimensions are range (or delay) and Doppler frequency. Because each complex element also has two dimensions, the graphical problem became one of presenting four data dimensions on a two-dimensional area. The data graphic solved this problem by creating a two-dimensional array or grid of two-dimensional areas. A square area represents the complex plane with the origin at the center for each complex element. Complex elements are arranged in rows and columns for the two major dimensions of the data. Range is on the vertical axis, and Doppler frequency is on the horizontal axis. The structure of the data presentation is similar in form to the structure of the data.

The arrangement of vectors, number of vectors, size, and scaling was also defined by the design. To avoid the visual confusion caused by crossing vectors, the data graphic was designed so that vector areas would not overlap. During the design process, it was realized that only a sub-

set of the data needed to be shown. The design balances the number of vectors that can be displayed on the screen against the size of the individual vectors. Because only a subset of the data can be seen in the data graphic area, the user has the option of selecting the area of interest. The maximum component of any vector appearing in the display area is used for scaling. Since scaling is affected by the maximum complex component value, the selection of the data area also affects scaling.

B. Details

The text presentation of radar configuration information in the first and last lines is shared with all other presentations in the new real-time display software. These standard text areas, showing time, target, and configuration, form a common visual framework for all presentations in the display set. Only the fundamental parameters are shown; other parameters can be derived from these to establish the configuration. To save screen space and display time, only the lower and left axes are shown. The ratio of the sides of the data area is 32/20 units, close to the ratio of the Golden Rectangle. Visual balance and aesthetics were considered in the definition of the presentation.

Color was used to emphasize some graphical elements and deemphasize others. For instance, transmit and receive range code offsets, critical to the management of the real-time system, are shown in red on the display screen for emphasis whenever they are active. They are also clustered at the top center of the presentation for the same reason. Axis labels and tick marks are displayed in cyan, a light graph-paper blue, to create a mild illusion of space between the data graphic and the closely surrounding text.

C. Interpretation

The data-based grid was employed in this design because it emphasizes patterns of local variation in data. In interferometry data, there are two significant features. The most noticeable is an abrupt change in magnitude that marks the return of the radar echo. The most important from the standpoint of interferometry is the appearance of phase fringes, characteristic local patterns of phase change. For the presentation to be useful, viewers should be able to recognize patterns in the data when significant relationships in phase or magnitude exist. When no important relationships between data elements are present, phase and magnitude dimensions of data elements in the graphic should appear to be unrelated.

IV. Early Results

The evaluation of the design took place in three stages. In the first stage, receiver noise was used to evaluate viewer response; an illustration of this is given in Fig. 2. Viewers saw noise as random. Performance, function, and usability were found to be well within design goals. Additional system testing with loop-back data simulating point-source signals would not have been adequate for evaluating the complex phase relationships that were likely to be received from a planetary target. The new display software architecture had anticipated this situation. It had been designed to accommodate the playback of recorded data. Since no interferometry observations were scheduled for several months, playback became the best option for gaining familiarity with the handling of the new display. The second stage began when existing data sets were reviewed to learn what the data would look like when presented in this way, how much data integration would be necessary to see and interpret the results, and whether the policy of coupling the selection of the data area to scaling would be as acceptable as predicted by the design. The third stage, evaluation during an observation, assessed the impact of the display on real-time operations and instrument management. A summary of results from the first real-time observations is given in Table 4.

A. Playback

1. Mars 1990. Playback of the Mars 1990 data demonstrated that the phase patterns of usable interferometry data could be recognized in the new presentation. Figure 1 is an example. In the original review of the data with the data reduction display, it was not possible to see whether phase angles in phase bands changed gradually or abruptly, because the data-reduction display was a color display that associated phase angles in 30-deg intervals with one assigned hue. In the phase–magnitude presentation, since phase angles are shown directly, the gradual change in phase around the front of the planet is apparent.

It is an advantage for the viewer to be able to see phase fringes in the display with integration times less than the maximum. The review of playback data established that fringes could be seen with the new presentation at least as soon as they could be seen with the nonreal-time data-reduction display. Figure 1 was produced from data integrated for 5 sec. To give a good indication of phase fringes, the data-reduction review of the same data had been set to a 20-sec integration. This demonstrated that the integration required for effective data representation by the phase–magnitude presentation was within functional goals.

After playing back the first part of the Mars data set, there was some concern about whether it would be possible to see phase patterns farther back than the first two or three range gates on the planet, given the scaling policy initially implemented for this presentation. For example, in Fig. 1, the phasors representing early returns are near maximum length at center frequency gate 64 and in range gates 16 and 17; in range gate 21, the vectors have become points. This concern was resolved after it was noted that an unusually strong initial echo was present in the data. The display was an accurate representation of the given data.

2. Mercury 1992. Several attempts to acquire Mercury interferometry data were made in 1992. On one occasion, data containing echoes were recorded from both stations in the interferometer, but because of a low SNR from the second station, DSS 13, and uncertainty about whether or not both stations were correctly aligned in range, the value of the interferometry portion of the data was doubtful. Playback of the data with the new presentation did not show phase patterns. This is illustrated in Fig. 3. The area of increased magnitude is the returned echo. Phase appeared to become even more random when data integration time was increased. Increasing integration also caused an apparent decrease in the overall change in magnitude between noise and returned signal. Both of these consequences of increased integration indicated that usable interferometry data had not been recorded.

B. Real-Time

1. 4179 Toutatis. The first real-time experience with the ranging interferometry display took place during the observations of asteroid 4179 Toutatis in December 1992. Station offsets for correct range alignment were unknown when the Toutatis interferometry experiment began. Figure 4(a) is the delay–Doppler display showing the initial range alignment. As with the interferometry presentation, range is the vertical dimension and frequency is the horizontal. After the correct station offsets were included in the system, the delay–Doppler display in Fig. 4(b) shows the final range alignment in range gate 41 for both stations. Phase fringes were recognized as soon as the first phase–magnitude display was received; that display is shown in Fig. 5. Interferometry data points from the most distant lobe of Toutatis appear on the left side, between frequency gates 65 and 76.

The option to select regions of the data to be displayed was utilized during this observation. In the 0.5-μsec range resolution configuration used for Toutatis, a large system artifact was present at dc (frequency gate 64) in all range gates. When the artifact was shown in the data graphic, surrounding data values were scaled to points. Moving the data display area to the left or right side of dc allowed normal scaling to the maximum data value.

Some experience with discovering and correcting problems during an observation was also obtained when science and engineering personnel at Goldstone and JPL noticed a degradation in the quality of the displayed interferometry data after several minutes of acquisition. Operations personnel began a search for the problem and noticed that a polarization switch had been set incorrectly. The problem was corrected and, subsequently, recorded data were of acceptable quality.

2. Mars 1992–1993. The first interferometry experiments for the 1992–1993 Mars opposition began 3 days after the Toutatis experiment. The range alignment procedure was repeated for Mars because the range alignment offsets for the Toutatis configuration with submicrosecond range resolution were not correct for the Mars configuration, which had a 6.0-μsec range resolution. The investigator reported that phase fringes were seen in real time during the first experiment, confirming correct range alignment. Figure 6 is an example of data taken later in the Mars series.

In subsequent Mars experiments, episodes of low SNR caused more switching between the delay–Doppler power spectra display (the two-channel Toutatis delay–Doppler displays shown in Fig. 4 are examples of this type) and the new interferometry display than had been expected. Because the delay–Doppler display emphasizes magnitude only, it is the presentation of choice when the SNRs of both stations in the interferometer are low.

3. Mercury 1993. The Mercury experiment used the same range alignment offsets as the previous Mars experiments, so the range alignment procedure was not repeated or checked. Science and engineering staff once again exercised the remote monitoring capability by remotely monitoring the returns from Goldstone from an office at JPL. A low SNR from the second station in the interferometer required long integration times to see the target, longer than the round-trip light time. Because of the low SNR, the acquisition of usable interferometry had been doubtful. However, phase fringes were recognized immediately when the first display was received. Figure 7 shows the first display.

During the observation, a drop in SNR was noticed in the interferometry display. The loss of SNR was pursued at the station, but the situation could not be corrected before

the end of the observation. Figure 8 shows an interferometry display after the drop in SNR. Maximum magnitude, shown on the right side, above the data display area, is a factor of three lower than the maximum shown in Fig. 7; the apparent magnitude difference between noise and echo has decreased in the data graphic, and indications of phase patterns have disappeared.

The next series of observations, during a favorable inferior conjunction in July 1993, showed that phase bands could be seen at least as far back as 17 range gates, in contrast to only 5 range gates in the Mars data shown in Fig. 1. Additional experience will be necessary to see what effects features on the surface of Mercury have on phase patterns.

C. Phase Features

The ability to read and understand detail in the displays is still developing. One experimenter has created the terms "block" and "swirl" for phase features that can be seen in the phase–magnitude interferometry displays. For planetary targets where the target is large when compared to the fringe spacing, blocks and swirls correspond principally to variations in surface altitude. For small bodies, where the target is small relative to the fringe spacing, blocks, or large areas of equal phase, are the expected form. The Toutatis data in Fig. 5, between range gates 43 and 49 and frequency gates 65 and 76, can be thought of as a block. In the 1990 Mars data, an area that has been referred to as a swirl can be followed in Fig. 9 as it moves from right to left on the display. The swirl is first seen in Fig. 9(a), in the area between range gates 10 and 13 and frequency gates 72 and 80. In Fig. 9(b), it has moved to frequency gates 69 through 76, between range gates 9 and 12. In Fig. 9(c), the swirl has moved toward the front of the planet in the area of frequency gates 66 through 74.

V. Conclusions

The early results from the Toutatis, Mars, and Mercury experiments with the phase–magnitude presentation of interferometry data have demonstrated that the design is efficient for real-time use with remote monitoring and that it is effective for presenting interferometry data to viewers. The implicit objective of this development was to improve the management of the radar during observations, making the acquisition of radar interferometry data sets routine rather than sporadic. Experience with the first three targets, summarized in Table 4, indicates that this objective was met. Use of the phase–magnitude display has given radar scientists a different and precise view of phase features. Although the phase–magnitude presentation was not designed for this purpose, it is being considered for data-reduction work as well.

The design accommodated the objectives summarized in Tables 1 through 3. The emphasis was on economy in performance and implementation. The displays have been frequently used for remote monitoring at JPL over phone lines transmitting data at 9600 bps; they complete in less than 8 sec. The implementation utilized low-cost, widely available microcomputers running widely available terminal emulator packages. The software that supported simple graphics commands for the radar control computer was already available on the system. Design and development time took less than 3 weeks for one software engineer working part time. The software was written to be maintainable and reliable; however, it has not yet required maintenance or modification. The only requested change has been minor, a modification to the way in which the "max" parameter is computed for display annotation.

Because this experimental design is part of an adaptive maintenance activity, the design also incorporated patterns of current use. User selections were simple and similar in structure to selection parameters for other presentation types. The controls were planned to minimize familiarization time and user errors. Experience with all user groups suggested that the display was straightforward and easy to use. The products of the new presentation were also designed to be used in established ways. For example, the circumstances that motivated the goal described in Table 2, that the display translate well to black and white for hard copy, occurred shortly after the first real-time observation with the new display method, when printouts were requested and sent by facsimile to Arecibo Observatory in Puerto Rico. Circumstances again tested the usability of the presentation in black and white during an interval when the station display device was only capable of black and white output.

This design experiment also demonstrated that new display methods could be included in the real-time software while minimizing degradation to the display software structure. The new software architecture was intended to enable engineering experimentation as part of the research and development charter of radar astronomy while maintaining the reliability of a major science instrument. The design of the phase–magnitude presentation for interferometry data reflects both the experimentation and maintenance interests of the project, and suggests that although R&D work is often associated with the early phases of a project, experimentation can also bring new results in the maintenance phase of the life cycle.

Acknowledgments

The author thanks Paul Willis, Carl Franck, and Ray Jurgens. Paul Willis's comments, suggestions, and perseverance improved this article. Carl Franck and Ray Jurgens assisted during system testing. Ray Jurgens has also provided insights and helpful observations while working with and interpreting the new interferometry presentation.

Reference

[1] E. R. Tufte, *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983.

Table 1. Summary of design objectives derived from the system.

Analysis	Derived design objectives
Real-time system with finite capacity for additional processing.	Minimize computing and transmission overhead of graphics commands.
Remote system testing and monitoring of observations reduces project costs.	Retain remote monitoring capability.
Time to draw standard delay–Doppler display can exceed 30 sec.	Design new display to complete in less than 30 sec on a terminal emulator receiving data from the host at 9600 bps.
Data sent to display every n sec, where n varies with the configuration from subsecond intervals to minutes (an architectural constraint, costly to change).	Stay within the system architecture. Plan to accept data every n sec.

Table 2. Summary of design objectives derived from the users.

Analysis	Derived design objectives
Users frequently make black and white hard copy from printers, faxes, and photocopiers.	Design the data graphics in the presentation to map well to black and white, without losing information.
Conference papers and publications that publish early results often restrict the use of color graphics.	Same as above.
In the past, only monochrome displays have been used in the real-time system.	Begin transition from monochrome screen graphics to color. Use color for emphasis and focusing viewer attention, rather than as a means of encoding data.
Equipment available at Goldstone is more likely to support monochrome displays; suitable equipment to support color is more difficult to acquire.	Be able to fall back to monochrome with minimal software adaptation if equipment availability becomes a problem.
Color vision defect can cause misinterpretation if red and green elements are adjacent.	Avoid designs that could generate red and green adjacent elements.
Graphics showing raw data have been published in the past.	Design publication-quality graphics with attention to good aesthetics.
Past displays omitted time, target, and basic parameters from which the configuration could be reconstructed.	Design within the constraints of the revised display process; include this supporting information in the display.
Observations are episodic; operators must often refamiliarize before each series.	Minimize the number of parameters that must be learned by operators to use the display.
Changes in the system can cause an increase in operator errors during initial uses.	Minimize the amount of adaptation that the operators must make to use the new display. Keep the system familiar.

Table 3. Summary of design objectives derived from the task.

Analysis	Derived design objectives
Front-end digital signal processing hardware undergoing modification.	Limit the scope of the presentation design and implementation.
Software development computer and file system to be removed or relocated.	Same as above.
One software engineer available part-time for development.	Same as above. Plan for development time of less than 1 workmonth.
Good software development practice has improved the reliability and maintainability of the software.	Use good software practice. Design for robustness, defect prevention, and maintainability.

Table 4. Interferometry results for the first three interferometry targets.

Radar target	Date, Universal Time	Significance	Phase fringes observed	Summary
Asteroid 4179 Toutatis	December 13, 1992	First opportunity to use new interferometry display. Opportunity for first Goldstone real-time ranging interferometry display.	Yes	Fringes seen immediately on receipt of initial display. (See Fig. 5.) Used delay–Doppler display for visual range alignment to determine correct range offsets. Positioned display off center to avoid dc system artifact.
Mars	December 15, 1992	First opportunity to use new interferometry display on Mars. Second interferometry target attempted with new display.	Yes	Reports of phase fringes from investigator.
Mercury	March 20, 1993	First opportunity to use new interferometry display on Mercury. Third interferometry target attempted with new display.	Yes	Fringes seen immediately on receipt of initial display. (See Fig. 7.) SNR dropped later; suspected station problem. No opportunity to look for influence of topography on phase patterns because of low SNR.

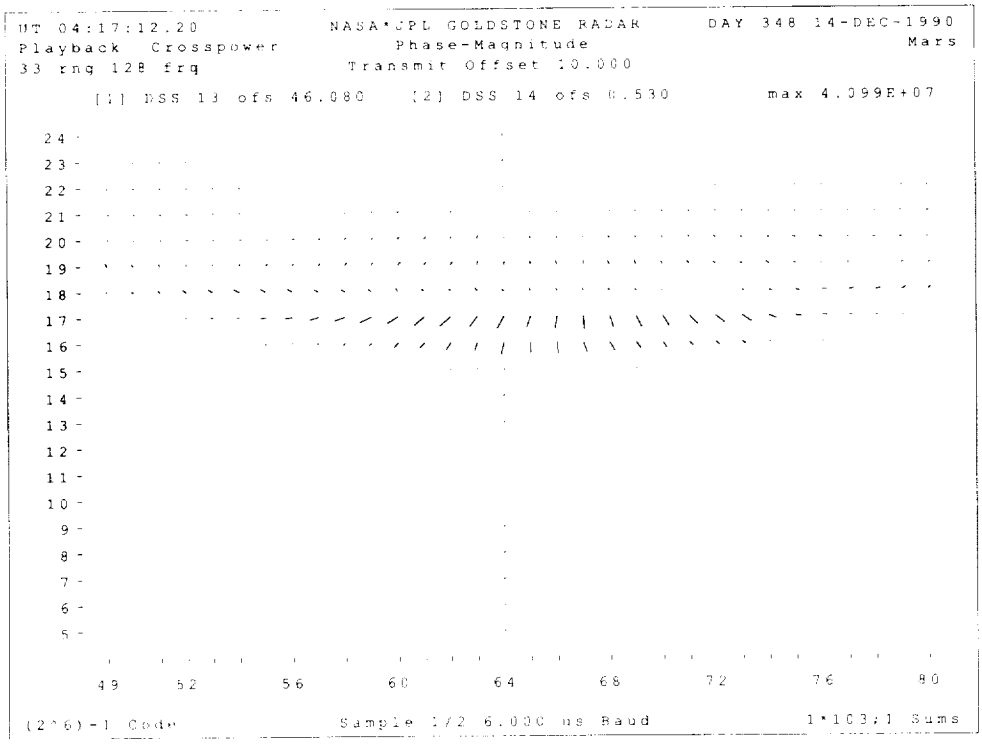


Fig. 1. The 1990 Mars interferometry playback.

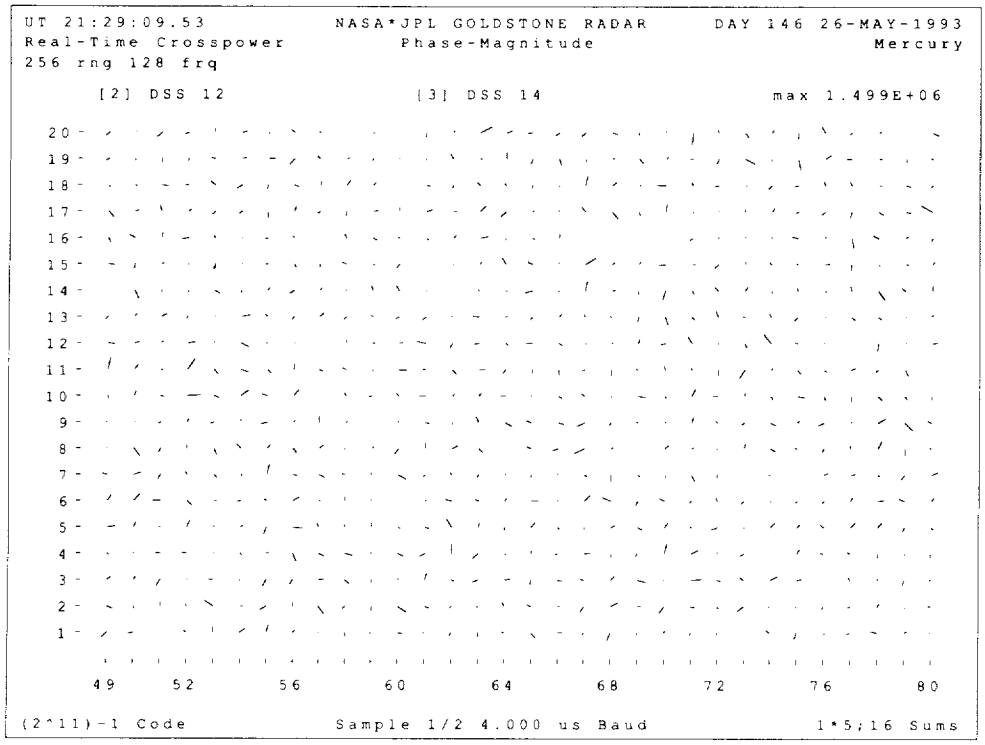


Fig. 2. Receiver noise.

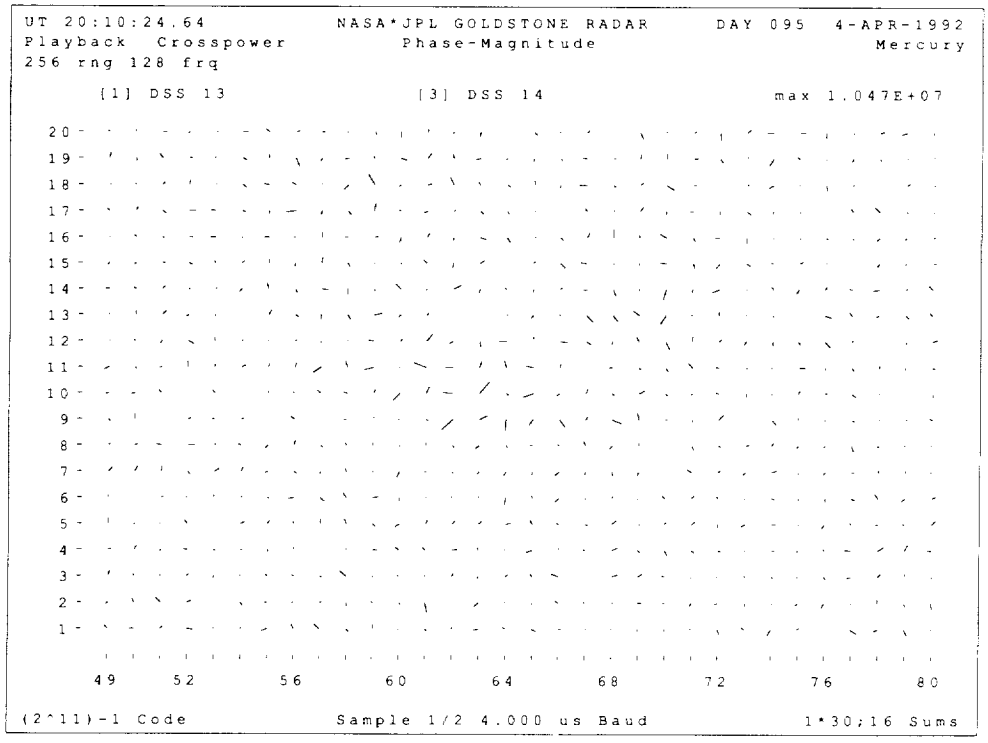
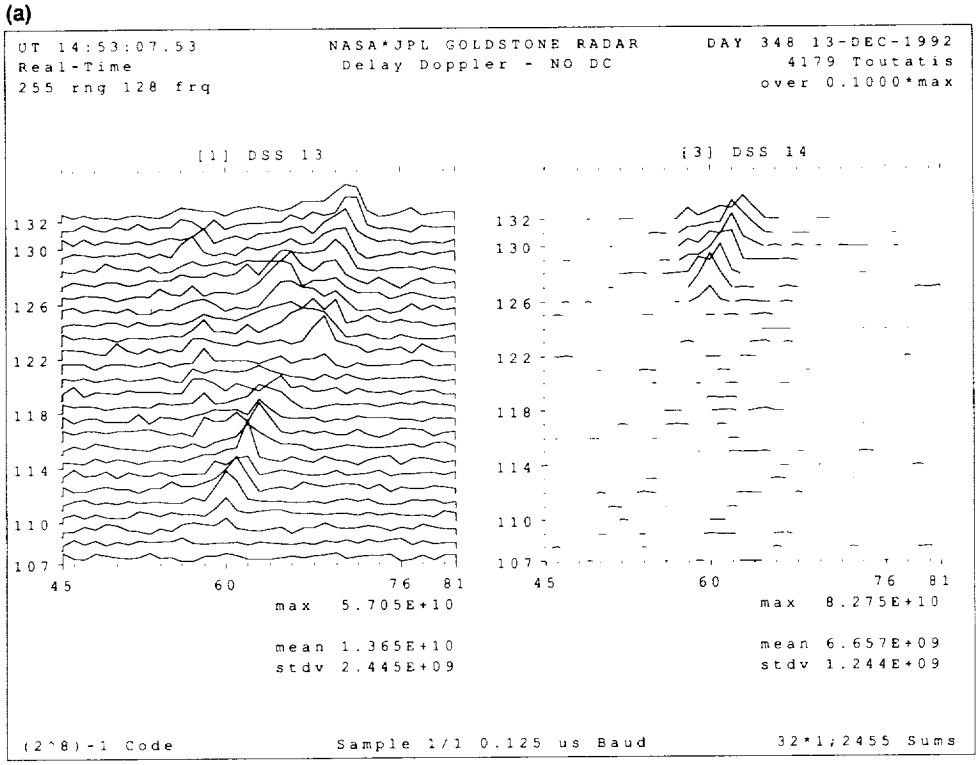


Fig. 3. The 1992 Mercury interferometry playback.

(a)



(b)

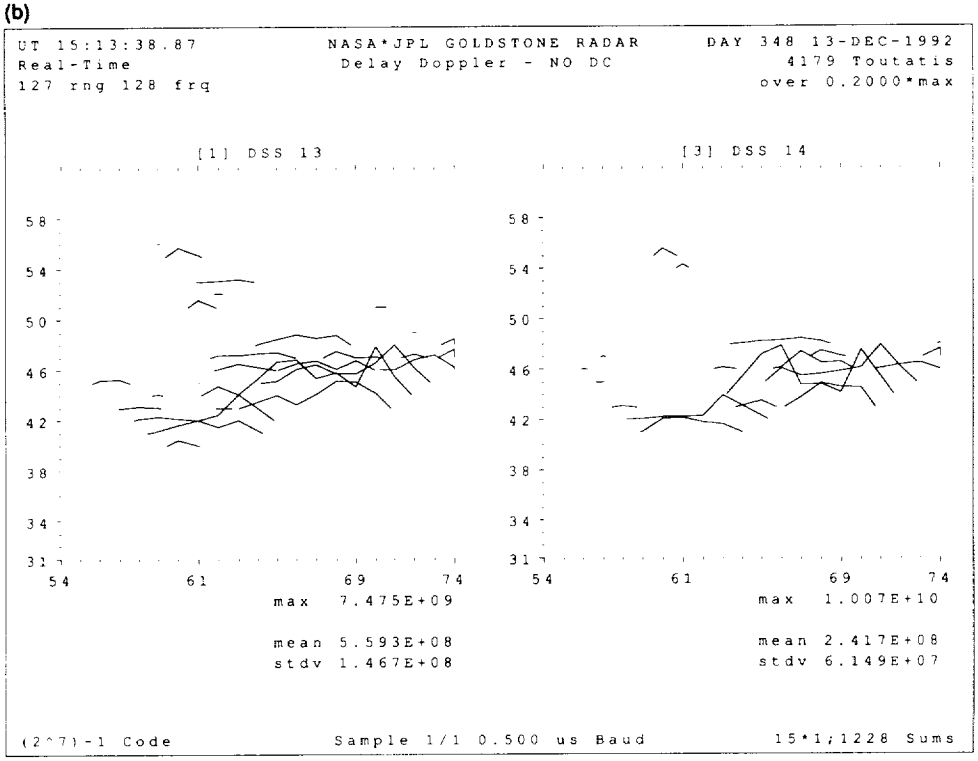


Fig. 4. Delay–Doppler displays of 4179 Toutatis (a) before range alignment and (b) after range alignment.

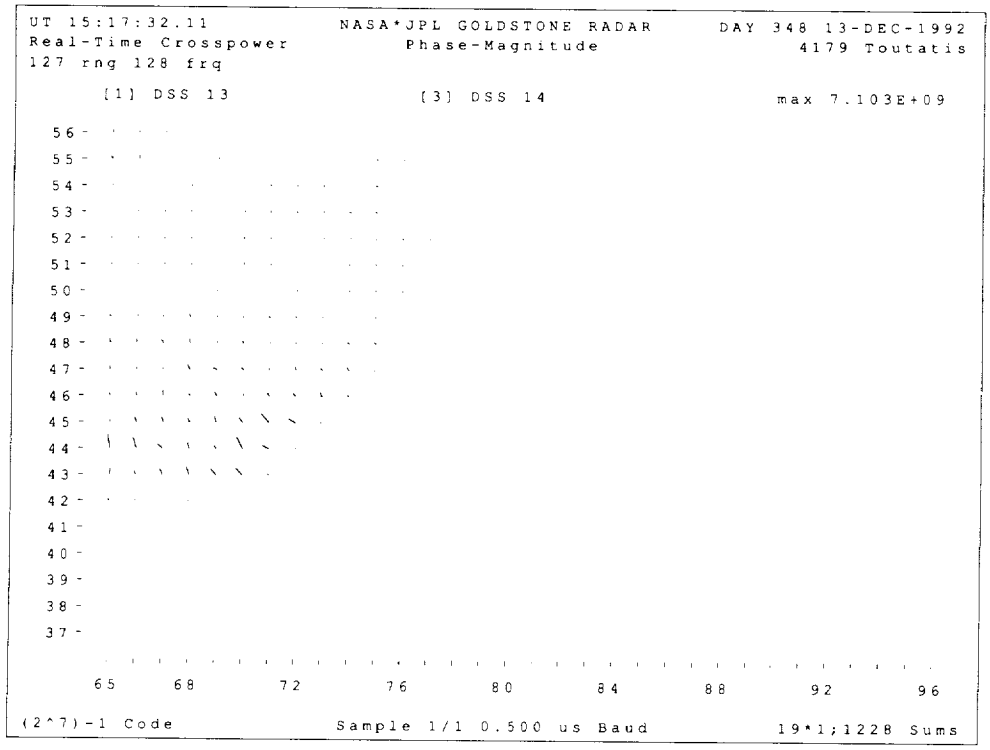


Fig. 5. The first real-time interferometry display: asteroid 4179 Toutatis.

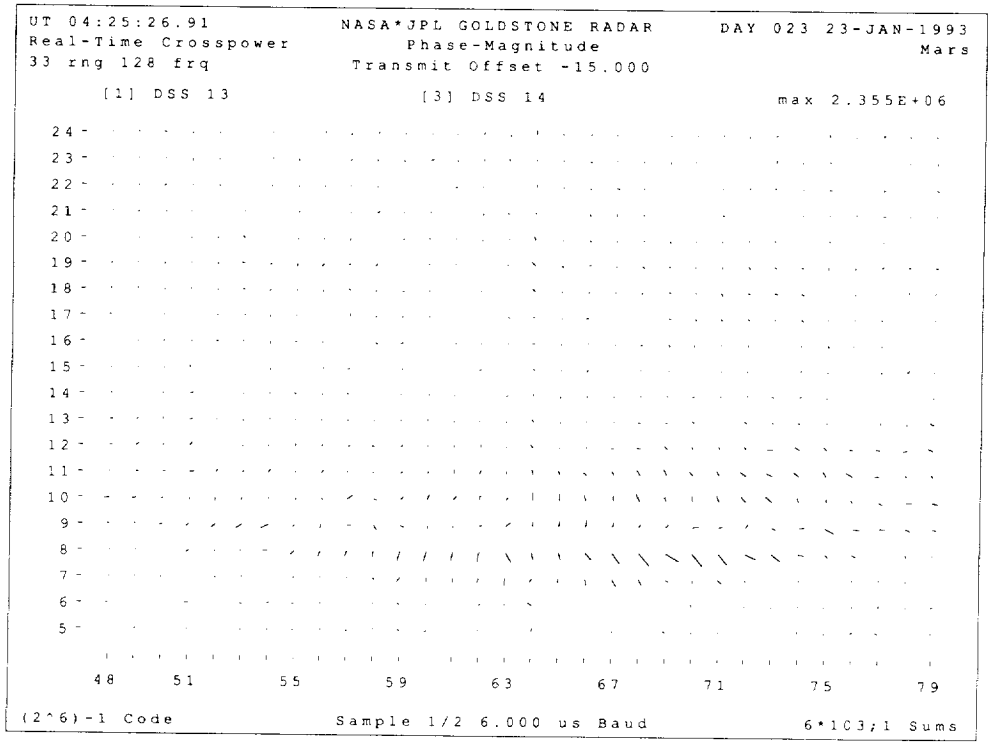


Fig. 6. Real-time interferometry presentation of Mars data (1992–1993).

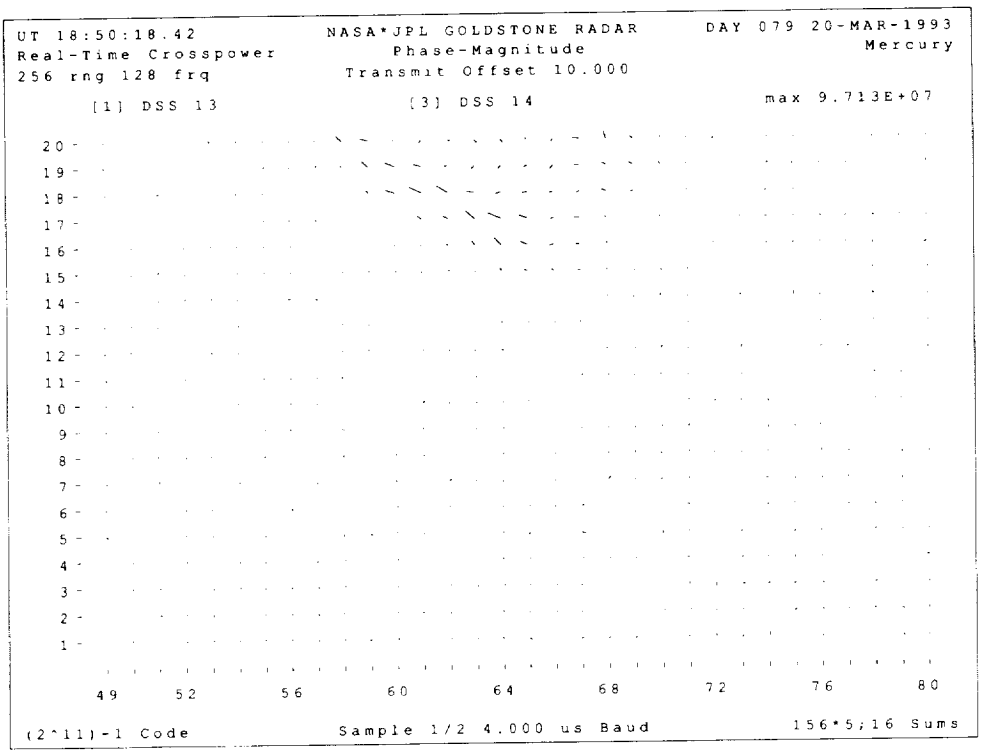


Fig. 7. The first real-time interferometry display of Mercury.

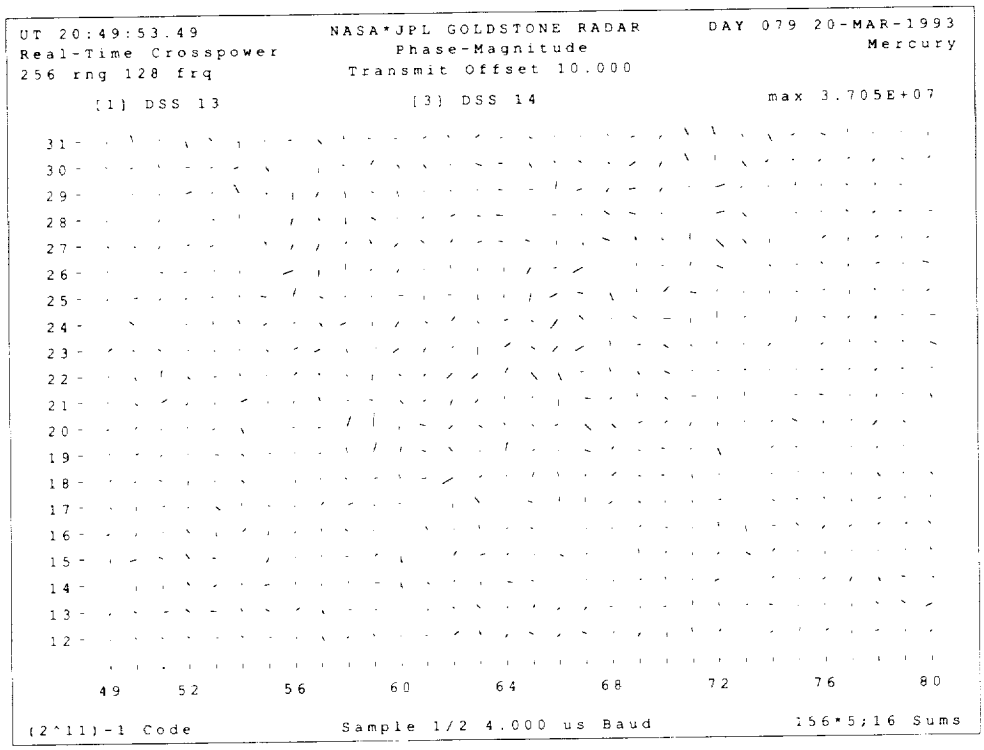
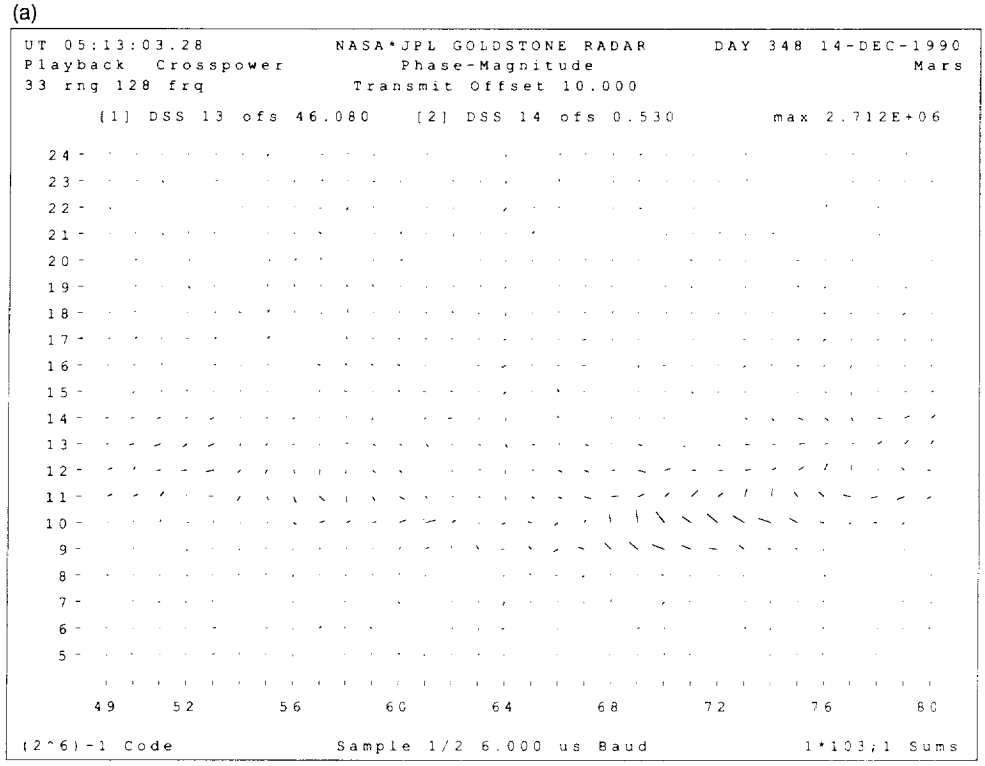


Fig. 8. Three-fold decrease in maximum power during Mercury data acquisition.

(a)



(b)

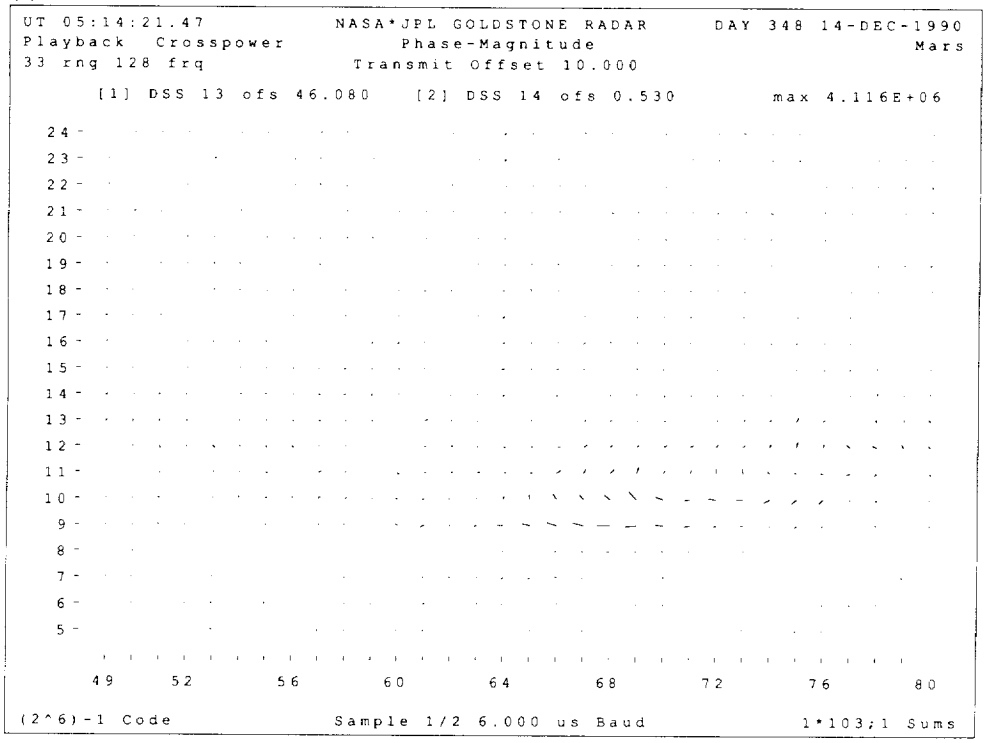


Fig. 9. A "swirl" seen in playback of the Mars 1990 data (a) on the edge of the echo, right-hand side of the display, (b) moving toward the front as the planet rotates, and (c) near the front of the planet.

(c)

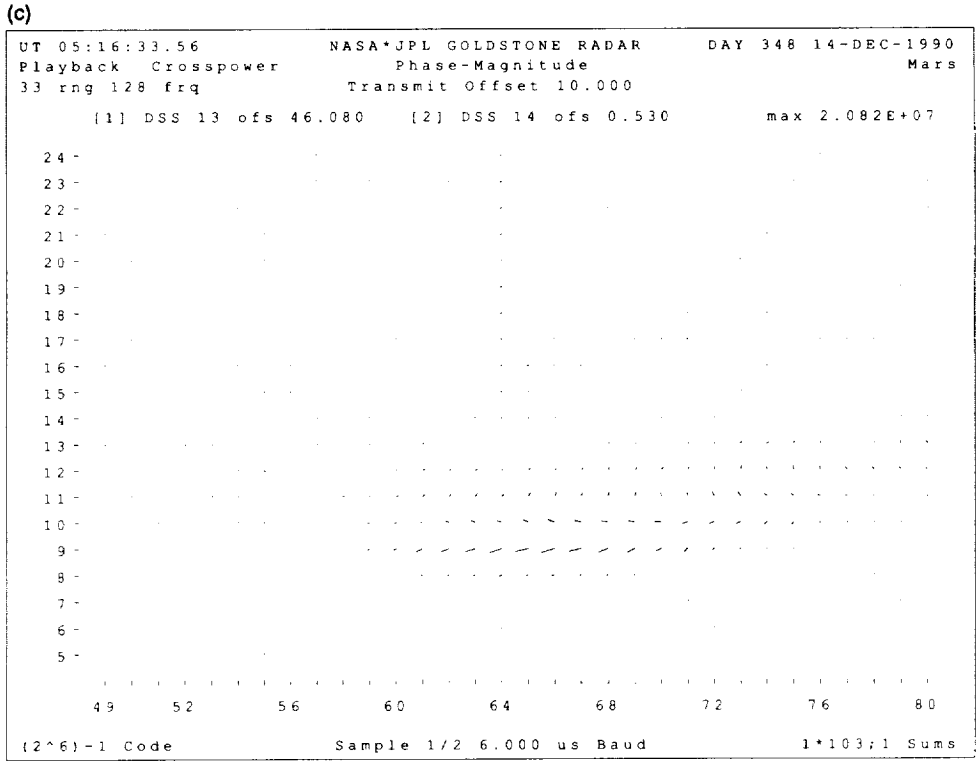

UT 05:16:33.56
NASA*JPL GOLDSTONE RADAR
DAY 348 14-DEC-1990
Playback Crosspower
Phase-Magnitude
Mars
33 rng 128 frq
Transmit Offset 10.000
[1] DSS 13 ofs 46.080
[2] DSS 14 ofs 0.530
max 2.082E+07
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
49
52
56
60
64
68
72
76
80
(2^6)-1 Code
Sample 1/2 6.000 us Baud
1*103;1 Sums

Fig. 9 (cont'd)

Appendix

Interferometry Data Type Description

Let $(t_1, \ldots, t_n)_{r,c}$ represent a vector of complex voltage samples, t_i taken on a channel $c, c = 0, \ldots, C-1$, where C is the number of complex channels for range gate $r; r = 0, \ldots, R-1$, where R is the number of range gates. The variable n is a power of 2, limited by implementation to be between 64 and 2048.

The expression $(f_1, \ldots, f_n)_{r,c}$ represents the result of applying a discrete Fourier transform to $(t_1, \ldots, t_n)_{r,c}$.

The complex spectra from selected channels l and m are combined by applying a complex conjugate multiply to corresponding elements from both spectra:

$$X_{r,1\times m} = (f^*_{1,r,m} f_{1,r,l}, \ldots, f^*_{n,r,m} f_{n,r,l})$$

Crosspower spectra for channels l and m are then computed for all range gates r in the given configuration to form the interferometry data structure.

N94-35561

TDA Progress Report 42-117

May 15, 1994

A New Presentation of Complex Voltage Data for Goldstone Radar Astronomy

S. D. Howard
Communications Systems Research Section

A presentation method similar to a vector field display or a data-based grid has been used to visualize complex voltage test data taken over multiple subchannels. Unlike simple plots of all data points on the complex plane, the position of the data in the time series is part of the presentation, providing additional information to aid in fault isolation during readiness testing. The "phase–magnitude" presentation, as it has come to be called, was designed for the demands of real-time data acquisition and processing with remote monitoring over low bandwidth communication links.

I. Introduction

A new real-time presentation of complex voltage data taken over multiple subchannels has been added as a display option to the Goldstone Solar System Radar ranging system. The new voltage data presentation was adapted from a design devised for complex interferometry data [1], referred to as a "phase–magnitude" presentation. It is similar in appearance to a vector field display or a data-based grid [2]. The voltage phase–magnitude presentation was intended to improve the effectiveness of readiness testing by enabling engineers to discover and correct defects before the data from an observation could be degraded or lost during acquisition due to component failures in the special-purpose signal processing hardware.

The ranging data-acquisition system was initially developed as a rapid prototype. Over the past 5 years, the software has undergone several significant upgrades to enhance reliability, add functionality, improve performance, and correct defects. An early example of one of these reengineering episodes is described in [3]. The most recent upgrade, a replacement of the software that manages the real-time data display, made complex voltage data available to the display and provided an extensible software architecture for experimenting with new methods of visualizing radar data. With the exception of the earliest version of the ranging system (1985) where power profiles were summed from complex voltage data and displayed, prior to the upgrade conjectures about the state of the system during readiness testing were made from integrated power spectra shown in the delay–Doppler display.

Readiness testing in the radar data-acquisition system typically takes the form of local loop-back testing. During a loop-back test, a point-source test signal is injected into the data-acquisition system before the analog-to-digital converters. Incoming test data are then digitized, autocorrelated, coherently summed, transformed to the frequency domain, detected by taking the magnitude squared, incoherently summed, and displayed in real time. The frequency and time delay of the test signal can be set by the operator. Engineers examine the real-time display as

a succession of data records is received, usually at about 30-sec intervals, for evidence of continuous or intermittent defects.

The interpretation of integrated power spectra of test data had two difficulties. First, in the hardware that digitizes, correlates, and sums data, failures occurred in the time domain. Interpretation of the data in the frequency domain was removed from the direct connection to the problem. Second, although integration reduced system noise and data rates during data acquisition, it could also reduce the visibility of intermittent defects during testing. The new presentation addressed these difficulties by presenting a user-selectable subset of time series data on multiple subchannels at selected intervals; integration was optional.

Experimentation with new data visualization methods had been a design goal of the ranging system display upgrade, so the incorporation of this presentation method into the ranging system was also another demonstration of the feasibility of an extensible research and development (R&D) software architecture.

II. Design

The radar test data shown in the phase–magnitude presentation in Fig. 1 are a series of complex voltage samples. As in the case of the interferometry presentation, each phasor represents one complex data point. Square areas representing the complex plane for each data point are nonoverlapping to avoid the visual confusion that might be caused by intersecting phasors. The in-phase and quadrature components correspond to the horizontal and vertical dimensions, respectively. Both phasor components are scaled to the maximum displayed in-phase or quadrature component. The rows in the phasor grid carry samples for a given subchannel or range gate. Columns are time samples taken within each subchannel; time increases from left to right. Only a region of the data is shown at any given time; however, the operator may change the region in the display at any time during the test.

To illustrate the effect of various frequencies on the appearance of the test signal in this kind of presentation, a series of local loop-back tests taken with the ranging system in a typical Mercury configuration (oversampled with a 4.0-μsec range resolution) is shown in Fig. 2. Because the configuration is oversampled, the test signal can be seen in range gates adjacent to range gate 11. As time increases from left to right, the phasor angle advances or retards, depending on the frequency of the test signal. The length of the phasor should be constant. In practice, since a small positive frequency is usually used to offset the point-source test signal from dc, a display similar to the 5-Hz case in Fig. 2(d) has become a familiar signature of good test data when viewed with the phase–magnitude presentation.

System noise is shown in Fig. 3. Gaussian noise looks random in the display. One of the early concerns about using a data-based grid in the design of the data graphic was that the viewer might be induced to perceive a pattern in the graphic when no such relationship actually existed in the data. Experience with the display has greatly reduced that concern, although assertions about subjective experience are difficult to quantify. In practice, engineers have been able to recognize and challenge subtle differences between pure system noise and noise that accidentally included other artifacts.

Expert test engineers rely on their ability to recognize patterns and subtle mismatches in patterns through visual, auditory, and sometimes tactile modes when they evaluate the behavior of complex systems. One of the goals of the phase–magnitude design was to provide not only readily interpretable signs about the basic state of the system but additional visual support for complex information gathering and decision making as well. A test engineer, experienced with the system and the presentation, should be able to glance at the display and immediately identify the appearance and source of familiar problems. In addition, if the problem did not match some previous experience, the display ought to provide a good representation of the available information to assist the troubleshooting process.

A common method of presenting sets of complex data samples graphically is to simply plot them all on a single complex plane. The advantage of this approach is that many data points can be included; this is useful for indicating whether or not a failure or anomaly has occurred. The disadvantage is that the association between each data point and its position in the time series is not defined by the graphic. The presence of a number of subchannels in the ranging system data further complicates the use of this kind of approach. The design of the new data graphic, by contrast, accepted the reduction in the number of data points plotted, in favor of retaining information about the position of data points in the time series. Since rapid fault isolation is the goal, the time dimension in the phase–magnitude presentation can show when the anomaly is occurring, giving additional information about the origin of the problem.

III. Experience

The first diagnostic use of the new method showed a defect that was later traced to a memory chip failure in a correlator–accumulator module. A printout of a typical display is shown in Fig. 4. Instead of the expected Gaussian noise display, similar to Fig. 3, a few large negative in-phase values in range gates 173, 185, and 187 dominated. The large negative values did not usually persist in fixed time-sample locations from data record to data record, as Fig. 4 might suggest, but occurred in various time-sample locations throughout the range gate, indicating that the defect was probably occurring after sampling. A display like Fig. 4 now suggests that this kind of failure needs to be investigated; memory chip failures are among the more frequently encountered failures in the radar modules.

The second experience, illustrated by the printout of the display in Fig. 5, alerted sustaining engineering to an unsuspected problem. Since the system is set for oversampling in this configuration, the test signal is expected to spread over three range gates. However, the test signal also apparently drifts forward in range from gates 10, 11, and 12 to gates 8, 9, and 10 in the first 12 time samples. After a timing adjustment of the baud-integrating analog-to-digital converters that supply data to digital channel 1, the test results showed a normal signal in range gates 10, 11, and 12, similar to that in Fig. 1.

IV. Conclusions

The key elements of the design were probably the retention of the time dimension in the data graphic and the acceptability of being able to view only selected regions of the data record at any given time. However, performance and aesthetics were also factors in the decision to incorporate this display method into the standard set. Because this design is an adaptation of the interferometry design, it places a low processing burden on the real-time system and retains the acceptable performance characteristics that allow remote monitoring over 9600-bps communication links. The display of a typical loop-back test with a small frequency offset exhibits a varying but symmetrical and aesthetically pleasing pattern. This was a fortuitous result. Unsatisfactory outcomes either in performance or in overall graphic design would have argued against pursuing this method as an operational tool.

Several measurements of how well the presentation conveys information might be of interest if the opportunity for more quantitatively based testing arises. First, some measurement might be made of the ability of users to identify a printout as being either of Gaussian system noise or noise combined with dc. There is probably a threshold below which users cannot distinguish the presence of dc. That threshold may vary among users and with experience. Second, it should also be possible to test the ability of users to identify printouts as being either from normal loop-back tests or from tests showing captured or induced system defects. If printouts are identified as being of system defects, then the user might be asked to identify a prioritized list of candidate problem sources. Again, some variation in ability among users is expected.

The initial informal evaluation of the voltage phase–magnitude presentation contributed to the preparation for the successful observations of asteroid 4179 Toutatis and of Mars in December 1992. Additional informal assessments of the power of this representation of complex voltage data to carry relevant information to users will continue as new situations arise in testing.

Acknowledgments

The author thanks Carl Franck and Ray Jurgens for assistance in setting up calibrated loop-back tests at DSS 14. Ron Winkler, Carl Elston, Larry Snedecker, and Paul Dendrenos have assisted with station support at DSS 14 on many occasions. Ray Jurgens provided insights and suggested new approaches during the exploration of the new display. Paul Willis made comments that improved this article.

References

[1] S. D. Howard, "The Real-Time Display of Interferometry Data for Goldstone Radar Astronomy Data-Acquisition," *The Telecommunications and Data Acquisition Progress Report 42-117, vol. January–March 1994*, Jet Propulsion Laboratory, Pasadena, California, pp. 213–227, May 15, 1994.

[2] E. R. Tufte, *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983.

[3] S. D. Howard, "Improving a Data-Acquisition Software System With Abstract Data Type Components," *The Telecommunications and Data Acquisition Progress Report 42-100, vol. October–December 1989*, Jet Propulsion Laboratory, Pasadena, California, pp. 248–251, February 15, 1990.

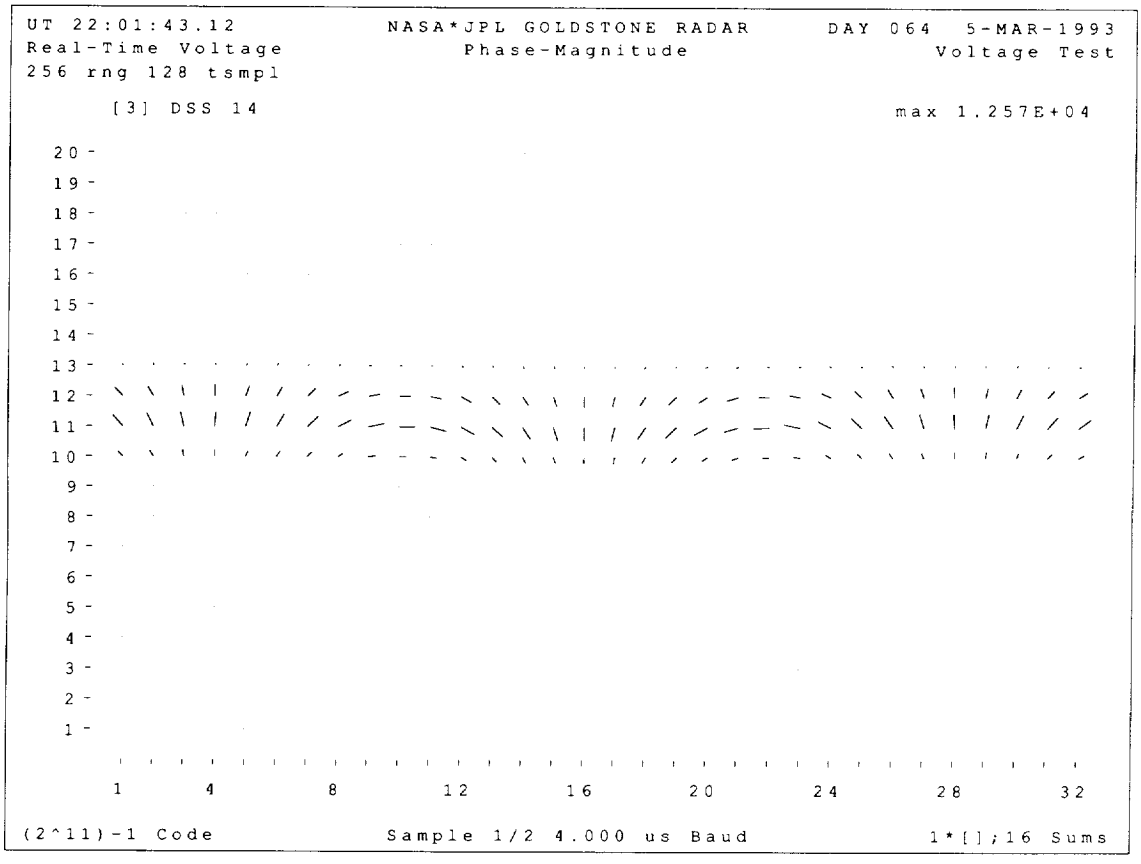


Fig. 1. Normal loop-back test of oversampled configuration.

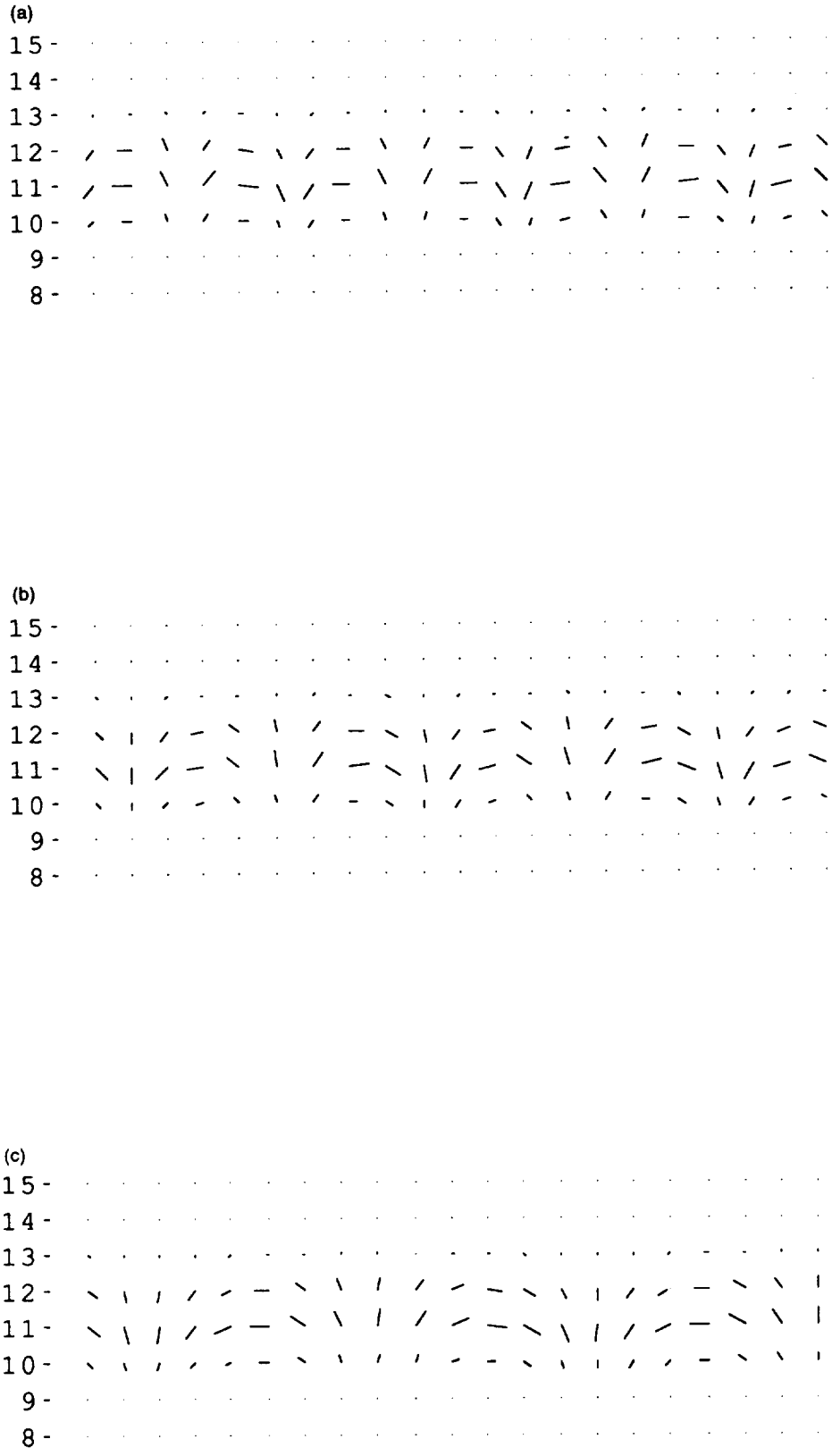
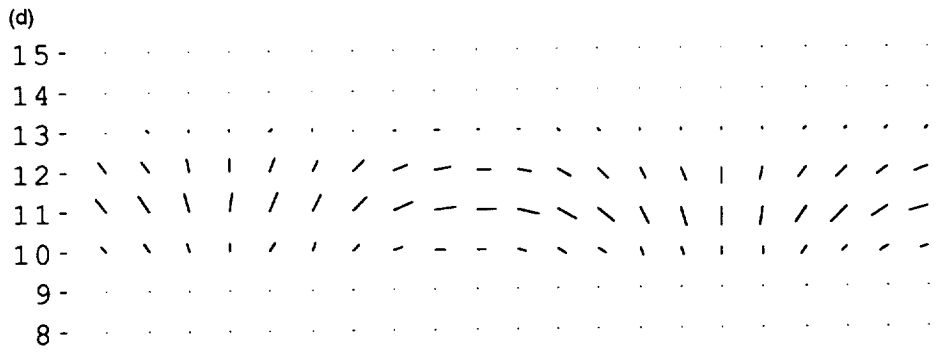
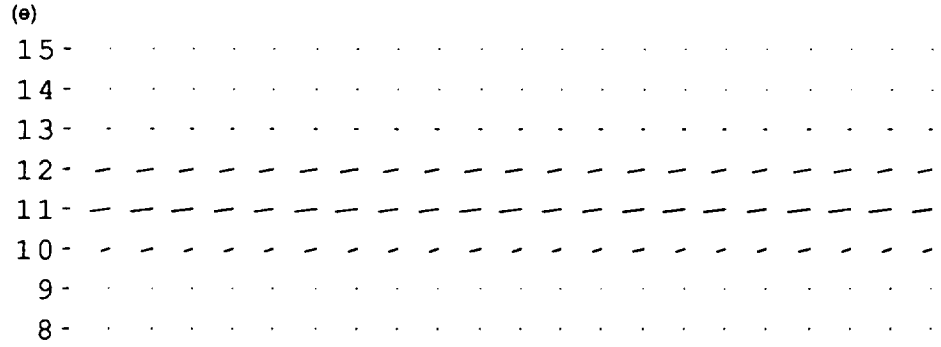


Fig. 2. Normal oversampled loop-back test at (a) 20, (b) 15, (c) 10, (d) 5, (e) 0, and (f) −5 Hz.





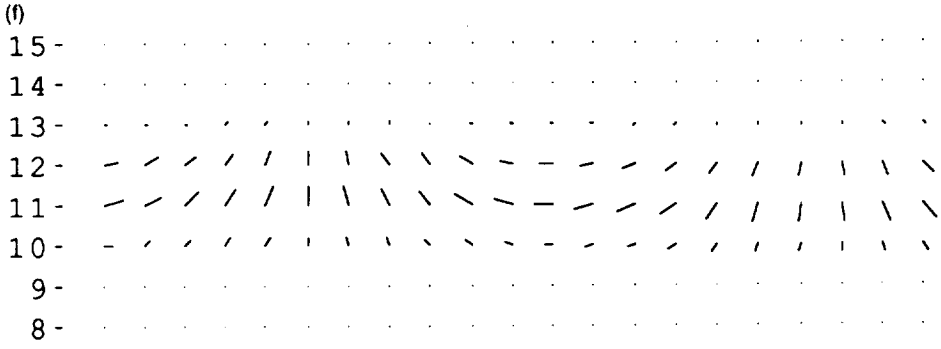


Fig. 2. (Cont'd)

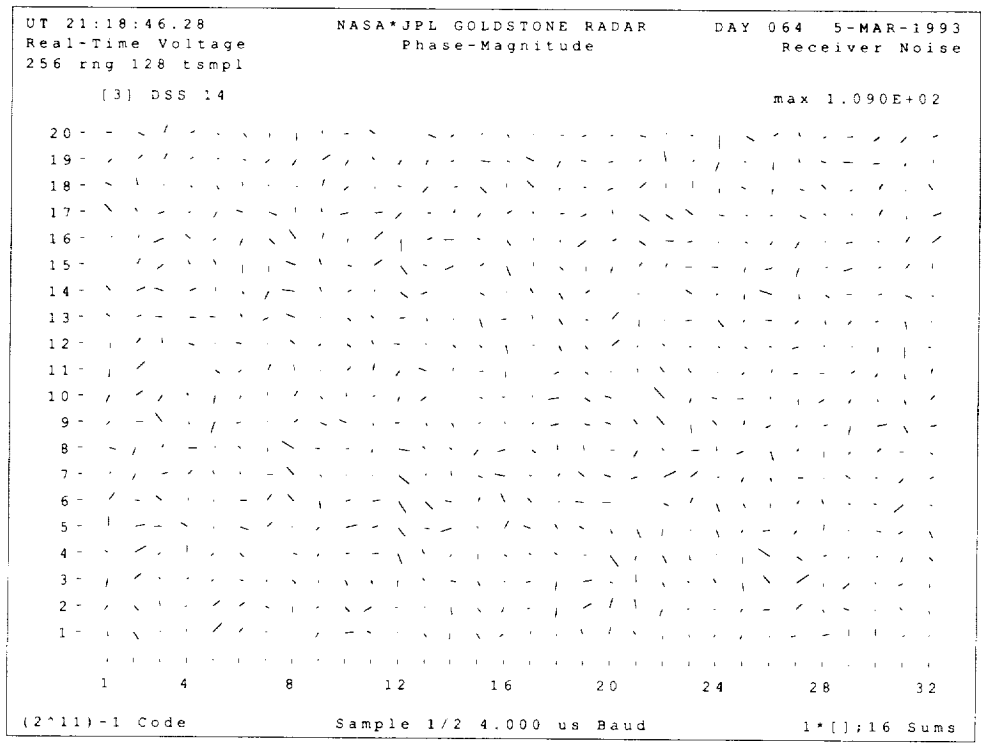


Fig. 3. Gaussian noise.

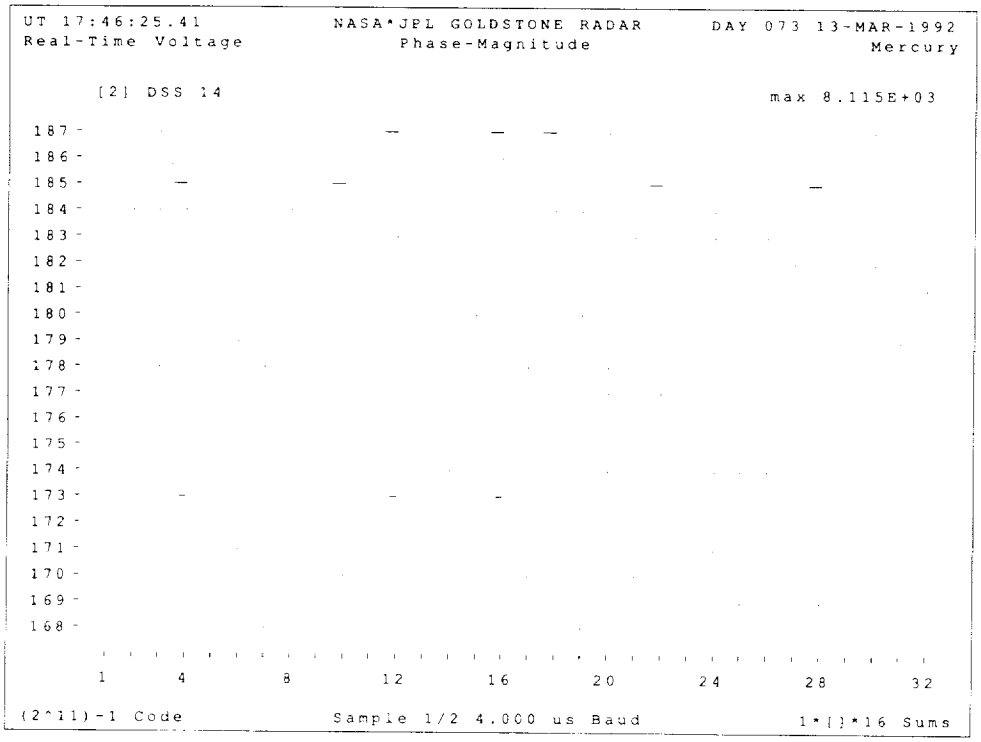


Fig. 4. Evidence of a memory failure in a correlator–accumulator module.

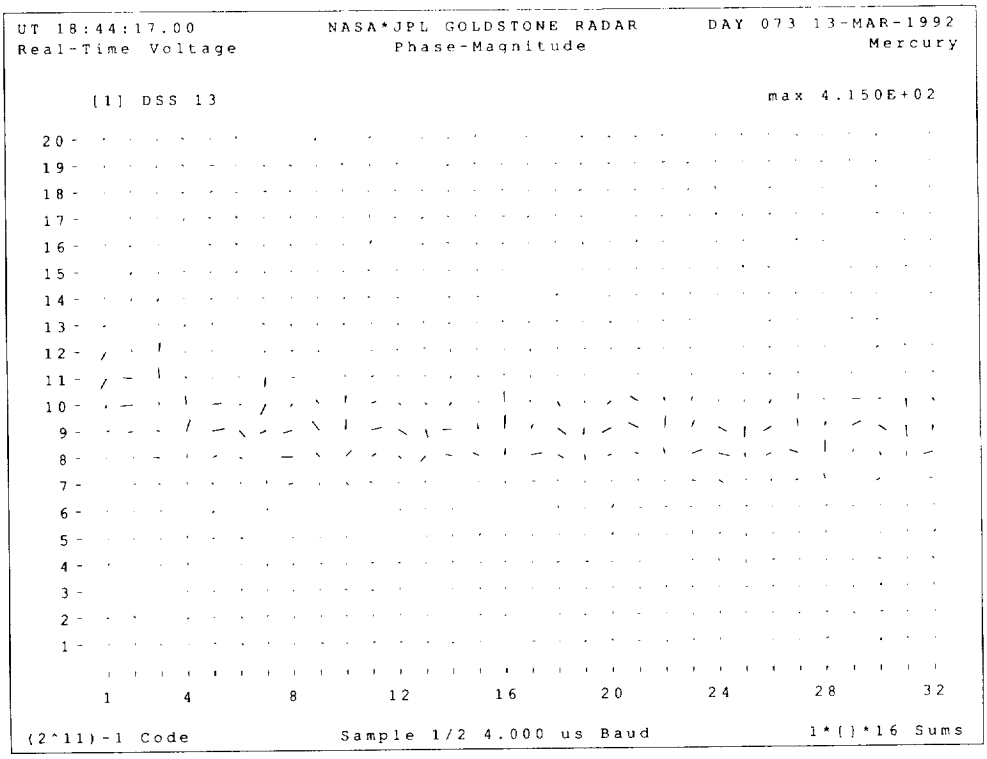


Fig. 5. Analog-to-digital converter adjustment problem.

S20-33
4107

N94-35562

TDA Progress Report 42-117

May 15, 1994

P-14

A 32-GHz Solid-State Power Amplifier for Deep Space Communications

P. D. Wamhof, D. L. Rascoe, K. A. Lee, and F. S. Lansing
Spacecraft Telecommunications Equipment Section

A 1.5-W solid-state power amplifier (SSPA) has been demonstrated as part of an effort to develop and evaluate state-of-the-art transmitter and receiver components at 32 and 34 GHz for future deep space missions. Output power and efficiency measurements for a monolithic millimeter-wave integrated circuit (MMIC)-based SSPA are reported. Technical design details for the various modules and a thermal analysis are discussed, as well as future plans.

I. Introduction

The Deep Space Network is developing telecommunication capability at 32 GHz (Ka-band) with expected improvement of as much as 8 dB over the current performance at 8 GHz (X-band). A proof-of-concept 32-GHz solid-state power amplifier (SSPA) with an output power of 1.56 W has been designed and demonstrated for future spacecraft applications.

The objective was to demonstrate useful RF output power of greater than 1.5 W with an efficiency greater than 15 percent at 32 GHz. We also wanted to use state-of-the-art components that were readily available from vendors and to build a compact system that could be interfaced with the deep space transponder.

The development of this SSPA provided practical hands-on experience in the incorporation of Ka-band monolithic millimeter-wave integrated circuits (MMICs) into an amplifier architecture. Other useful information was obtained in the areas of mechanical and electronic system integration and high-power RF testing. The completed hardware from this effort is currently being used as a stepping stone toward 32-GHz SSPAs having higher efficiencies (>25 percent) and output power capacity (1.7 W).

II. Subsystem Design

Some initial goals and requirements for the 32-GHz SSPA were established based upon future deep space mission requirements. An RF output power level of 1.5 W having greater than 15 percent power added efficiency was selected, which would make the SSPA a good candidate for insertion into the Pluto Fast Flyby mission. The SSPA would also require 31 dB of RF gain in order to interface with the deep space transponder that provides 1 mW of input power to the SSPA.

We surveyed industry to find 32-GHz high-power devices (>0.5 W) having efficiencies greater than 25 percent in order to meet these requirements and came up with three possible vendors. GE Aerospace has 1.0-W 32-GHz pseudomorphic high electron-mobility transistor (PHEMT) discrete devices with 30 percent power added efficiency, but these were not available for use outside of GE Aerospace. Raytheon initially claimed to have

0.5-W PHEMT devices with 30 percent power added efficiency; however, sample parts only delivered 0.4 W and fell short of our requirements. Alpha has 0.5-W metal-semiconductor field effect transistor (MESFET) MMIC devices having 25 percent power added efficiency and 11-dB gain, which were available for immediate purchase. We selected the Alpha AA035-P2 0.5-W MMIC as our power amplifier component.

A. SSPA Architecture

The output from the MMICs can be combined in different ways to meet the 1.5-W output power requirement. Several architectures for the SSPA were considered, but the topology shown in Fig. 1 provided the best performance in terms of gain and power added efficiency, and used the fewest number of parts (based on the parts available at that time). The output stage for the SSPA consists of four 0.5-W power amplifier MMICs (Alpha AA035-P2) where the outputs of these four devices are combined in a corporate, in-phase, Gysel [1] microstripline combiner. The input signal to the SSPA is amplified by a preamplifier MMIC (Alpha AA035-P3) followed by a high-power driver MMIC (Alpha AA035-P2) and split through a four-way Gysel microstripline divider to feed each of the four output stages. A further consideration in determining the optimal architecture is the distribution of gain through the cascaded chain. Our SSPA design ensured that each amplifier would enter gain compression at the same unit input drive level. If the driver amplifier were to saturate before the output stages, the desired output power would not be achieved. Table 1 shows the gain distribution and efficiency calculation for this 1.5-W RF amplifier architecture, and Fig. 1 shows the associated block diagram.

The mechanical interfaces were determined by considering the other system components (transponder and antenna) as well as performance and overall size. A waveguide provides the lowest loss, and minimizing the loss (especially at the output) is important in maximizing the efficiency at the expense of larger size. The transponder output port is a coaxial connector, and the antenna input port is a waveguide. At both the input and output of the SSPA, a transition from waveguide to microstrip is required. At the input, a transition from coax to waveguide is required to mate to the transponder.

B. MMIC Amplifiers

The preamplifier MMIC (Alpha AA035-P3, the first device in the chain) provides over half the RF gain of the SSPA (17.9 dB). The frequency response and return loss plot for the preamplifier is shown in Fig. 2. This MMIC contains three identical cascaded MESFET stages with associated matching and bias circuitry to operate from a single supply, all on a single 3.9-mm by 1.9-mm chip.

The power amplifier MMICs and driver amplifier MMIC (Fig. 3) each deliver 0.5 W of RF power. The frequency response and return loss plot for an individual MMIC power amplifier is shown in Fig. 4. The output power and power added efficiency for this power MMIC is shown in Fig. 5, while the corresponding gain compression curve is shown in Fig. 6. This chip contains an input stage of two MESFETs in parallel, each driving two of the four MESFETs that make up the output stage. Power dividers, interstage matching networks, and bias networks are all integrated onto the 3.9-mm by 3.1-mm chip.

An important design criterion when combining amplifiers together in parallel to increase RF power is phase matching within the amplifier channels. For a worst-case phase imbalance between MMIC amplifiers of ϕ, the combined voltage adds vectorially so the total normalized power of four amplifiers would be $4\cos^2\ \phi$. The power lost due to phase imbalance is $4(1-\cos^2\ \phi)$. Ideally, we would like each amplifier to have the same insertion phase. For each of the power amplifiers we purchased, Alpha Industries provided us with scattering parameters with which we determined the insertion phase for each amplifier. We carefully selected amplifiers to maintain a phase balance of ± 5 deg to keep this power loss less than 0.03 dB.

C. Power Dividers

The power combiners allow us to combine the power of the four relatively low-power devices to achieve the desired output power level, which is significantly higher than what can be obtained with a single device. The power divider requirements for this SSPA were small size (smaller than a waveguide), low loss (less than 1.2 dB), good phase match (less than 2 deg), and good port isolation (greater than 15 dB). Typical microstrip power combiners include tee dividers, branchline and ratrace couplers, Wilkinson dividers, and Gysel dividers. Wilkinson and Gysel dividers are the only type that provide both phase match and isolation. A Gysel divider is capable of handling higher RF power levels while maintaining phase match and isolation because the isolation resistors are positioned to provide a heat conduction path to the ground plane, whereas in the Wilkinson divider, they do not.

The initial power divider shown in Fig. 7 was designed using EEsof's Touchstone circuit analysis software. Like the Wilkinson, it divides the power through two 75-ohm quarter-wave sections. The two isolation resistors are joined into the remaining 75-ohm quarter-wave

ring through 100-ohm quarter-wave lines. Compensation for mismatches at the tee-junctions was incorporated into the geometry of the divider by using 30-deg notch cutouts [2]. A single two-way divider was fabricated at the David Sarnoff Hybrid Processing Facility. Thin-film deposition processes were used to fabricate the microstripline circuit and the thin-film isolation resistors onto the 0.25-mm (10-mil)-thick alumina substrates. The measured insertion loss, isolation, and phase match for the two-way divider were 0.5 dB, 12 dB, and 2 deg, respectively. The insertion loss is shown in Fig. 8. This circuit was also analyzed using finite difference time domain (FDTD), a full-wave analysis method for solving Maxwell's equations that include electromagnetic effects not accounted for in Touchstone's circuit simulator. FDTD analysis predicted a 0.4-dB insertion loss as compared to 0.2 dB on Touchstone (Fig. 9).

III. SSPA Integration

The SSPA was designed such that two separate amplifier modules, a preamplifier/driver module and an output module, could be tested and then dropped into a housing to be integrated with the input and output waveguide ports and bias supply connections.

A. Amplifier Modules

The amplifier modules consist of carriers that integrate the MMIC devices with passive transmission line circuits, RF decoupling capacitors, and dc bias ports. Molybdenum was selected as the carrier material because its thermal expansion coefficient is well matched to the alumina substrates mounted to it. In addition, molybdenum provides higher heat conduction for the thermal management of the power devices. Substrates and devices are mounted with conductive silver epoxy into recessed channels in the carrier, which helps suppress potential RF moding. Bias voltages are injected through the bottom of the carrier with miniature dc feed throughs. Two parallel bypass capacitors are located adjacent to the MMIC dc bonding pads. The small 0.6-pF bypass capacitors located closest to the MMIC device filter frequencies above 20 GHz, while the 100-pF capacitor provides higher-level bypassing at the lower frequencies. This filtering effect was modeled using EEsof's Touchstone and is shown in Fig. 10, where the filter response is shown with and without the 0.6-pF capacitor.

B. Housing

The SSPA housing was designed as a double-sided unit with two end pieces containing the waveguide ports. The preamplifier and output power amplifier modules are located on the top side of the housing (Fig. 11), and dc wiring from the Cannon-D connector to the miniature feed throughs on the modules are routed on the bottom side (Fig. 12). Large 0.1-μF ceramic bypass capacitors that stabilize the amplifiers against low-frequency oscillation are also located in the bottom compartment.

The end pieces are WR-28 waveguides incorporating E-plane waveguide probes. These probes are of the same design as those used in the DSS-13 experimental 32-GHz low-noise HEMT amplifiers. The probes are connected to the microstrip transmission lines via glass bead feed throughs (the same as those used in Wiltron's K-connectors) soldered into the housing. Figure 13 shows a cross-sectional view of the transition and E-plane probe assembly.

C. Thermal Issues

A thermal analysis was performed for the SSPA; it indicated that the maximum expected junction temperature for a 50 deg C baseplate (preliminary Pluto specification) would be 82 deg C provided that gold/tin eutectic solder is used to secure the MMICs to a molybdenum carrier. To speed up the assembly process time, we used conductive silver epoxy to secure the MMICs, bypass capacitors, and microstrip circuits. By using conductive epoxy, the thermal analysis predicted that the junction temperatures would change from 82 deg C to somewhere between 92–115 deg C depending upon the epoxy thickness and consistency. For our initial benchtop demonstration, the breadboard SSPA is held at room temperature (20 deg C), so the MMICs operate with junction temperatures of 62–85 deg C, and the use of silver epoxy is not a thermal concern. Typically, junction temperatures should be kept at or below 110 deg C for acceptable reliability in long-term spacecraft missions.

The thermal model used actual layout dimensions for all of the heat producing field effect transistors (FETs) within the MMICs placed on carriers within the housing. Table 2 shows the thermal conductivity values of the different materials used in the electronics package.

IV. SSPA Performance

The complete SSPA performance parameters are summarized in Table 3 and discussed in detail below. The test setup used for integration and test is shown in Fig. 14. In order to make the single-frequency (31.5-GHz) measurements to generate the compression curves (Figs. 15 and

16), the sweep oscillator source was used in conjunction with a power meter and calibrated waveguide coupler only.

Output power was measured to be 1.56 W with a corresponding power added efficiency of 15.4 percent. Figure 15 shows output power and power added efficiency plotted against input drive level. The amplifier was tuned to have maximum power at 31.5 GHz; however, it can be tuned for the same performance at 32 GHz.

Figure 16 shows SSPA gain and power added efficiency plotted against drive level. Note that in order to achieve the 1.5-W output power level, the SSPA is driven well into compression. The corresponding gain compression level at 1.5-W output power is 3.5 dB. At the maximum output power of 1.56 W, the corresponding compression level is 8.7 dB.

Figure 17 shows a swept frequency response from 27.5–37.5 GHz. As mentioned above, the peak power occurs at 31.5 GHz and is tunable from approximately 31–33 GHz by using gold ribbon stubs located on the power divider transmission lines. The active power devices (Alpha AA035P2-00) have power gain variations of ± 1.2 dB from 30–34 GHz, while the input and output power combiners have minimum insertion loss in this frequency region. The preamplifier (Alpha AA035P3-00) has consistent gain flatness from 30–36 GHz and does not affect the response as severely as the output devices. The input return loss was measured to be 19 dB looking into the input waveguide port. The output return loss was 12 dB and was measured at the output waveguide flange.

V. Summary

A solid-state power amplifier operating at 31.5 GHz with greater than 1.5 W total RF power and greater than 15 percent power added efficiency has been demonstrated using existing devices. Fabrication of the hardware provided valuable practical experience in the area of MMIC insertion. System-level issues for SSPAs were also addressed, including the impact of output losses on efficiency, phase matching within the parallel amplifier channels, bias circuit design, and system gain distribution.

Several areas have been identified for potential future work. The advanced high-power discrete devices from GE Aerospace (now Martin Marietta) having power added efficiencies of 30 percent have been integrated into an SSPA output module with an output power of 1.7 W (25 deg C), a corresponding power added efficiency of 27.8 percent, and an RF gain of 7.3 dB. A discrete preamplifier section could be designed to interface with this output module and boost the RF gain up to 32 dB with an efficiency of 27 percent. Additionally, in the near term, the SSPA described herein is being planned to be tested at low temperatures (-100 deg C) to study improved output power and efficiency behavior. This temperature range would be easily achieved using passive radiative technology on a spacecraft.

Acknowledgments

The authors gratefully acknowledge Dr. G. Birur (JPL) for the thermal analysis, Dr. S. Gedney at the University of Kentucky for electromagnetic simulation, C. Cruzan (JPL) and S. Chavez (JPL) for assembly assistance, and P. Pelka and S. Bennett at the David Sarnoff Research Laboratory for manufacturing assistance.

References

[1] R. Knochel and B. Mayer, "Broadband Printed Circuit 0°/180° Couplers and High Power Inphase Power Dividers," *IEEE MTT-S Symposium Digest*, pp. 471–474, 1990.

[2] R. Chadha and K. C. Gupta, "Compensation of Discontinuities in Planar Transmission Lines," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-30, no. 12, pp. 2151–2155, December 1982.

Table 1. SSPA gain distribution and predicted performance parameters.

Component	dc power consumption, W	Compressed gain, dB	Net gain, dB	Output power, dBm
Preamplifier	1.128	17.23	17.23	20.23
Driver amplifier	1.971	7.64	24.87	27.87
Two-way divider	0.000	−3.60	21.27	24.27
Two-way divider	0.000	−3.50	17.77	20.77
High-power amplifier	7.816	7.19	24.96	27.96
Two-way combiner	0.000	2.50	27.46	30.46
Two-way combiner	0.000	2.40	29.86	32.86
Waveguide transition	0.000	−0.25	29.61	32.61
Isolator	0.000	−0.15	29.46	32.46

Input parameters

Input power, 3.00 dBm
RF phase alignment, 10 deg

Calculated parameters

RF output power, 1.71 W
SSPA dc power, 10.92 W
SSPA efficiency, 15.66 percent

Table 2. Thermal conductivities used in thermal analysis calculations.

Material	Thermal conductivity, W/m-deg C
GaAs MMIC	44.1
AuSn eutectic solder	57.1
Silver epoxy	2.0–5.9
Molybdenum	133.8
Aluminum	157.4

Table 3. SSPA performance parameters.

Design parameter	Measurement
Output power, W	1.56
Power added efficiency, percent	15.4
Small signal gain, dB	42.7
Total dc power, W	10.1
Input return loss, dB	19
Output return loss, dB	12

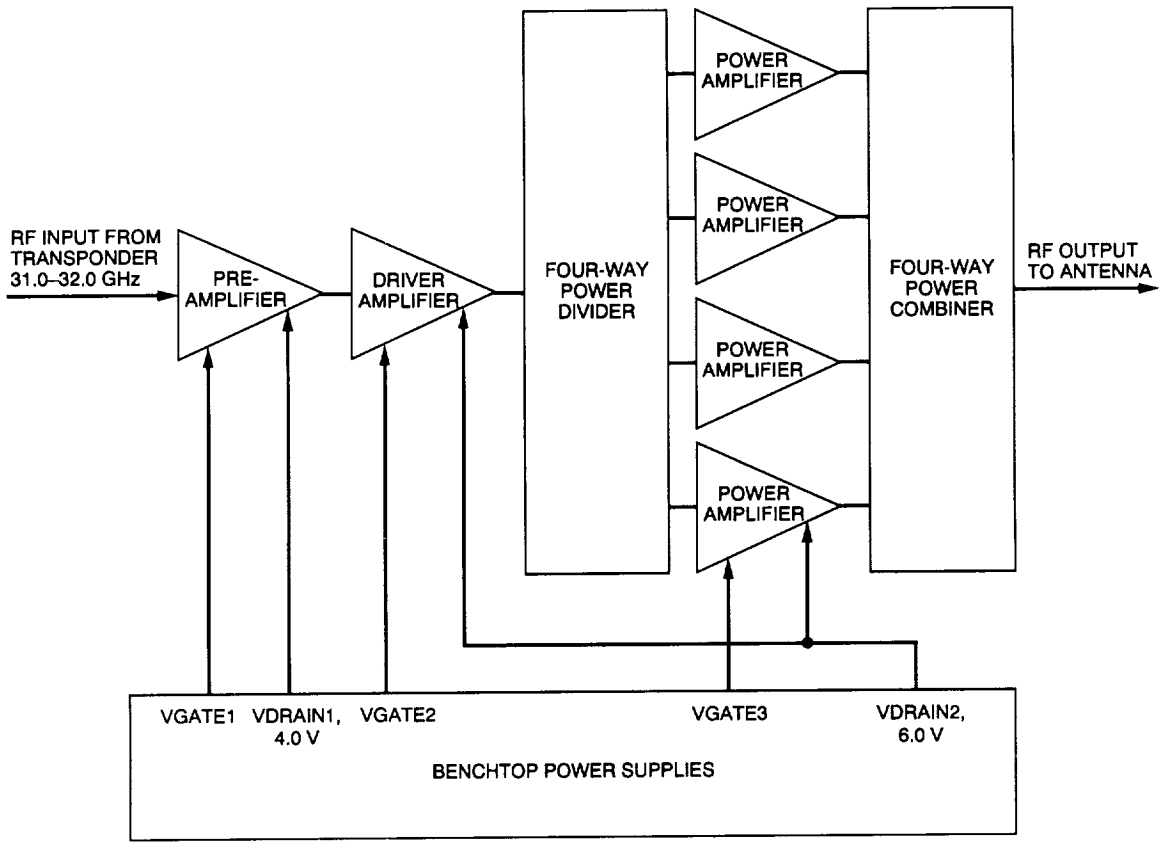


Fig. 1. The 32-GHz solid-state power amplifier.

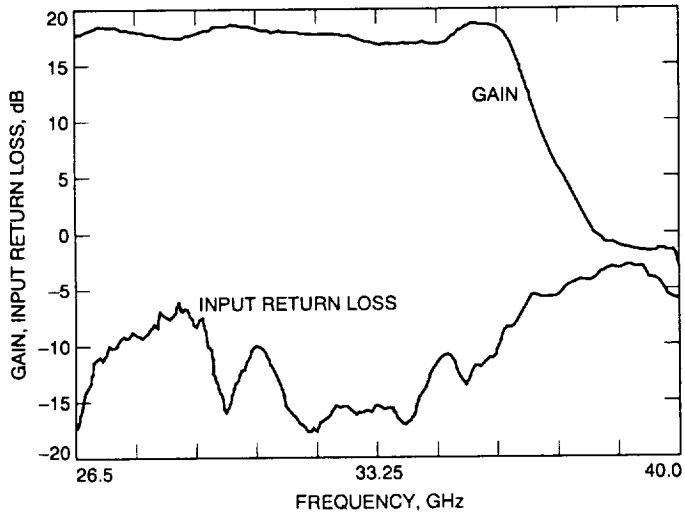


Fig. 2. Alpha MMIC preamplifier frequency response and return loss plots showing a 17.9-dB gain at 32 GHz and an associated return loss of 17 dB.

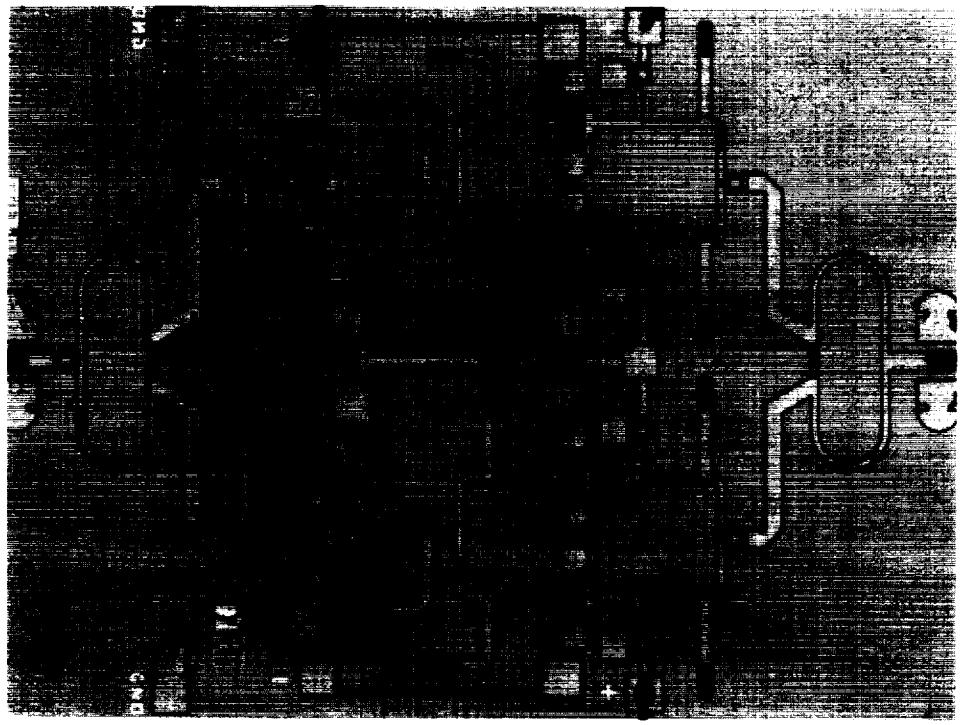


Fig. 3. Alpha 0.5-W MMIC power amplifier chip.

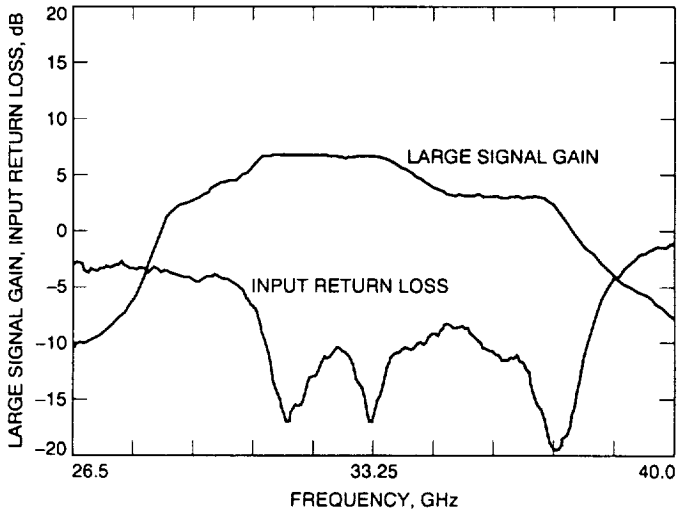


Fig. 4. Alpha MMIC power amplifier frequency response and return loss plot showing a 7.0-dB gain (compressed) at 32 GHz with an associated return loss of 12 dB.

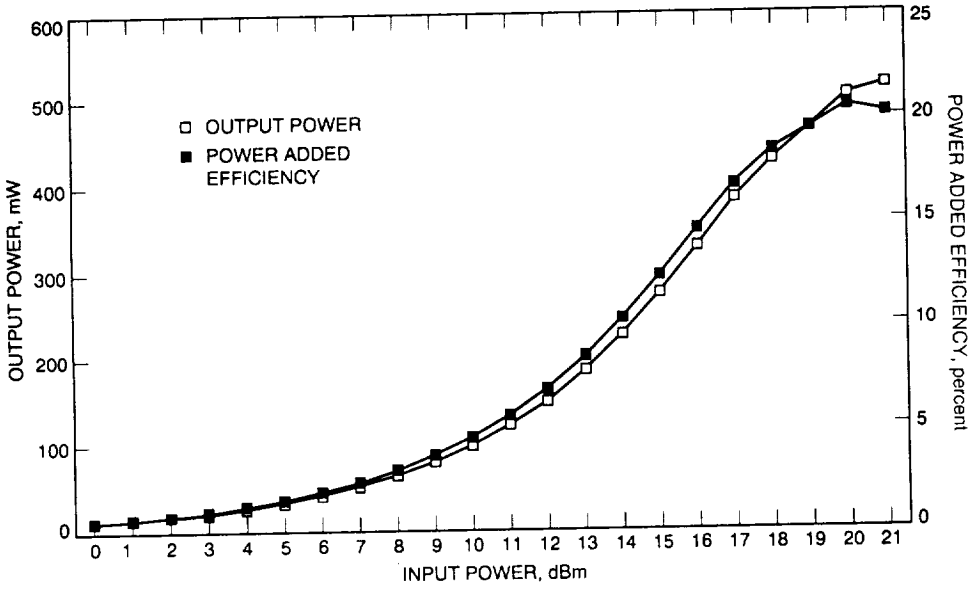


Fig. 5. Output power and power added efficiency plotted against input drive level for the Alpha MMIC power amplifier.

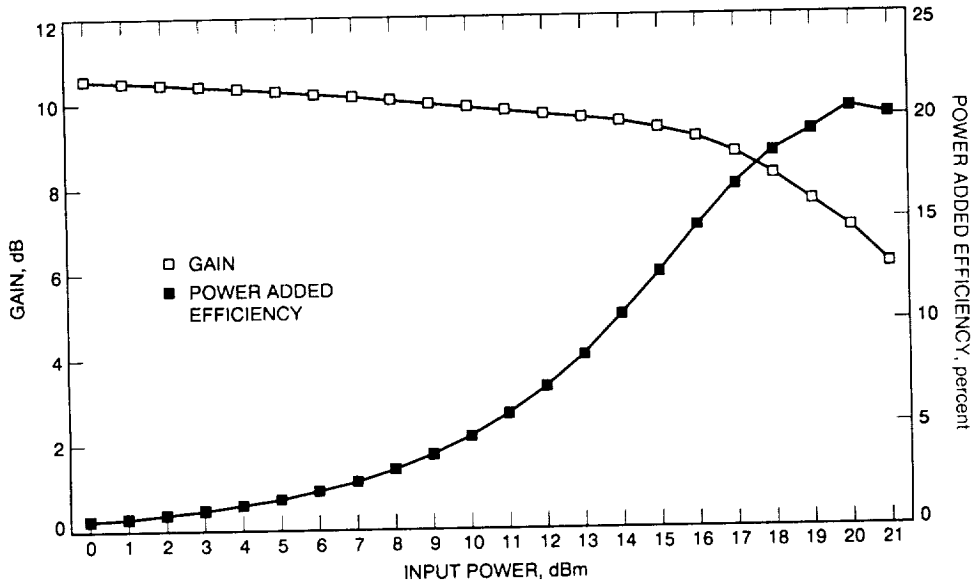


Fig. 6. Amplifier gain and power added efficiency plotted against input drive level for the Alpha MMIC power amplifier.

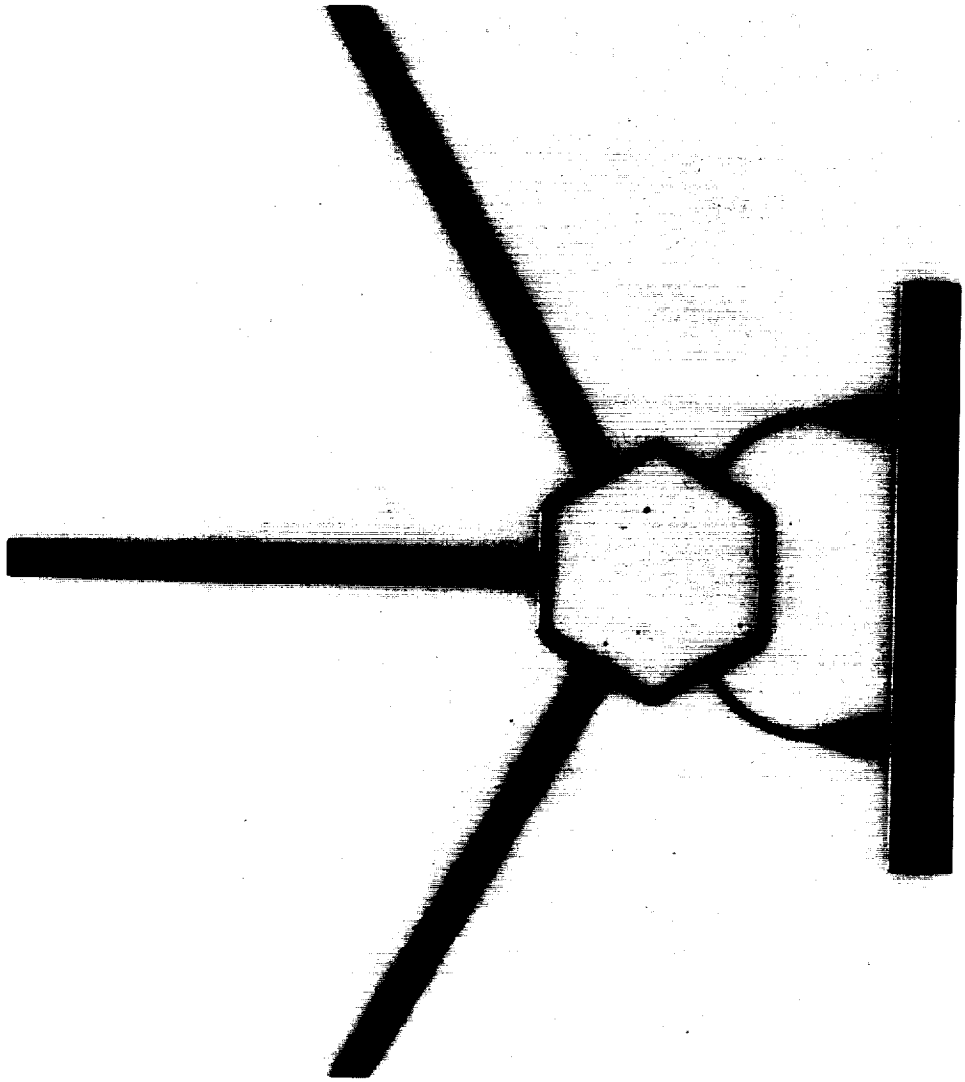


Fig. 7. The Gysel two-way power divider.

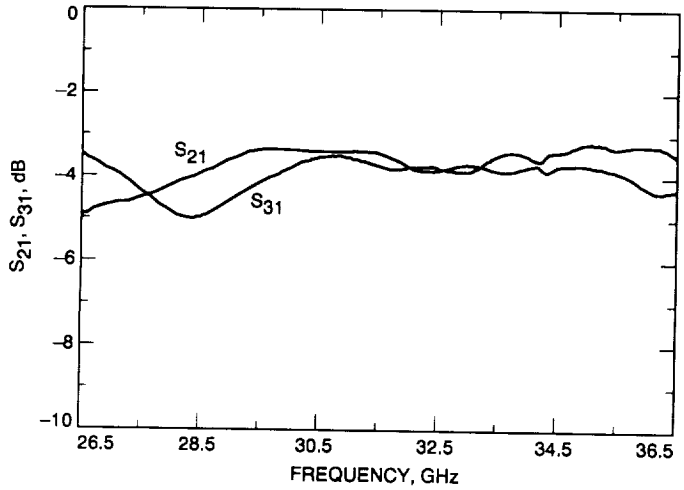


Fig. 8. Measured insertion loss of a two-way Gysel power divider.

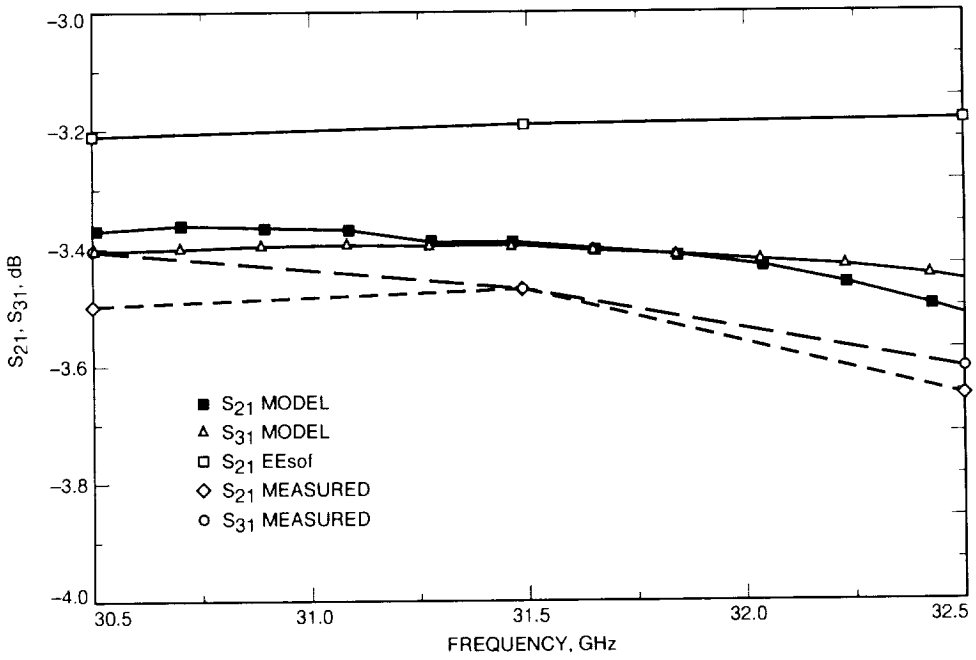


Fig. 9. Comparison between measured data, EEsof Touchstone model predictions, and FDTD model predictions for the Gysel two-way power divider. The FDTD model gave a more accurate insertion loss prediction.

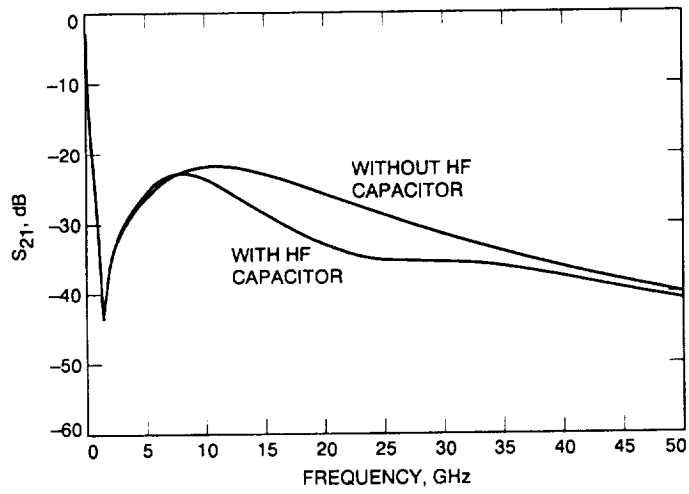


Fig. 10. The dc bypass filter response characteristics, showing the effect of adding a secondary 0.6-pF high-frequency (HF) capacitor in the network. An additional 5–10 dB of rejection is obtained between 10–37 GHz.

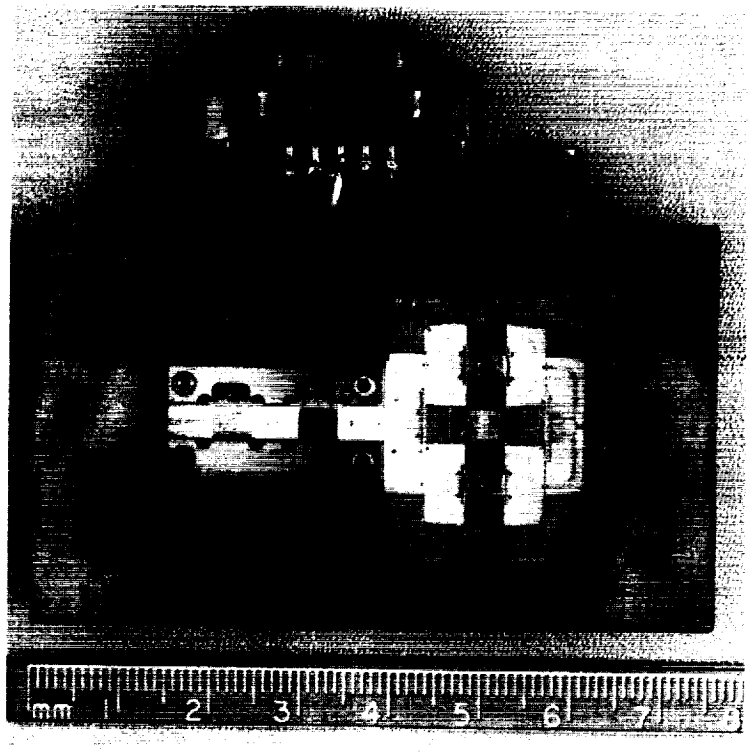


Fig. 11. Photograph of the 32-GHz SSPA showing the RF component side.

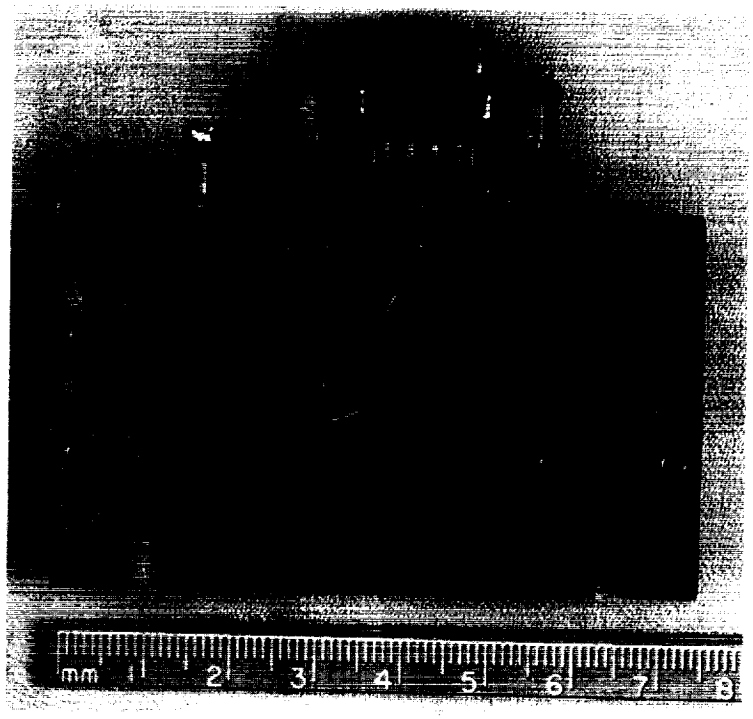


Fig. 12. Photograph of the 32-GHz SSPA showing the dc component side and the waveguide input and output ports.

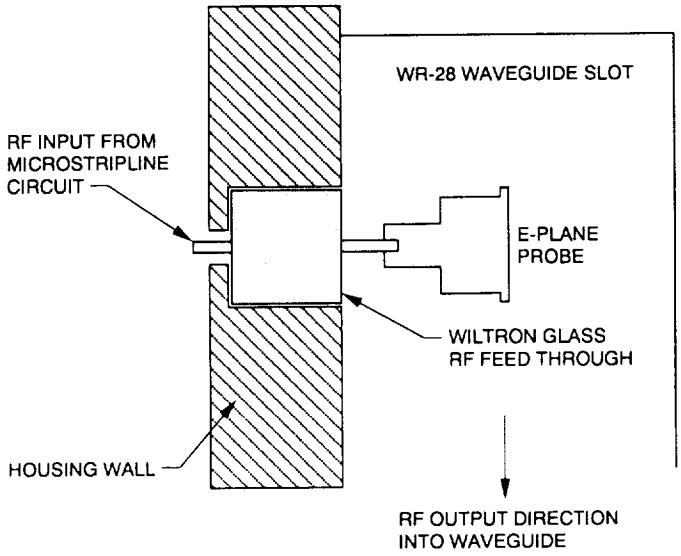


Fig. 13. E-plane microstrip-to-waveguide transition cross-sectional view.

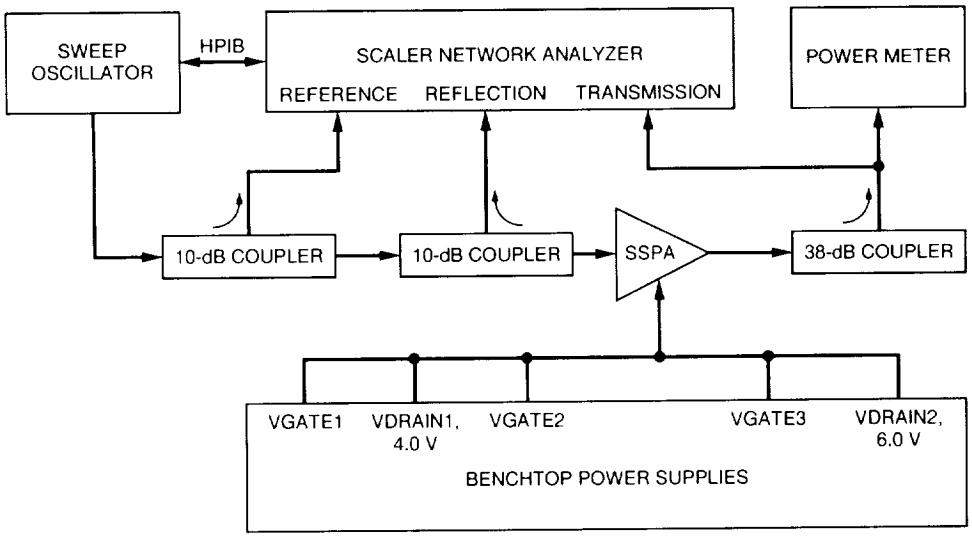


Fig. 14. The measurement system.

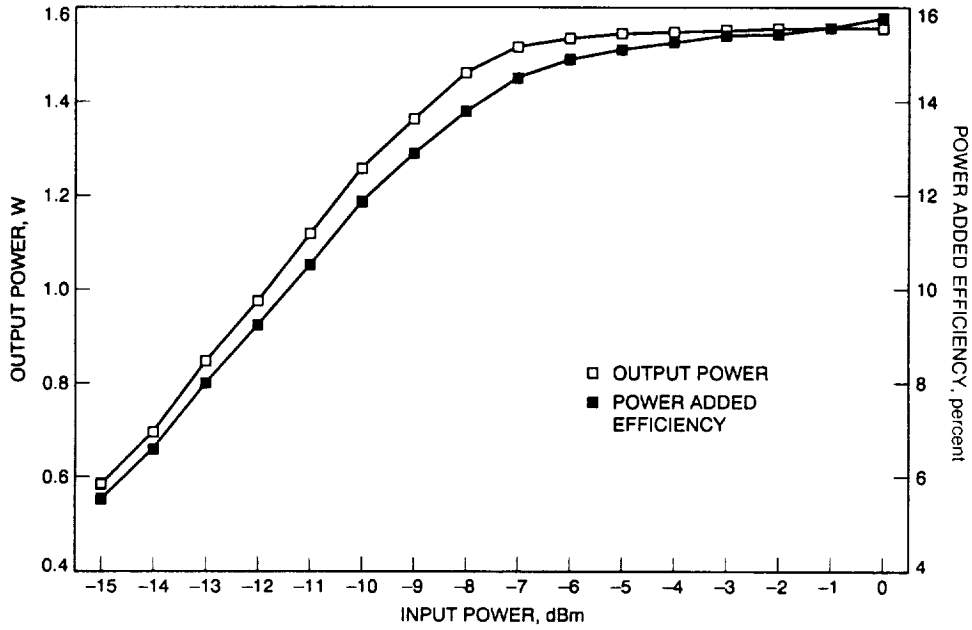


Fig. 15. Output power and power added efficiency plotted against input drive level for the 32-GHz SSPA.

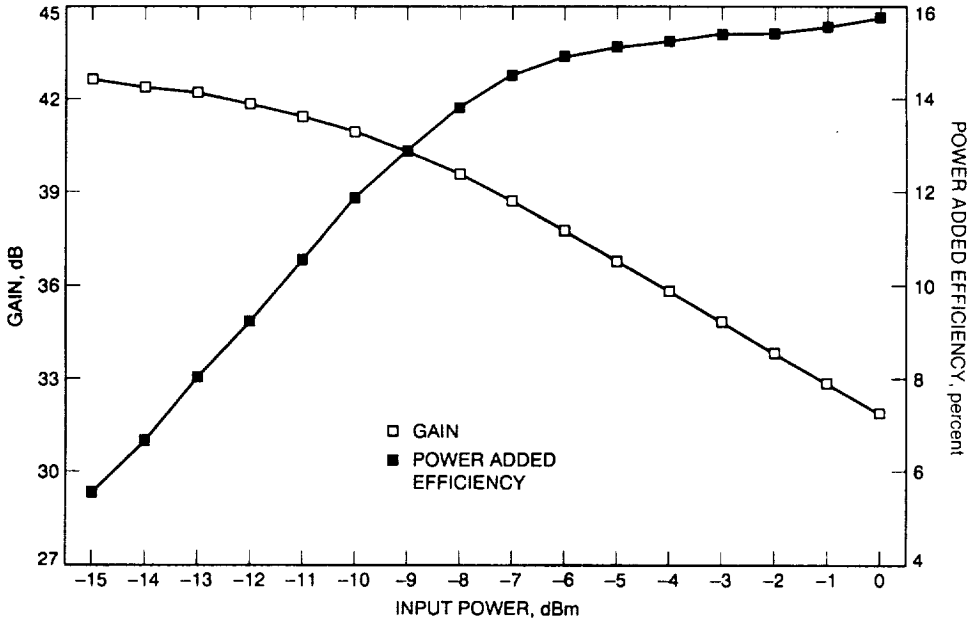


Fig. 16. Amplifier gain and power added efficiency plotted against input drive level for the 32-GHz SSPA.

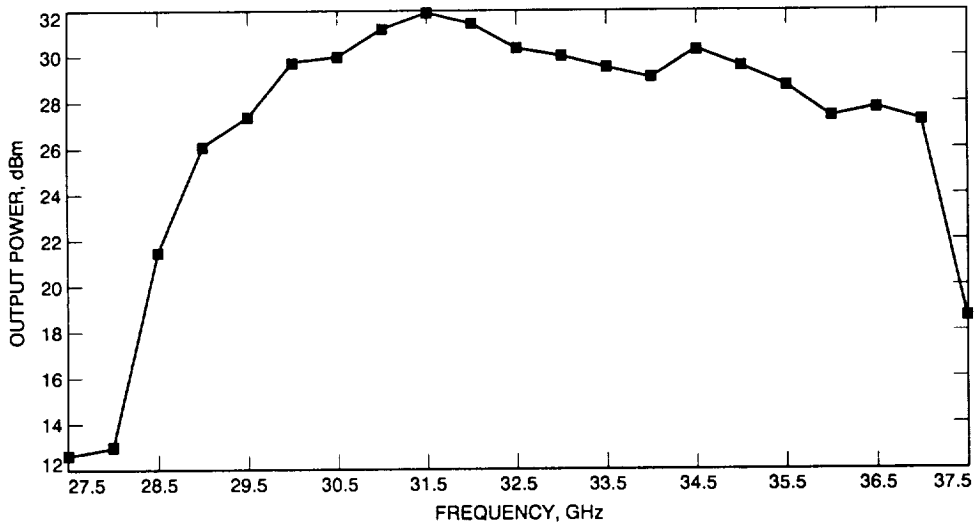


Fig. 17. Frequency response for the 32-GHz SSPA. The peak power was tuned at 31.5 GHz.

The Mars Observer Ka-Band Link Experiment

T. A. Rebold, A. Kwok, and G. E. Wood
Telecommunications Systems Section

S. Butman
TDA Technology Development

The Ka-Band Link Experiment was the first demonstration of a deep-space communications link in the 32- to 35-GHz band (Ka-band). It was carried out using the Mars Observer spacecraft while the spacecraft was in the cruise phase of its mission and using a 34-meter beam-waveguide research and development antenna at the Goldstone complex of the DSN. The DSN has been investigating the performance benefits of a shift from X-band (8.4 GHz) to Ka-band (32 GHz) for deep-space communications. The fourfold increase in frequency is expected to offer a factor of 3 to 10 improvement (5 to 10 dB) in signal strength for a given spacecraft transmitter power and antenna size. Until recently, the expected benefits were based on performance studies, with an eye to implementing such a link, but theory was transformed to reality when a 33.7-GHz Ka-band signal was received from the spacecraft by DSS 13. This article describes the design and implementation of the Ka-Band Link Experiment from the spacecraft to the DSS-13 system, as well as results from the Ka-band telemetry demonstration, ranging demonstration, and long-term tracking experiment. Finally, a preliminary analysis of comparative X- and Ka-band tracking results is included. These results show a 4- to 7-dB advantage for Ka-band using the system at DSS 13, assuming such obstacles as antenna pointing loss and power conversion loss are overcome.

I. Introduction

Theoretical studies in DSN telecommunications have shown that utilizing Ka-band (32 GHz) over X-band (8.4 GHz) on a spacecraft-to-ground link would yield a benefit of 3 to 10 times the telemetered data for a given spacecraft transponder weight, antenna size, and power allocation [1,2].[1] The enhancement comes from the increased antenna gain at smaller wavelengths, but it is reduced by other factors such as higher atmospheric noise, antenna performance deficiencies, and weather susceptibility at Ka-band.

Just as past transitions in the DSN from L-band (0.96 GHz) to S-band (2.3 GHz) and later to X-band yielded almost 20-dB improvements in link capacity, the transition to Ka-band appears to be the next logical step in the evolution of the DSN. The possibility of enhanced performance led to the proposal of a Ka-band link experiment

[1] R. L. Hortorr, ed., *Ka-Band Deep Space Communications*, JPL D-4356 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 15, 1987.

to verify these studies [3] and to discover any obstacles that may prevent this gain from being achieved.

The experiment was taken up by the Mars Observer Project, originally on a best-effort basis and later with full commitment to carry on board the spacecraft a small Ka-band transmitter that could meet the proposed goals. As it became a reality, the Ka-Band Link Experiment (KaBLE) acquired the following specific objectives:[2]

(1) Verify the performance level of a Ka-band link versus an X-band link.

(2) Obtain 250-bit/sec telemetry data over the Ka-band link.

(3) Determine X-/Ka-band differential range to the spacecraft using the ranging code modulated on the X- and Ka-band carriers.

(4) Obtain experience in actual operations of a Ka-band link at a 34-meter antenna.

Mars Observer (MO) was launched on September 25, 1992. According to the KaBLE plan,[3] comparative X- and Ka-band observations would be performed weekly beginning with the Outer Cruise phase in January of 1993, when the spacecraft switched from the low-gain X-band only antenna to the high-gain antenna (HGA) and began transmitting both at Ka-band and X-band. Observations would continue beyond Mars Orbit Insertion to the completion of the Mars Observer mission in 1996. Due to the very weak signal level, the telemetry (TLM) demonstration could be performed only during the first two weeks of the experiment. Ranging was originally planned for the Mars opposition in January of 1995, but was later rescheduled to the period immediately following the telemetry demonstration. This was done for three reasons: first, to take advantage of signal strength due to relative nearness to Earth; second, to practice for a possible test of general relativity during solar conjunction starting mid-December 1993; and third, to diminish the chance of missing the ranging opportunity entirely, should the as-yet unproven KaBLE system on the spacecraft fail before the first Mars opposition in 1995.

This article summarizes the main accomplishments of the experiment, including the design of a Ka-band downlink, the implementation of a Ka-band tracking station at DSS 13, the successful acquisition of Ka-band telemetry and ranging data, and the routine carrier tracking performed from January 1993, to the unexpected loss of Mars Observer in August. A preliminary analysis explores some of the data processing that has been accomplished to relate the performance of Ka-band to the performance of X-band. Finally, future Ka-band performance studies in the absence of Mars Observer will be discussed.

II. Ka-band Link Design

A. Spacecraft Contribution

Mars Observer was equipped with a special 0.033-W transmitter and a 28-cm antenna to support the KaBLE with minimal impact on spacecraft power consumption and antenna design. The Ka-band primary reflector was actually the reverse side of the X-band subreflector, as seen in the spacecraft photograph and block diagram in Fig. 1. The Ka-band downlink was obtained from the 20-W X-band downlink using a 14-dB coupler and a times-four frequency multiplier. The output of the multiplier was physically routed via waveguide along a subreflector support leg to a focal-point feed on the Ka-band antenna. This design, evolved by Martin Marietta Astro Space (Astro), the spacecraft contractor, underwent several changes to increase the Ka-band power output from 0.012 W and to prevent overheating of the multiplier while in near-Earth cruise. The additional 4.4 dB of power margin proved to be very important in compensating for unexpected pointing loss and adverse weather conditions during the telemetry demonstration.

Table 1 derives the effective isotropic radiated power (EIRP) for Mars Observer at X- and Ka-bands, including nominal circuit losses and antenna pointing errors. Note that because of the difference in antenna diameters at X- and Ka-bands, the Ka-band antenna beam is actually wider than the X-band beam (1.3 deg at Ka-band versus 0.9 deg at X-band). Although the two antenna beams were about 0.4 deg off center, they were nevertheless aligned to less than 0.1 deg. As long as the high-gain antenna was pointed to Earth, the full X- and Ka-band power would be received. As it turned out, the high-gain antenna upon deployment was mispointed by 1 deg, resulting in a significant loss of received signal power at both frequencies. This will be discussed further in Section VI.

Because the Ka-band transmitter was essentially a times-four frequency multiplier, the modulation indices

[2] S. A. Butman and J. G. Meeker, *DSN Advanced Systems-Mars Observer Ka-Band Link Experiment Plan*, JPL-D 8799 (internal document), Jet Propulsion Laboratory, Pasadena, California, August 1991.

[3] Ibid.

for telemetry and ranging are four times higher at Ka-band than at X-band. For example, a modulation index of 45 deg at X-band becomes 180 deg (or pure carrier) when the signal frequency is multiplied by four. Spectra taken from Mars Observer during the electromagnetic emissions testing in Fig. 2 show the Ka-band signal that results from X-band telemetry modulations of 0, 44.8, 59.9, and 80 deg. As seen in the figure, the Ka-band spectrum at 44.8 deg (X-band modulation index) is nearly identical to the spectrum at 0 deg. This was particularly helpful during most of the KaBLE tracking, because it allowed the transmission of X-band telemetry from the spacecraft without suppressing the Ka-band carrier. For the telemetry demonstration, the optimum X-band modulation index for obtaining 250-bit/sec data at Ka-band was 59.9 deg, which multiplies to 239.6 deg (with the equivalent suppression of 60.4 deg) at Ka-band. However, it was necessary to gain the confidence of the mission designers and the approval of project management to operate the spacecraft in these modes.

B. Ground Station Contribution

The DSS-13 ground station is a prototype for a future DSN operational subnet of beam-waveguide antennas. Unlike Mars Observer, which was modified to carry a small Ka-band reference beacon, DSS 13 was designed mainly to focus on Ka-band performance. One of the reasons for building a beam-waveguide antenna was to allow locating the Ka-band front-end in the protected, easily accessed environment of the antenna's pedestal room. As will be seen, this feature also proved critical to the experiment's success during the adverse weather of the telemetry demonstration.

Figure 3 shows a high-level block diagram of the DSS-13 feed system from the antenna structure to the receivers. Table 2 shows the antenna gain/temperature (G/T) budget, which includes contributions from the maser, dichroic plate, antenna structure, atmosphere, and cosmic background radiation. A theoretical plot of G/T versus time (and elevation angle) is shown in Fig. 4 for a KaBLE pass in January 1993 (telemetry experiment) under weather conditions ranging from 0- to 90-percent worst-case likelihoods.

The complete link budget for the last day of the telemetry demonstration is shown in Table 3 for both X- and Ka-bands, using the EIRP derived in Table 1. This budget predicts a link margin of 2.1 ± 1.4 dB for telemetry reception. The 36-dB difference between the two downlinks becomes apparent when the total power–signal-to-noise ratio (P_t/N_0) at Ka-band is compared to X-band. Part of the task of the KaBLE is to normalize the results to form a fair comparison between the two downlink frequencies.

The above link budget depends to a large extent on spacecraft-to-Earth range, which varies considerably over the experiment. Figure 5 shows the expected variation in received Ka-band signal-to-noise ratio during the planned duration of the experiment. The dotted line at 29 dB-Hz indicates the minimum level required to obtain 250-bit/sec data at a modulation index of 60 deg; other marks indicate the minimum level for tracking the carrier under various spacecraft configurations. Even though carrier tracking could be performed throughout the 3-year experiment, the entire telemetry demonstration had to fit within a narrow 2-week window beginning the day after the high-gain antenna was activated. Further along, in January of 1995, an opportunity to obtain more data would have occurred, but only at a reduced 10-bit/sec rate.

III. Ground Station Design

The 34-m beam-waveguide antenna at DSS 13 was developed in two phases. In phase I, the antenna was designed, constructed, and rigorously tested for performance at X- and Ka-bands.[4] Transforming the antenna into an operational tracking station for KaBLE took place under Phase II using the requirements summarized in Table 4. Essentially, the station was required to simultaneously track the dual X- and Ka-band downlinks arriving from Mars Observer, demodulate and decode the telemetry data, and store the relevant tracking statistics for future analysis.

A block diagram of the KaBLE system is shown in Fig. 6. From a system level, the ground station can be broken into several subsystems: antenna, microwave, receiver, data acquisition, monitor and control, and frequency and timing. The highlights of each subsystem will be described below.

A. Antenna Subsystem

A primary concern for operating a 34-m antenna at Ka-band is the need for improved pointing accuracy. Whereas for X-band an accuracy of about 8 mdeg is sufficient, for KaBLE an accuracy of 2 mdeg was needed to limit pointing loss to less than 0.5 dB. In the past, pointing calibrations

[4] M. Britcliffe, ed., *DSS-13 Beam Waveguide Antenna Project, Phase I Final Report*, JPL D-8451 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 15, 1991.

had been shown to correct errors to about 5 mdeg rms in a "blind" or open-loop mode.[5] The task of reducing the pointing errors fell to two approaches: (1) improvement of blind-pointing calibration models and (2) use of an automatic boresight algorithm known as AUTOBORE, which corrects pointing errors in real time.

To improve the blind-pointing model, calibrations were performed primarily on two radio sources lying in close proximity to the Mars Observer declination at the start of KaBLE. Since the antenna beams for X- and Ka-bands are not perfectly co-aligned, data were taken at both frequencies. The X-band calibrations, completed one month prior to the start, yielded very promising results of approximately 2.8 mdeg rms (at X-band).[6] However, several factors—including the length of time needed to collect data, last-minute corrections to the system, and bad weather—prevented the completion of a thorough Ka-band calibration. Instead, the initial KaBLE tracks utilized the X-band model. Refinement of the Ka-band pointing model was left to the period after the telemetry experiment, when the station was in a more stable configuration.

AUTOBORE performs a scan in both the elevation and cross-elevation axes, centered on the target, and uses measurements of signal level to determine and remove the pointing error. For natural radio sources, the signal-level estimate is simply the temperature of the source as provided by the total power radiometer (TPR). For spacecraft tracking, the parameter of interest is received carrier power (P_c). The boresight algorithm was modified to obtain P_c by reading carrier-to-noise ratio estimates (P_c/N_0) from the advanced receiver (ARX-II) and multiplying these by noise temperature estimates (N_0) from the TPR.

During normal operations, both the pointing model and the boresight algorithm were employed. A typical worst-case acquisition scenario (during heavy overcast or when the Ka-band signal was very weak) was handled in several steps: (1) the antenna was pointed using the Ka-band pointing model, (2) the ARX-II acquired X-band, (3) automatic boresighting removed X-band errors, (4) the second ARX-II acquired Ka-band, (5) automatic boresighting removed Ka-band errors, (6) the antenna tracked the target with corrected pointing offsets, and (7) automatic boresighting removed errors every hour. When the signal was strong, acquisition of Ka-band usually took place in step 2.

B. Microwave Subsystem

The microwave subsystem at DSS 13, as shown in Fig. 3, takes advantage of the pedestal room layout to allow switching among different feed packages arranged in a ring around a rotating ellipsoid reflector. The feed package used for KaBLE consists of an X-/Ka-band dichroic plate, an X-band high-electron mobility transistor (HEMT) low-noise amplifier (LNA), and a Ka-band maser. The main accomplishments in this area were the fabrication of the dichroic plate and development of JPL's first Ka-band cavity maser.

The X-/Ka-band dichroic plate is essentially transparent at Ka-band and reflective at X-band. It was designed primarily to study performance at the DSN-allocated Ka-band frequencies of 32 GHz (to receive) and 34 GHz (to transmit), and only secondarily for KaBLE. The KaBLE downlink of 33.66 GHz fell at a rather poor location between the two well-matched bands. Thus, out of the 11 K that the dichroic plate contributes to the KaBLE Ka-band system noise temperature, perhaps 6 or 7 K would be eliminated at a more optimal frequency.

Probably the most critical task of the KaBLE was the construction of JPL's first Ka-band cavity maser. To maximize the sensitivity of the system at Ka-band, an ultralow-noise amplifier (ULNA)—a cavity maser with a cryogenically cooled feedhorn, waveguide components, and a postamplifier—was developed with a noise temperature of 5 K at the feedhorn aperture. The maser itself was cooled to 1.6 K with a double-Dewar, open-cycle cryogenic system. The system noise temperature at Ka-band using the ULNA was about 41 K on a clear day. This is somewhat higher than the requirement in Table 4 of 33 K, mainly due to the extra contribution from the dichroic plate. Figure 7 shows a cross section of the maser cryostat with the actual maser embedded in the inner Dewar. The design and performance of the KaBLE ULNA are described in detail in a separate article [4].

For X-band downlink reception, a simple HEMT LNA was installed, yielding a system noise temperature of about 37 K. One additional LNA, a Ka-band HEMT with a system noise temperature of 65 K, was used to back up the maser.

[5] L. Alvarez, "Pointing Accuracy Measurements," in *DSS-13 Beam Waveguide Antenna Project, Phase I Final Report*, JPL D-8451 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 15, 1991.

[6] L. Alvarez and C. Racho, "Estimated DSS-13 Ground Antenna Pointing Performance During the KaBLE Telemetry Demonstration," Interoffice Memorandum 3324-93-032 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 18,1992.

C. Receiver Subsystem

The receiver subsystem for both X- and Ka-band consists of an RF–IF downconverter, a fiber-optic transmission link from the pedestal room to the control room, IF distribution in the control room, and an ARX-II—a digital receiver used for prototyping advanced receiver concepts (Fig. 6). For the X-band path, downconversion is performed using the 8.1-GHz first local oscillator (LO) that is a standard for the DSN. For the Ka-band path, a two-stage downconversion is employed, mixing the incoming 33.7-GHz nominal downlink signal with a 25.2-GHz first local oscillator followed by an 8.2-GHz second local oscillator. All local oscillators are coherent with the station frequency and timing subsystem (FTS).

It was found early on in the design of the system[7] that the Doppler jerk (rate of acceleration) from the spacecraft while in orbit around Mars, together with the weak 10-dB-Hz signal level expected, would (at Ka-band only) exceed the tracking capability of any third-order phase-locked loop, including that of the ARX-II. To remedy this problem, a device known as the Doppler tuner was developed to remove the predicted Doppler profile from the Ka-band signal prior to the ARX-II.

At the heart of the Doppler tuner lies a programmable oscillator (PO) designed by the Goldstone Solar System Radar Group (Fig. 8). The PO works in conjunction with a synthesizer, accepting time-coded Doppler predict files and adjusting a reference to the synthesizer to match the Doppler frequency at any given time. The Doppler correction converter, after an initial upconversion, mixes the IF with the output of the synthesizer, thus providing a 300-MHz fixed-frequency Ka-band IF to the ARX-II.

To fulfill KaBLE's requirement for two advanced receivers, a second receiver, intended for permanent location at DSS 13, was built. Known as TP-13 (for DSS-13 telemetry processor), this receiver was equipped with faster signal-processing and monitor hardware, making it the receiver of choice for Ka-band. In terms of function, the two receivers are essentially interchangeable and can be used to back up each other: in the event of a hardware failure, the functioning unit would be used for Ka-band.

Finally, one extra level of redundancy was provided by the multi-tone tracker (MTT), a digital receiver particularly suited for KaBLE. Because of its multichannel capability, the MTT can track X- and Ka-bands simultaneously, and also use the phase of the X-band carrier to aid the tracking of the Ka-band carrier. This technique takes advantage of the coherent, factor-of-four relationship between the X- and Ka-band downlink frequencies, allowing the detection of very weak Ka-band signals and tracking with very narrow loop bandwidths. Carrier aiding is even more effective than the Doppler Tuner in removing Doppler from the Ka-band signal, because the strong, easily tracked X-band carrier offers real-time information on what the signal is actually doing, rather than what it is predicted to do.

D. Data Acquisition Subsystem

Most of the remainder of the ground subsystem falls under the broad category of data acquisition. Here the relevant link parameters are measured and recorded for future analysis. The instruments included in this subsystem are the TPR for measuring system noise temperature; the water vapor radiometer (WVR) for measuring the quantity of water in the atmosphere; the weather station; and the data handling terminal (DHT) for measuring telemetry performance, displaying data from the other instruments, and recording all KaBLE-specific data to disk.

The TPR operates together with the microwave switch controller (USC) in the Microwave Subsystem to perform calibrated system noise temperature (SNT) measurements on a continuing basis during a track. Measurements of the total IF noise power are made with two HP436A power meters—one each for the X- and Ka-bands. A local terminal reads these measurements and converts them to SNT values at a rate of once per second, based on a transfer function derived from a previous calibration.

Calibrations can be initiated by operator command at any time during a track. A total of six calibrations are performed with the antenna at zenith before and after a track, and the results are averaged to form a single calibrated transfer function. During each calibration, the TPR measures the total IF noise power while the USC switches the maser input from sky, to noise diode, to ambient load, to both noise diode and ambient load. The measurements are then converted into a transfer function (SNT as a function of total noise power) and an estimate of system nonlinearity.

The WVR is a small, stand-alone package positioned some distance from the antenna, which calculates the presence of atmospheric water in vapor and droplet form along the antenna beam by measuring noise levels at 20.7 and 31.4 GHz. It consists of a platform with a small feedhorn and mirror assembly that can observe any point in the sky

[7] T. Rebold, "Tracking KaBLE Doppler Accelerations," Interoffice Memorandum 3393-91-167 (internal document), Jet Propulsion Laboratory, Pasadena, California, November 11, 1991.

with a beamwidth of 7 deg. It performs periodic tipping curves from zenith to 20-deg elevation and, when provided with pointing predicts, tracks along the line of sight of the 34-m antenna.

Data collected by the WVR can be used to understand the Ka-band tracking results by indicating the condition of the atmosphere in the direction the antenna is pointing. When a cloud passes overhead, for instance, the WVR will show the increase in sky temperature, from which the expected drop in Ka-band signal level can be estimated. Eventually, the WVR will be used to remove atmospheric effects, such as fluctuating path length delays, from the tracking data.

The weather station samples and records a range of meteorological parameters, including temperature, humidity, wind speed, and wind direction. Like all the instruments in this subsystem, data are stored locally and routed through monitor and control to a central storage device: the data handling terminal.

During operations, the DHT is routinely monitored for insight into the state of the downlink. Such data types as carrier-to-noise ratio, SNT, antenna elevation angle, and wind speed can be displayed simultaneously versus time on a single monitor. In addition, for tracks demonstrating Ka-band telemetry reception, the DHT calculates and displays the telemetry bit-error rate (BER) of the Ka-band data using the virtually noise-free X-band data for reference. Most data arrive at the DHT through the station data recorder (SDR), a gateway terminal that serves to back up data archiving and relieve some of the real-time load from the DHT. A second interface provides the DHT with two channels of 250-bit/sec telemetry data from the receivers' convolutional decoders.

E. Monitor and Control Subsystem

Tying all subsystems together into an operational system with a high degree of flexibility is the task of the monitor and control (M&C) subsystem. The main strength of the DSS-13 M&C subsystem is its user interface, a window-driven software display configured by the operators rather than the developers. The window approach provides both a high-level view of the entire system and a detailed configuration table for each individual subsystem. In this approach, flexibility is guaranteed: When a new project is assigned to the station (such as KaBLE), a new set of displays are developed by the operators for the particular subsystems used by a project. Then, at the start of a new track, the appropriate display package is loaded into memory and configuration is begun. Naturally, this approach lends itself well to automation, and an effort is under way to allow automated remote operations of the station after normal shift hours.

The M&C subsystem currently runs on a 486-platform. Eventually, the entire software package will be transferred to a Sun workstation. Interfacing to the other subsystems is provided by a local area network (LAN), a token bus easily modified to include new subsystems as they are added to the system. Each subsystem's host computer taps into the LAN through a special board plugged into its chassis, and a driver for handling real-time commands is loaded into the host computer's RAM at power-up.

F. Frequency and Timing Subsystem

KaBLE was able to benefit from the maturing fiber-optic technologies on two fronts. In the area of frequency and timing, all of the sensitive reference frequency distributions were made with fiber. In fact, even though DSS 13 has a hydrogen-maser frequency standard, the prime standard for the station is located about 20 miles away in the control room of Signal Processing Center (SPC)-10. With fiber optics, the degradation of the standard over that distance is practically immeasurable. Distribution of the 100-MHz first local oscillator to the downconverters in the pedestal room is also performed using fiber. Even many instruments in the control room are now given fiber-optic frequency references. The result is an exceedingly stable frequency reference for the entire system, which happens to be coherent to SPC-10 as well.

The second use of fiber optics is in the transmission of broadband IF signals. A fiber line connecting the pedestal room to the control room improves upon the instability normally obtained from coaxial cable. But the real advantage came with the installation of an IF link from DSS 13 to SPC-10, allowing use of the DSS 13 X-/Ka-band front-end with the two Pioneer-10 receivers located at SPC-10, where they could be interfaced to the ranging and Doppler equipment. Thus, for the modest cost of four fiber-optic transducers, a number of valuable possibilities were enabled, including a backup telemetry string and a viable Ka-band ranging system.

IV. Testing

System testing for KaBLE began in July of 1992 and continued in blocks as individual subsystems were installed

or upgraded all the way to the start of the telemetry demonstration in January of 1993. Antenna time for KaBLE was sandwiched between the numerous other tasks supported by DSS 13, including the High Resolution Microwave Survey (HRMS) and the Array Feed Project.

System performance tests, covering the period from July to August of 1992, focused on the basic functions of the KaBLE system: verification of system linearity and radiometric measurements, acquisition and tracking of frequency-ramped carriers, and acquisition and decoding of telemetry data. An important side benefit of these tests was getting acquainted with subsystem peculiarities and bringing software bugs to the surface for troubleshooting and repair.

A series of tracks were performed in August using the Voyager and Magellan downlinks as test signals. These tracks, although X-band only, were close approximations to a KaBLE track in terms of Doppler rate, signal level, and antenna pointing. As the first tracks with real signals, they were very helpful in proving not only receiver and telemetry performance but also the accuracy of the predict generation software. In addition, they offered the first opportunity to verify the AUTOBORE pointing correction algorithm on spacecraft signals. As the experiment start date approached, tracks of Mars Observer at X-band over the low-gain antenna were repeated for the operations exercise approximately once every two weeks.

Among the problems that arose during testing, one of the more serious was the lack of proper subsystem interfacing over the station LAN. This resulted in subsystem crashes whenever a complex task was performed, in particular during the AUTOBORE sequence where the antenna monitor and control (AMC) configures and reads data from the advanced receiver and the total power radiometer at the same time. Since most of the subsystems were able to function perfectly well in stand-alone mode, it was not a fatal problem. But to get the antenna to boresight properly, a great deal of time was invested in debugging the interfaces.

Shortly after the installation of the Ka-band ULNA in late November, a gain oscillation was observed in the Ka-band path, with a period on the order of 2 min and an amplitude of about 5 K. The source of the oscillation, first thought to be the ULNA, was later isolated to the downconverter. It turned out that a temperature controller in the Ka-band downconverter assembly had, due to a power glitch, reset its control values to a stored default, causing the temperature, and hence gain, in the assembly to oscillate. (As seen in Section VI, the maser gain and noise temperature were also prone to fluctuate.)

All this had the effect of delaying quality antenna calibrations at Ka-band until several days prior to the start of the experiment. By then the weather was already routinely cloudy and wet, and good data could not be taken at Ka-band. As a result, the X-band pointing model was used during the first several KaBLE tracks until accurate calibrations could be made at Ka-band. Because of this, pointing losses at Ka-band grew rapidly between boresights—about 0.75 dB/hr on average.[8]

V. KaBLE Operations

A. Predicts Generation and Delivery

The task of routine spacecraft tracking at an R&D station like DSS 13, not to mention the unique requirements of tracking at Ka-band, prompted the development of a special predicts-generation software package known as PG-13. The basic requirements for PG-13 were to provide time-tagged antenna pointing predicts accurate to 0.1 mdeg (to be consistent with the 2-mdeg overall budget) and X- and Ka-band frequency predicts accurate to within 100 Hz (to fall within a convenient acquisition range of the ARX-II).

Testing of PG-13 was performed both by comparing its output with Network Operations Control Center support system (NSS)-derived predicts and by measuring predict errors during test tracks of Voyager, Magellan, and Mars Observer. Eventually the level of accuracy was refined well beyond requirements: antenna predicts to much less than 0.1 mdeg and frequency predicts to about 1-Hz accuracy. This level of accuracy requires close attention to detail and the maintenance of a database of station locations, planetary ephemeris, spacecraft trajectory, spacecraft ultrastable oscillator (USO) frequencies, and uplink exciter ramps, which must be updated on a regular basis.

In the case of one assembly, the Doppler tuner, the frequency predicts provided by PG-13 were transformed into a series of Chebychev polynomial expansions using a standard algorithm [5]. Another assembly, the multi-tone tracker, was provided predicts directly by its developers.

The delivery of predicts from JPL to the DSS-13 M&C was nearly instantaneous over the NASA Science Internet (NSI) connection to Goldstone. From there, predicts were either downloaded by the station operator to the destination subsystem or typed into the appropriate window display on monitor and control.

[8] L. Alvarez and C. Racho, op. cit.

B. Station Operations

During a track, station operations were governed by a sequence-of-events (SOE) file, listing initial station and spacecraft configurations as well as the key events for the track. A typical track consisted of configuring the station for KaBLE, downloading predict information to the appropriate subsystem, performing noise temperature precalibrations at zenith, acquiring the signal, boresighting the antenna, and occasionally reacquiring the signal during uplink transfers with the spacecraft. After the track, data files were transferred from the data handling terminal to M&C, where they could be accessed over Internet from JPL.

The goal was for system operations to be under the control of one person. Ideally, all commands would be sent from M&C over the station LAN to the various subsystems. In practice, however, actual operations were more labor intensive than planned because of the many problems associated with a new system; typically, these operations required the attention of two operators and the experiment engineer. Just about every subsystem had its bad days, resulting in some unique station problem that had to be debugged before the track could proceed. Until the problematic AMC/M&C interface was refined, antenna boresights were routinely commanded from the local AMC terminal in the pedestal room. Fortunately, the requirement that each subsystem be able to operate locally made the system flexible enough to handle most contingencies.

Even so, KaBLE operations sometimes presented unusual challenges. In one track, a power glitch during precalibrations kept the operators busy recooling the Ka-band HEMT to bring it back online as quickly as possible. In another, the first local oscillator lost lock with the frequency reference and ran noncoherently with the rest of the station. On one occasion, because of a receiver problem, the operator at DSS 13 had to perform manual boresights of the antenna using carrier estimates from a Pioneer-10 receiver located at SPC-10.

Eventually, through much perseverance on the part of station personnel and subsystem engineers, the system irregularities and interface problems were eliminated. Towards the end of KaBLE, operations took on a more routine nature, and preparations were underway to handle the more operationally challenging mapping phase of the Mars Observer mission when Doppler dynamics would increase a hundredfold and carrier reacquisitions would occur about eight or nine times per track.

VI. Experiment Results

As discussed earlier, KaBLE was divided into three tasks: a telemetry demonstration, a ranging demonstration, and a tracking experiment. The goal of the telemetry demonstration was to prove the feasibility of a telemetry link at Ka-band by collecting a minimum of 1 million bits of quality 250-bit/sec data. The goal of the ranging experiment was to perform a similar demonstration of Ka-band ranging. The goal of the tracking experiment was to develop a database of Ka-band link statistics, including weather, noise temperature, and antenna performance, for use in the development of future missions and ground stations at Ka-band. For the most part, all three goals were achieved by KaBLE. Both telemetry and ranging demonstrations were successful. However, because of the untimely end of Mars Observer, the tracking experiment was able to acquire only a partial database of antenna performance. The results of the three phases of KaBLE and a preliminary analysis are discussed below. Further study of Ka-band antenna performance is covered in the next section.

A. Telemetry Demonstration

A certain amount of difficulty is normally expected whenever a new technology is demonstrated, so it is not surprising that a number of obstacles arose during the KaBLE telemetry demonstration. The first obstacle was the weather. By an unfortunate coincidence, the telemetry demonstration took place during one of the worst weather systems to ever hit the Goldstone area. Heavy rains, flashfloods, and dense fog were daily occurrences. During one track, a blanket of wet snow covered the antenna and ground.

This type of weather has a major impact on Ka-band performance, since the Ka-band is adjacent to the water absorption lines in the RF spectrum. Noise temperature at Ka-band seldom dropped below 90 K during these 2 weeks, and sometimes, at low elevations, reached as high as 250 K. This, in itself, should not eliminate the use of Ka-band in the DSN, since weather of this nature is so rare (99.9-percent probability of not occurring compared to the 90-percent weather tolerance allocated in the KaBLE link budget). However, for the duration of the telemetry demonstration, weather was a major limiting factor in data collection.

Given the extreme weather conditions, the use of a beam-waveguide antenna for KaBLE proved to be a wise choice. Even though the dish was saturated most of the time, the Ka-band front-end, in particular the feedhorn

cover, was kept dry in the shelter of the pedestal room. Had an extra layer of water been allowed to collect on the feedhorn cover, as would be the case in a Cassegrain feed antenna, the telemetry demonstration would have collected even less data at Ka-band, perhaps none.

A second obstacle involved the spacecraft. Due to an anomaly in the Attitude Control Subsystem (ACS), Mars Observer delayed switching on the HGA (and hence the Ka-band downlink) until 3 days into the 14-day telemetry demonstration. Furthermore, once the HGA was turned on, an antenna calibration maneuver revealed a large (1.1-deg) pointing error. This resulted in an unplanned loss of signal level—about 6.7 dB at X-band and 3.5 dB at Ka-band as derived in Section VI, which the spacecraft team could not correct until after orbit insertion.

On the first night of tracking, a final obstacle was encountered: a polarization mismatch resulting in a 10- to 15-dB loss in signal level at Ka-band. It took several days of testing to recognize the problem and isolate the source. The tests that were performed included injecting right- and left-hand circularly polarized (RCP and LCP) signals into the KaBLE ULNA, switching to the backup Ka-band HEMT, and inserting a half-wave plate into the signal path while tracking to reverse the polarization of the downlink signal. The last test proved the most conclusive, resulting in about 10 dB of signal gain with the half-wave plate installed. But to achieve the desired minimum noise temperature, a more permanent solution was needed. With only 4 days remaining before the end of the demonstration, the ULNA was warmed, reconfigured to the correct RCP polarization, and brought back on line, all within a 36-hr period. Again, easy access to the Ka-band front-end provided by the layout of a beam-waveguide antenna proved critical to the eventual success of the demonstration.

While the polarization problem was being corrected, meetings were held at JPL with the Mars Observer team to incorporate "on-the-fly" KaBLE input into the configuration of the spacecraft. Because of the weather and the increased spacecraft pointing loss, the signal hovered very near the 250-bit/sec threshold. A plan was implemented to allow the KaBLE principal investigator to decide, based on the weather forecast at 4 p.m., whether to proceed with a 250-bit/sec track or a reduced 10-bit/sec track (for higher signal margin). A phone call to the spacecraft team at 4 p.m. would result in a change in spacecraft configuration 2 hr later; this change would then have to remain constant for the duration of the track.

During the remaining 3 days, the weather continued wet and cold, but the 250-bit/sec configuration was chosen for all 3 days. On the first day, the signal never cleared the 250-bit/sec threshold. On the second day, station personnel pulled the dichroic plate in the hope that some data might clear the threshold because of the reduced Ka-band noise temperature, even though no real-time confirmation of Ka-band data quality would be possible without X-band data. Shortly after the dichroic plate was reinstalled, the weather eased and the first burst of good data was collected. In spite of 3 dB of extra spacecraft pointing loss and persistently bad weather, a total of 3 hr of good data (about 3 million bits) was collected on that day. On the last day, the signal remained below the 250-bit/sec threshold for the duration of the track.

A plot of tracking statistics for the successful telemetry track is shown in Fig. 9. The carrier-to-noise ratio (P_c/N_0), SNT, and antenna elevation are shown in comparison with the received bit-error rate. The gap in X-band SNT covers the period when the dichroic plate was removed. The first burst of good data, which came after the plate was reinstalled, lasted only about 6 min, at which point the spacecraft changed mode from three-way operation with DSS 65 uplinking to one-way operation; the weather worsened, and antenna efficiency dropped as the elevation approached 70 deg. It was not until the weather eased and the elevation neared the rigging angle (50 deg) that the data improved again, this time for 3 hr. The bit-error rate ranged from 50 percent (or -3 on the log scale) at the start of the track to about 1 percent (or -20) during peak performance, which is adequate proof of the successful acquisition of Ka-band telemetry data.

B. Ranging Demonstration

Work on the KaBLE ranging demonstration began shortly after the telemetry demonstration ended. The goal of this effort was to demonstrate the demodulation of DSN ranging codes at Ka-band. Also, it was hoped that the X-/Ka-band differential range could provide a new data type for radio-science observations near solar conjunction (a hope that was not realized due to the loss of Mars Observer).

A dual-channel ranging system for KaBLE was obtained at virtually no extra cost to the experiment by sending the IF signals from DSS 13 over the existing optical-fiber link to the ranging equipment at SPC-10. The SPC-10 equipment included two Pioneer-10 (P10) receivers, which are prototypes for the Block-V receiver, and a modified sequential ranging assembly (MSRA), a standard SRA modified to interface with the P10 receivers.

Figure 10 shows the ranging system configuration for KaBLE. The heavy lines represent the ranging signal path

for the demonstration. While DSS 15 performed normal two-way ranging with Mars Observer at X-band, the P10/MSRA assembly performed three-way ranging using the X- and Ka-band signals received at DSS 13. Ranging data were logged on a personal computer, which interfaced with the MSRA through a maintenance port.

Two successful ranging passes were obtained for KaBLE. On day of year (DOY) 56, 32 valid range points were collected, while on DOY 76, a total of 60 points were collected, 14 of which were discarded due to antenna mispointing. The ranging modulation was identical for both passes: a 20-sec integration time for the clock component (1 MHz) followed by 10 sec at each of the seventeen successive components for a total cycle time of 3 min and 30 sec per range point. Analysis of the data from these two passes, summarized below, indicates good agreement between the measured ranging signal-to-noise ratio (SNR) and the accuracy of the range estimates.

The ranging SNR (P_r/N_0) for the two passes is shown in Fig. 11. Because the modulation index is four times higher at Ka-band, the difference in X- and Ka-band ranging SNR is only about 20 dB (compared to the 35-dB difference in carrier power). Nevertheless, on DOY 76 the increased fluctuation in P_r/N_0 estimates indicates that the code strength neared threshold for the measurement integration time.

Figure 12 shows the X-/Ka-band differential range for the two passes measured in range units (RU) of the round-trip light time (1 $RU \approx 0.952$ nsec). The bias in the data on the first chart (DOY 56) shows that the Ka-band path was 26 RU longer than the X-band path, which is consistent with the layout of the DSS-13 system. The bias in the second chart shows that the offset increased to 67 RU on DOY 76, which is also consistent with a cable bypass in the X-band path on that day.

The range residuals for each channel are shown in Fig. 13. The residuals are relative to a sixth-order curve fit made on the X-band data. The difference in delays measured for the X- and Ka-band paths can again be identified by the 26-RU and 67-RU offsets between the two curves. The standard deviation of the residuals is listed in Table 5 and compared with the expected standard deviation based on the average ranging SNR (P_r/N_0) measured during the pass. At X-band, the measured sigma is about twice the expected sigma, which may be due to errors in the model used to form the prediction. At Ka-band, there is close agreement between the average P_r/N_0 and the measured sigma, although on DOY 76 the measurement is slightly better than expected. This is probably due to the large fluctuations in P_r/N_0 during that pass.

C. Tracking Experiment

The tracking experiment covered the period from deployment of the Mars Observer HGA in January 1993 to just prior to the Mars Orbit Insertion in August. During this 9-month period, 27 tracks were performed for a total of 291 hr. However, due to the large number of problems that occurred while tracking, only about 164 hr of Ka-band tracking data were obtained.

During the 9 months of the tracking experiment, the KaBLE system underwent frequent changes as new equipment was installed and subsystems were upgraded. The Ka-band maser was removed a few weeks after the telemetry demonstration and the Ka-band HEMT—with higher noise temperature but lower maintenance costs—was installed in its place. The TP-13 Advanced Receiver was swapped in and out to support experiments with the Galileo spacecraft. The Antenna Monitor and Control subsystem was improved with hardware and software upgrades to allow boresighting from the control room and faster response time. Through it all, a viable KaBLE configuration was maintained.

Table 6 summarizes the results of each track. In general, the problems that occurred were due mainly to subsystem failures or interfacing problems, and so the shortfall in data collection does not reflect any limitations in the performance of a Ka-band downlink. As discussed in Section V, the problems that arose during KaBLE tracks were quite varied and numerous. In fact, experience gained with these problems was a valuable by-product of KaBLE and will go a long way toward improving the design and implementation of a Ka-band tracking subnet in the DSN.

On some occasions, KaBLE tracks were limited by a combination of bad weather and poor signal strength. For instance, almost all of the tracks during the telemetry demonstration (DOYs 8 through 17) were affected by the heavy rains. Also, on DOYs 76 and 95, the wind was so high (and the carrier signal so low) that proper tracking was impaired. (On DOY 95 the antenna had to be stowed for several hours until the winds died down.) As an example of the data set collected by KaBLE, tracking results for three separate tracks—DOYs 18, 56, and 197—are shown below.

1. DOY 18. This track, which followed the last day of the telemetry demonstration, was performed with the spacecraft configured for carrier only at Ka-band (i.e., no telemetry). Also, the spacecraft transponder was referenced to the DSS-15 uplink. At the time, a problem with the DSS-15 exciter led to cycle slips in the X- and Ka-band downlinks whenever DSS 15 was uplinking. These cycle

slips, which show very clearly in Fig. 14 as 10- to 20-dB dropouts in received carrier power, tended to interfere with efforts to boresight the DSS-13 antenna. Consequently, much of the track was performed in a blind-pointing mode. The descending arc from hour 27 to 31 represents the accumulation of pointing errors in the blind-pointing model until the boresight at hour 31 corrected the pointing error.

The weather and noise temperature statistics for this track are shown in Fig. 15. On this day, the weather was similar to that during much of the telemetry demonstration in that a layer of heavy clouds rolled in shortly after the track began, bringing fog and rain with them. The weather signature can be seen clearly in the Ka-band system-noise-temperature curve.

2. DOY 56. DOY 56 was the day of the first successful Ka-band ranging demonstration. The tracking results in Figs. 16 and 17 complement the ranging data presented in Part B of this section. As in DOY 18, only a few boresights were performed, so as not to interfere with the ranging measurement. The scalloped signature of the Ka-band received carrier power in Fig. 16 is essentially due to the blind pointing error of the DSS-13 antenna.

The weather for the track was clear with only a thin, high-altitude haze. Humidity is seen to be rising throughout the track, but its effect on carrier power is negligible. The Ka-band noise temperature is about 70 K, due to the higher noise Ka-band HEMT that was installed in place of the ULNA.

3. DOY 197. This track came at the end of a difficult week of tracking. The intention was to obtain a large quantity of high-quality data during one week to make up for a 2-month hiatus in tracking over the summer. To this end, system health was verified during a track on DOY 172, and the Ka-band maser was reinstalled to obtain optimum noise-temperature performance.

During the first day of this week, DOY 193, problems arose with a gateway that had been installed to buffer the AMC from LAN traffic. Efforts to debug the problem took up the next 3 days, during which the antenna could not be boresighted. On the second day, the maser began to experience some difficulties, including a higher than normal system noise temperature (60 to 80 K) and recurring gain fluctuations, the source of which is under investigation. But on Friday, DOY 197, the system temperature returned to about 41 K and the AMC gateway problem was corrected.

Figure 18 shows the tracking data from the DOY 197 track. Minicalibrations of system noise temperature and antenna boresights were performed on the hour, as indicated by the downward spikes in received carrier power. Note that even with the maser, the Ka-band carrier-to-noise ratio was only about 10 dB-Hz, since the spacecraft at this time was approaching opposition. The low signal level results in about 2 to 3 dB of spread in carrier-to-noise ratio estimates.

In Fig. 19, the weather and noise temperature statistics are shown. Here the hourly minicalibrations show as vertical lines in the system noise temperature curves. The remnants of the Ka-band maser gain fluctuations can be seen in the 2- to 3-K difference in noise temperature before and after each minicalibration. The most accurate SNT measurements are those following the minicalibrations. In terms of weather, the track was performed under nearly ideal conditions: clear skies, low humidity, and light winds.

D. Preliminary Analysis of Tracking Data

The basic goal posed at the outset of KaBLE was to determine how Ka-band ranks in comparison to X-band for use in deep-space communications. The question is complex, since there are many different ways of evaluating a communications link. This section reviews some of the analysis that has been performed to place the KaBLE results in a suitable context for further comparisons with X-band.

1. Spacecraft Pointing Loss. The first analysis for KaBLE involved the determination of the spacecraft pointing loss, a critical factor needed for further analyses. Pointing losses were estimated using DSN tracking data during the HGA calibration maneuver on DOY 8. During the maneuver, the spacecraft was offset 0.5 deg from the Earth line and allowed to rotate one full turn (lasting 100 min) while the uplink and downlink signals were observed. Had the HGA been properly pointed along the Y-axis there would be no significant variation in signal level other than an initial step of about -1 dB when the 0.5-deg offset was inserted.

However, what was actually observed in the tracking data from DSS 65 was a sinusoidal signature in carrier power with an amplitude variation of about 15.3 dB. This observation provides an accurate means of determining pointing offset. Figure 20 shows the Mars Observer HGA gain at X- and Ka-bands along with the variation in X-band power that would be observed during the maneuver, assuming a given pointing error. Note that the Ka-band beam pattern is actually wider than X-band because of the smaller Ka-band antenna. Based on this chart, the

measured variation of 15.3 dB implies a pointing offset of 1.1 deg, or a pointing loss of 6.7 dB at X-band and 3.5 dB at Ka-band. This derivation agrees well with similar data on measured uplink variations during the maneuver, and is accurate to about 0.2 dB or less.

2. Comparisons With Theory. One way to gain understanding of the Ka-band downlink, in light of the unanticipated spacecraft pointing losses, is to use measurements of signal level from a KaBLE track to estimate the spacecraft EIRP. This is done in Table 7 for the track on DOY 197. The procedure for deriving EIRP is to work backwards through the link budget, starting with measurements of P_c/N_0 and adjusting for the effect of each parameter. Both P_c/N_0 and system noise temperature were sampled at 50 deg elevation, the rigging angle of the DSS-13 antenna, and inserted into the table along with the pointing-loss estimates derived above, modulation loss estimates (due to ranging and telemetry modulations) obtained from laboratory data,[9] and the rest of the link budget parameters.

The resulting EIRP estimates come very close to those derived in the link budget of Table 3. The X-band estimate is 0.3 dB higher than theory, while the Ka-band estimate is about 1 dB higher. Although somewhat higher than planned, the Ka-band value does fall within the tolerance of the link budget estimate and may explain why the telemetry demonstration was successful in spite of the extra pointing loss. The uncertainty in this estimate is about 1.2 dB at Ka-band (0.4 at X-band) and can be reduced by sampling data from other tracks.

A similar analysis was used to derive the plot in Fig. 21. Here a theoretical estimate of total signal-to-noise ratio (P_t/N_0) versus time was derived from the link budget by varying the antenna G/T as a function of elevation angle. Then the theoretical curve was overlaid upon the actual tracking data, in this case the track on DOY 17 for 0- and 90-percent weather conditions. The result shows a strong correlation between measured P_t/N_0 and theory, with the actual signal appearing about 1 dB below the 90-percent line. At either end of the curve, the measurement falls farther away from theory, which may be due to deterioration of the weather. (See Fig. 9 for more DOY 17 tracking statistics.)

3. Link Normalization. Because of the difference in X- and Ka-band radiated power from the Mars Observer spacecraft, there is a large difference in received carrier power at X- and Ka-bands. One of the first steps toward obtaining a fair comparison is to normalize the two downlinks. This requires the development of correction factors for both ends of the link: the spacecraft and the receiver.

Spacecraft normalization is derived in Tables 8 and 9. Column 2 of Table 8 shows the EIRP budget for the Mars Observer X-band transmitter. Column 3 develops the estimated EIRP for a normalized Ka-band transmitter: in other words, a Ka-band beacon that transmits the same power through the same-sized antenna as X-band. Comparisons to the actual KaBLE beacon can be made with the numbers in column 4.

Several important assumptions have been made in normalizing the Ka-band transmitter. The first and most important is that the power amplifier efficiency on board the spacecraft is the same at X- and Ka-bands. Although current technology does not support this assumption, new developments in power amplifier design are making frequency less of an issue. The second is that antenna efficiency and pointing losses will be the same at both frequencies. All of the assumptions depend upon spacecraft design, and so cannot be considered exact. The point of this exercise is to offer a reasonably fair comparison for the sake of understanding relative link performance.

Table 9 derives the overall spacecraft correction factor by combining the change in EIRP with spacecraft pointing and telemetry losses. The pointing loss was calculated in section VI.D.2. The modulation losses, both ranging and telemetry, were derived from measurements of the actual flight unit in the laboratory.

The normalization of the receiver is simpler than the spacecraft, consisting only of a noise temperature correction for obtaining a common reference: the best possible performance at DSS 13. For Ka-band, this requires removing the dichroic plate and optimizing the beam waveguide. For X-band, this requires switching to a ULNA. These correction terms are shown in Table 10.

Figure 22 shows the results of applying the correction factors to the tracking data from DOYs 18, 56, and 197. In spite of the wide scatter in Ka-band power estimates (due to the very low SNR) the potential improvement of Ka-band can be seen to range from 4 to 7 dB. Also, the relative advantage on a day like DOY 18, which was cloudy and wet, is seen to be about 1 dB less (disregarding ground station pointing errors) than a day like DOY 197, which was clear and calm.

[9] J. F. Weese, "Third Week Report for the Mars Observer Flight Panel Testing Accomplished in the Telecommunications Laboratory (TDL) from 4/29/91 through 5/3/91," Interoffice Memorandum 3395-91-054 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 10, 1991.

From here, the analysis can be taken still further, for instance by integrating the performance difference over an entire track. This would provide a single measurement of the Ka-band performance improvement for the track, including the effects of pointing errors and weather degradations while smoothing the effects of measurement scatter. Other goals for KaBLE data analysis include the development of a model of Ka-band link degradation as a function of weather and a study of the DSS-13 antenna performance. These goals will be carried out by the successor project to KaBLE, the Ka-Band Antenna Performance (KaAP) Experiment.

VII. Future Work

KaBLE ended due to the loss of communications with the Mars Observer spacecraft on August 21, 1993. However, the primary goal of KaBLE, characterizing Ka-band link performance, was far from complete. Compilation of accurate statistical weather models and ground-antenna performance parameters require much more data than that obtained during the 8 months of the experiment. To continue the work begun by KaBLE, a new experiment, the Ka-band antenna performance study was begun. The objective of the KaAP study is to characterize the antenna performance of DSS 13—its efficiency, G/T, pointing accuracy, and weather sensitivity—using quasars and other natural radio sources at X- and Ka-bands. In addition, studies and proposals are under way to equip future deep-space and Earth-orbiting spacecraft with Ka-band transmitters to conduct follow-on experiments with coherent signals.

Other plans for Ka-band include the Cassini Mission to Saturn, which will conduct radio science experiments at X- and Ka-bands, and the Gravitational Wave Experiment planned for the late 1990's, which will utilize dual X- and Ka-band signals coherent with a ground-to-spacecraft uplink. At that time, Ka-band radio links in deep space would become an operational aspect of the DSN.

The data analysis presented in this article is preliminary. Work will continue to further process and analyze KaBLE's data set to improve on the statistical aspect of the results.

Acknowledgments

Many people from JPL and Astro contributed to the success of this experiment—more than can fit on this one brief list. On the spacecraft side, thanks to Lance Riley and Dan Rascoe of JPL for initiating the Ka-band beacon with Astro; Bob Ostovarpoor of Astro for doubling the power of the beacon; Jim Weese of JPL for flight and ground compatibility testing at JPL, Astro, and Cape Canaveral; Jack Meeker, John Koukos, and Victor Lo for analysis and planning; and the Mars Observer Flight Project—especially Glenn Cunningham, Charles Whetsel, and Suzanne Dodd—for accommodating the needs of the experiment.

On the ground side, we thank Mark Gatti for overseeing the DSS-13 implementation; George Rinker for predicts generation; Farzin Manshadi for the dichroic plate; Leon Alvarez, Caroline Racho, and Martha Berg for antenna pointing; Jim Shell, Gary Glass, and Mark Fiore for the Ka-band maser; Bob Clauss for clinching the polarization problem; Eric Paulsen and Loris Robinett for receiver and Doppler tuner development; Ramin Sadr, Roland Bevan, Remi La Belle, and Robert Lee for the advanced receivers and data handling terminal; Sumita Nandi for tone tracker support; and Randy Heuser and Richard Chen for the monitor and control subsystem. The ranging demonstration could not have been performed without the spare Pioneer-10 receiver provided by Jack Childs and Jeff Berner, the new fiber-optic transceivers provided by George Lutes, and the ranging interface modifications made by Hal Baugh and Miguel Marina.

The authors particularly thank Lyle Skjerve and the staff of DSS 13, especially Gary Bury, Jerry Crook, Ron Littlefair, and Juan Garnica, for excellent service and

for finding various ways to improve the design and operation of the KaBLE ground system.

Finally, we thank Sami Asmar and Trish Priest for help in the preparation of this article, and David Morabito for his meticulous review.

References

[1] J. G. Smith, "Ka-Band (32-GHz) Downlink Capability for Deep Space Communications," *The Telecommunications and Data Acquisition Progress Report 42-88, vol. October–December 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 96–103, February 15, 1987.

[2] J. W. Layland and J. G. Smith, "A Growth Path for Deep Space Communications," *The Telecommunications and Data Acquisition Progress Report 42-88, vol. October–December 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 120–125, February 15, 1987.

[3] A. L. Riley, D. M. Hanson, A. Mileant, and R. W. Hartop, "A Ka-band Beacon Link Experiment (KABLE) With Mars Observer," *The Telecommunications and Data Acquisition Progress Report 42-88, vol. October–December 1986*, Jet Propulsion Laboratory, Pasadena, California, pp. 141–147, February 15, 1987.

[4] J. Shell and R. B. Quinn, "A Dual-Cavity Ruby Maser for the Ka-Band Link Experiment," *The Telecommunications and Data Acquisition Progress Report 42-116, vol. October–December 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 53–70, February 15, 1994.

[5] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes*, New York: Cambridge University Press, pp. 147–151, 1989.

Table 1. Mars Observer Spacecraft EIRP derivation for X- and Ka-band.

Parameter	X-band	Ka-band
Antenna diameter, m	1.5	0.28
1-dB beamwidth	0.9	1.3
Beam alignment, dB	0 (Ref.)	<0.1
EIRP budget		
Power, dBm	46.5	15.3
Circuit loss, dB	−3.4	−0.8
Antenna gain, dBi	39.8	37.0
Pointing loss, dB	−0.9	−0.5
Total *EIRP*, dBmi	82.0	51.0

Table 2. Ka-band G/T budget for DSS 13 at the rigging angle (50 deg el), assuming 90-percent weather conditions.

Parameter	Gain, dB	Physical temperature, K	Effective temperature, K
Maser and follow-on	0.0 (Ref.)	5.2	5.2
Cooled horn and waveguide	−0.27	1.8	1.8
Dichroic plate	−0.11	7.5	7.0
Antenna at F3 (at 50 deg el)	78.28	15.6	14.3
Atmosphere (90 percent at 50 deg el)	−0.28	17.9	16.4
Cosmic background	0.00	2.0	1.7
Subtotal	77.62	—	46.4
	−16.66		(16.66 dBK)
G/T	60.95		

Table 3. KaBLE link budget at end of telemetry window, January 18, 1993.

Parameter	X-band	Ka-band
Spacecraft $EIRP$, dBmi	82.0	51.0 ± 1
Space loss (0.37 AU), dB	−267.8	−277.9
DSS-13 G/T (above 30 deg el), dB	54.1	60.5 ± 0.7
DSS-13 pointing and polarization loss, dB	−0.1	−0.6 ± 0.4
Boltzmann's constant, dBm-K-Hz	198.6	198.6
P_t/N_0, dB-Hz	66.8	31.6
Telemetry modulation loss (59.9 deg), dB	−1.3	−1.5 ± 0.2
Processing loss, dB	−0.3	−1.0 ± 0.5
Data rate (250 bits/sec), dB-Hz	−24.0	−24.0
Decoder threshold, dB	−3.0	−3.0
Link margin, dB	38.2	2.1 ± 1.4

Table 4. DSS-13 technical requirements for KaBLE.

Item	Requirement
Antenna pointing error, mdeg rms	< 2.0
Microwave band, GHz	33.66 to 33.7
System noise temperature, K	33 ± 2 (90-percent weather)
Telemetry data rate, bits/sec	250
Ranging capability	Functional
Data acquisition	Various tracking statistics

C-4.

Table 5. Standard deviation of ranging measurements at X- and Ka-band compared with expectation from average P_r/N_0.

Track		Average P_r/N_0, dB-Hz	Expected sigma, RU	Measured sigma, RU
DOY 56	X-band	24.3	1.9	4.3
	Ka-band	3.4	21.2	25.1
DOY 76	X-band	20.9	2.8	4.6
	Ka-band	−1.6	37.7	34.4

Table 6. Summary of KaBLE tracking results.

DOY, 1993	Track, hr	Ka-data, hr	Weather	Remarks
8	12	8	Clouds and rain	Polarization mismatch
9	12	1	Clouds and rain	Polarization mismatch
10	12	10	Clouds, wind, and rain	Polarization mismatch
11	12	10	Clouds and rain	Polarization mismatch
12	12	1	Clouds, rain, and snow	Polarization testing
13	12	2	Clouds and rain	Decision to reconfigure maser
14	12	3	Heavy clouds and rain	Ka-band HEMT
15	12	6	Heavy clouds and rain	Maser back on line
16	12	11	Scattered clouds, rain	Telemetry demonstration success
17	12	10	Clouds and rain	Cannot clear TLM threshold
18	12	10	Clouds and rain	Carrier tracking only
28	8	4	Clear	Receiver problem
35	8	0	Clear	First LO free running
42	8	4	Clear	Ka-HEMT glitch (warm)
49	8	1	Clouds and rain	Power out–receiver crash
56	8	6	Clear	Ranging demonstration success
76	8	3	Clear and windy	Range data, interface problem
89	8	6	Clear and windy	DHT problem
95	8	3	High winds	Antenna wind limit (stow)
96	9	8	Clear	Interface problems
172	8	4	Clear	DHT problem
193	5	2	Clear	AMC problem
194	9	4	Clear	AMC/maser problems
195	10	8	Clear	AMC/maser problems
196	10	6	Clear	Maser G/T instability
197	10	10	Clear	AMC ok, maser stable
207	10	6	Clear	DHT problem
211	10	9	Clear	Ka-HEMT (low SNR)
217	7	4	Clear	LO tuning

Table 7. Derivation of KaBLE EIRP from DOY 197 tracking data shown with predicted EIRP from Table 1.[a]

Parameter	X-band	Ka-band
Measured P_c/N_0 (at 50 deg el), dB-Hz	**43 ± 0.2**	**11 ± 1**
Modulation losses, dB	3.3 ±0.2	4.1 ± 0.2
Boltzmann's constant, dBm-K-Hz	−198.6	−198.6
Polarization and circuit loss, dB	0	0.2
Measured SNT, dBK	**15.4 ± 0.1**	**16.4 ± 0.2**
DSS-13 pointing loss, dB	0.1	0.5 ± 0.2
DSS-13 gain (at 50 deg el), dB	−68.11 ± 0.1	−77.9 ± 0.5
Atmospheric attenuation, dB	0	0.3
Space loss, dB	280.5	292.5
Spacecraft pointing loss, dB	6.7 ± 0.2	3.5 ± 0.2
Estimated *EIRP*, dBmi	82.3 ±0.4	52.0 ± 1.2
Predicted *EIRP*, dBmi	82.0	51.0

[a] Tracking data are in boldface.

Table 8. Normalized Ka-band EIRP budget for performing X- and Ka-band link comparisons, assuming the same power amplifier efficiency, circuit loss, antenna diameter (and efficiency), and pointing loss as that for X-band.

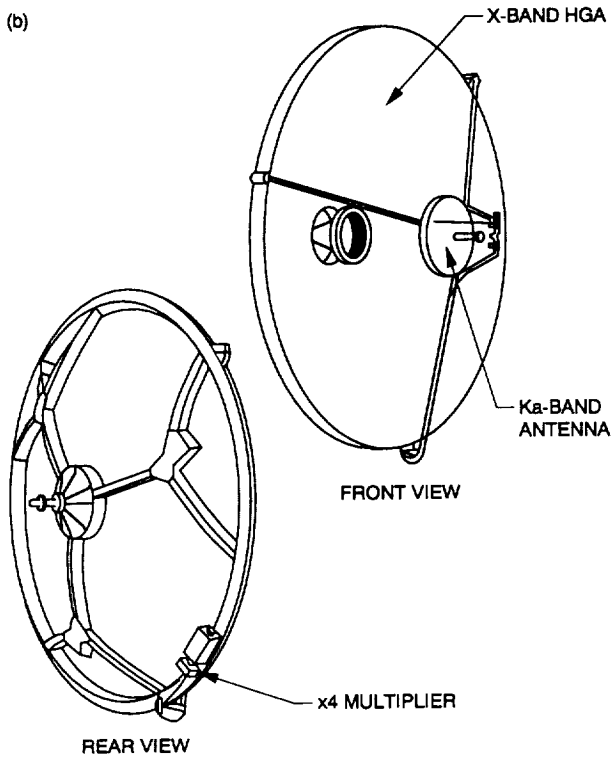
Parameter	X-band	Normalized Ka-band	KaBLE
Dish diameter, m	1.5	1.5	0.28
Power, dBm	46.5	46.5	15.3
Circuit loss, dB	−3.4	−3.4	−0.8
Antenna gain, dBi	39.8	51.8	37.0
Pointing loss, dB	−0.9	−0.9	−0.5
Totals			
EIRP, dBmi	82.0	94	52 (DOY 197
$\Delta EIRP$ (normalized KaBLE)	0.0	42	estimate)

Table 9. Correction factor for converting measured X- and Ka-band carrier SNR (P_C/N_0) into a comparative total SNR (P_t/N_0) based on EIRP normalization in Table 8, spacecraft pointing loss estimated in Section VI.D.1., and modulation losses from laboratory measurement.

Parameter	X-band	Ka-band
$\Delta EIRP$, dBmi	0.0	42
S/C pointing loss (estimate), dB	6.7	3.5
Modulation loss (ranging + TLM), dB	3.3	4.1
Correction factor, dB	10.0	49.6

Table 10. Receiver system noise temperature corrections to obtain common reference point, i.e., best possible SNT obtainable at DSS 13 for performing link comparisons.

Band	Best possible DSS-13 SNT	Actual		Correction	
		SNT_{ULNA}	SNT_{HEMT}	ΔT_{ULNA}	ΔT_{HEMT}
X	18	18	33	0	15
Ka	25	41	65	16	40



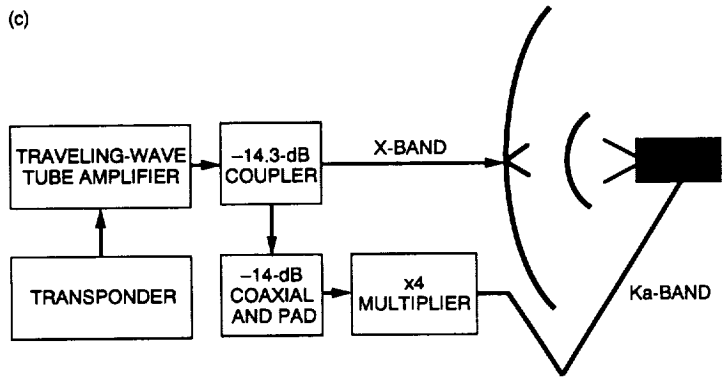


Fig. 1. KaBLE equipment on the Mars Observer spacecraft: (a) the spacecraft during preflight inspection; (b) HGA showing Ka-band antenna and times-four multiplier; and (c) transponder RF components.

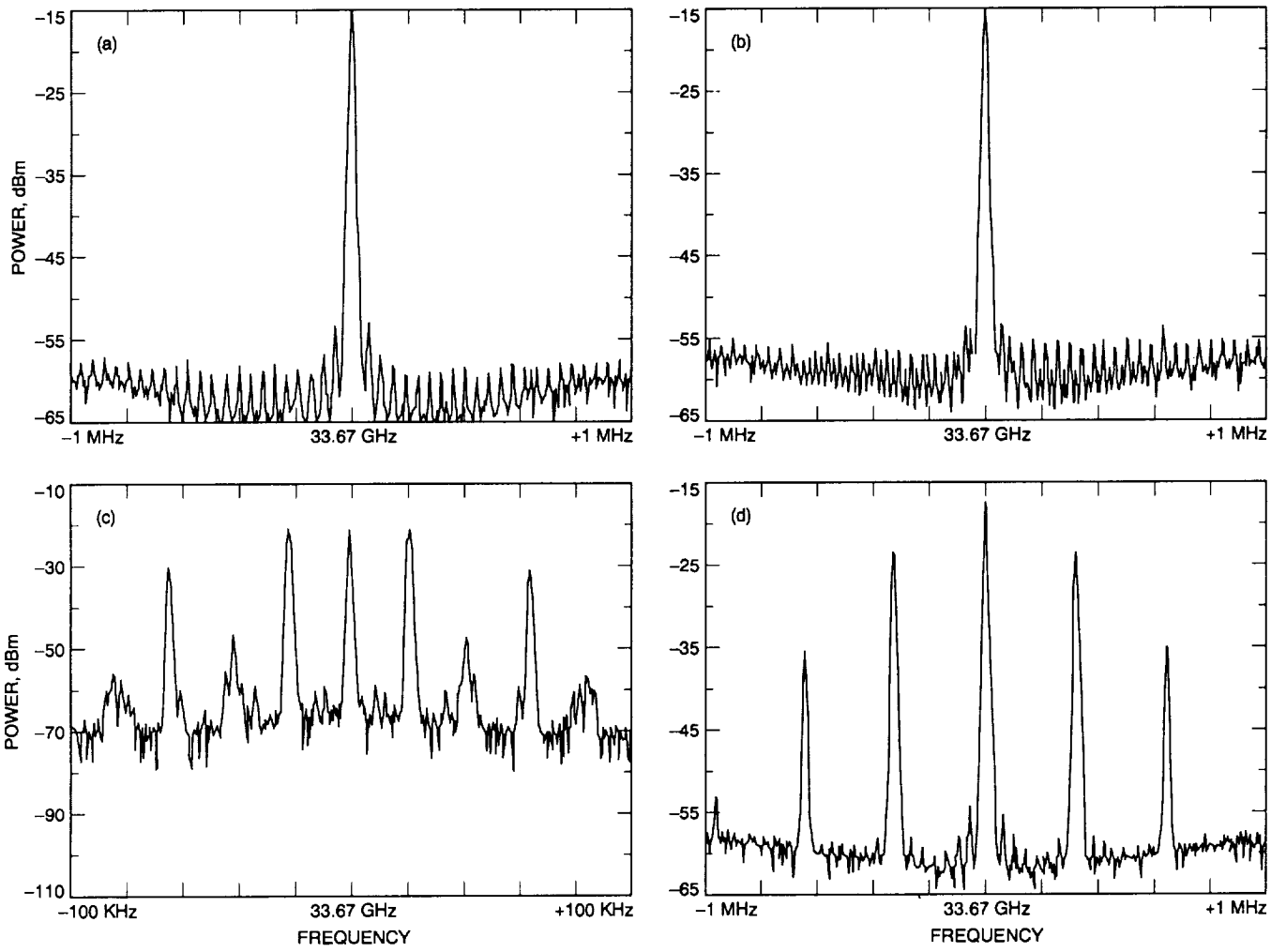


Fig. 2. Ka-band spectra from Mars Observer electromagnetic emissions test, showing the Ka-band carrier supression for X-band modulation indices of (a) 0 deg, (b) 44.8 deg (used for tracking experiment), (c) 59.9 deg (used for telemetry demonstration), and (d) 80 deg (used for mapping orbit tracking).

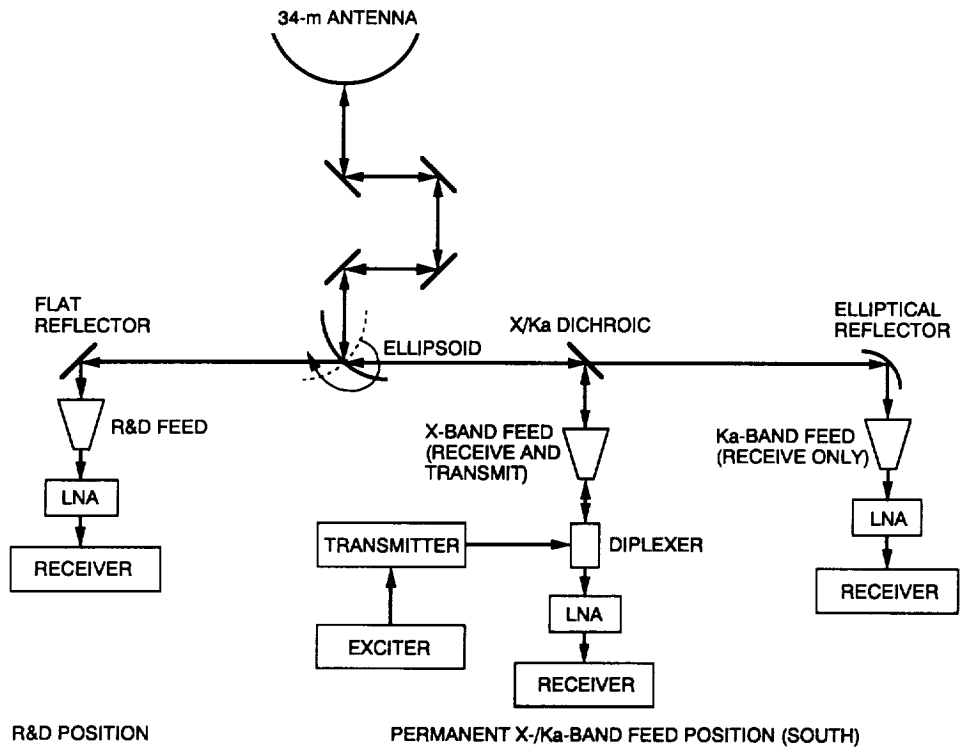


Fig. 3. DSS-13 microwave subsystem with the components included in the *G/T* budget of Table 2.

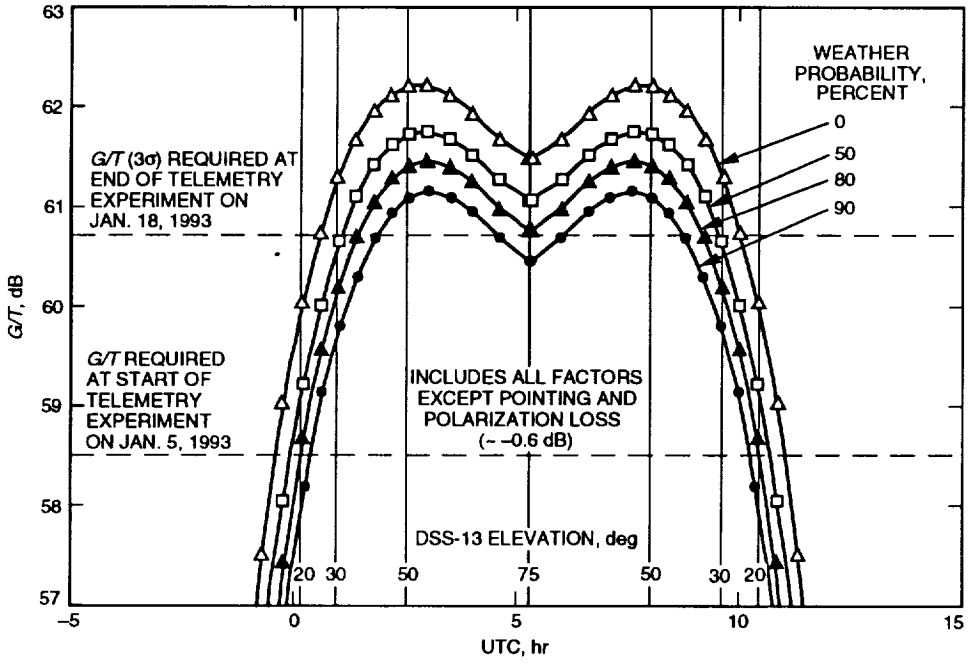


Fig. 4. DSS-13 *G/T* at 33.7 GHz versus time for a track in January 1993, showing the minimum requirements for telemetry reception and the effects of various weather conditions.

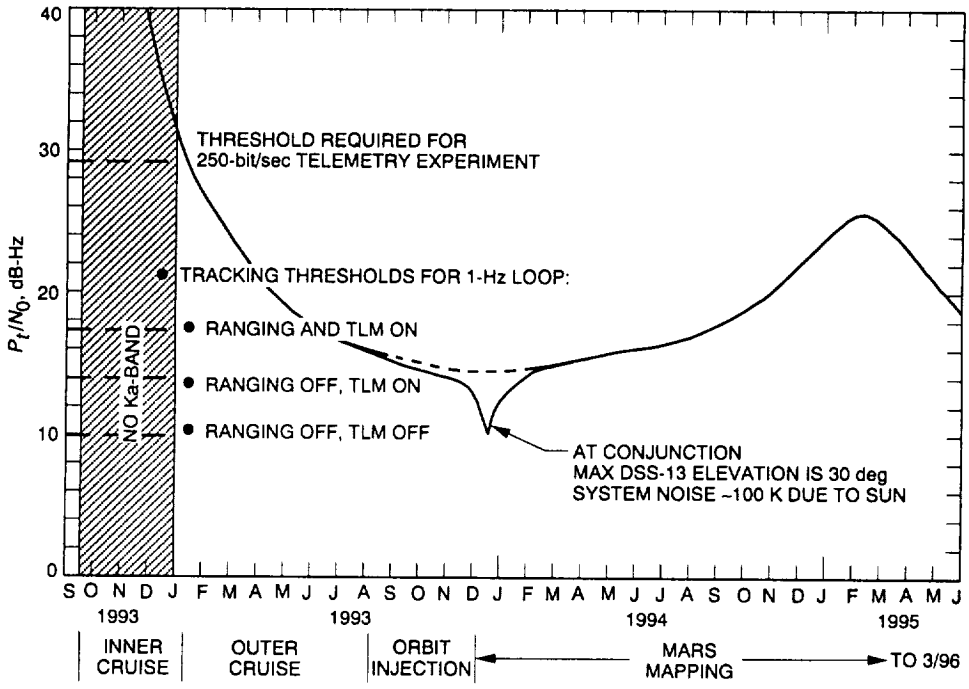


Fig. 5. Received Ka-band signal strength at DSS 13 expected over 2 years of the KaBLE experiment.

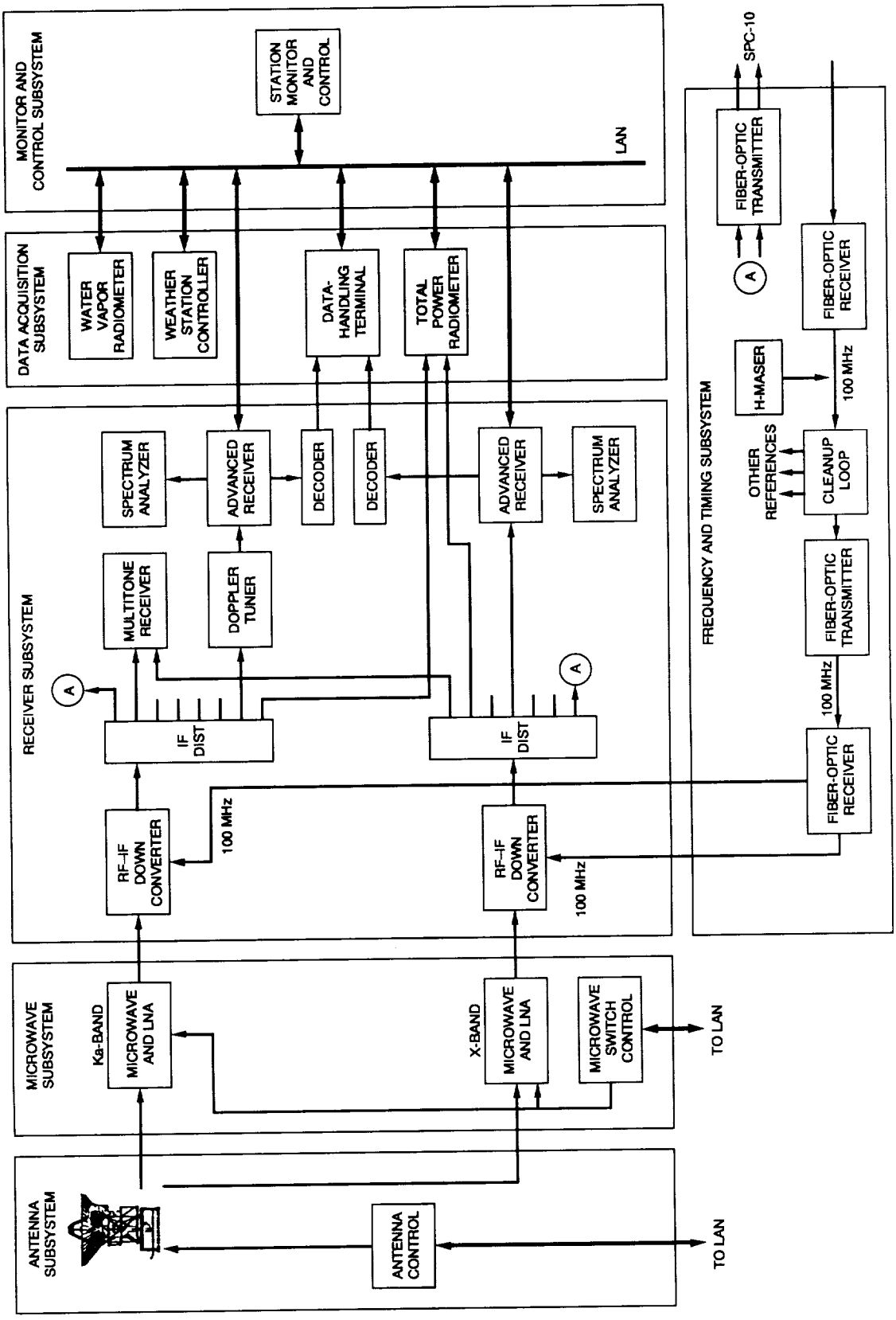


Fig. 6. KaBLE system installed at DSS 13.

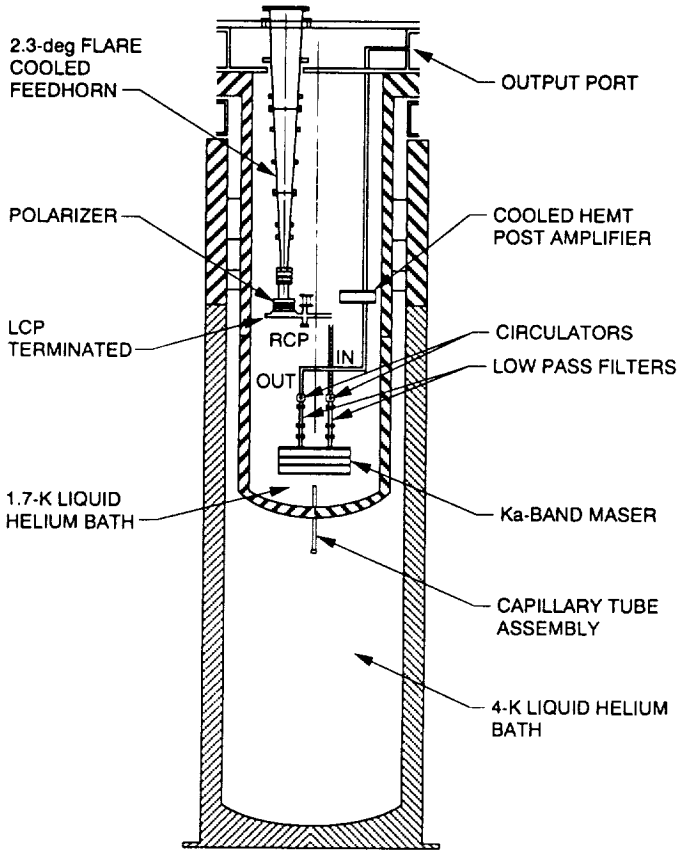


Fig. 7. Cross section of Ka-band maser cryostat, showing double-Dewar cryogenic system, cooled feedhorn, and maser.

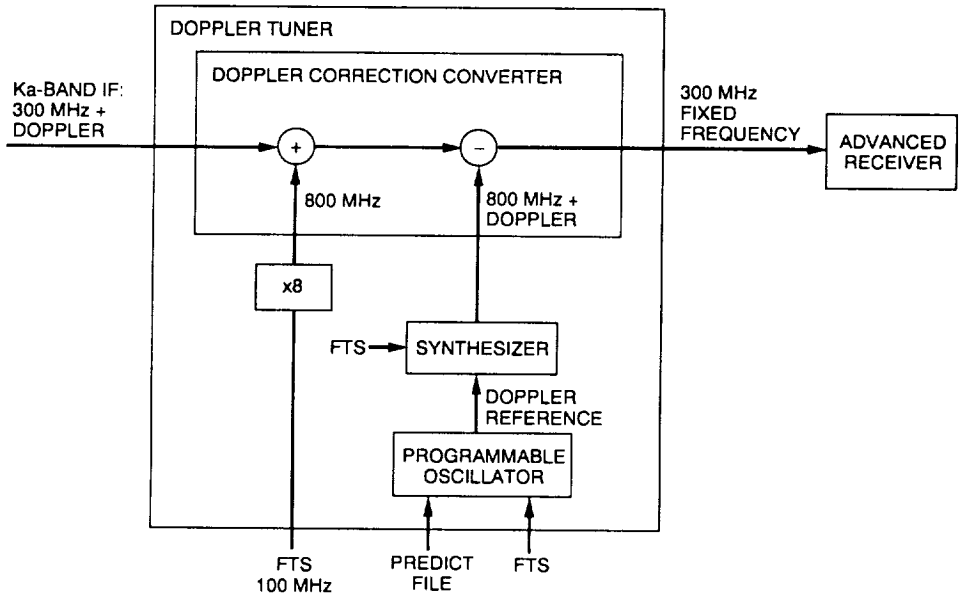


Fig. 8. Detail of KaBLE Doppler tuner.

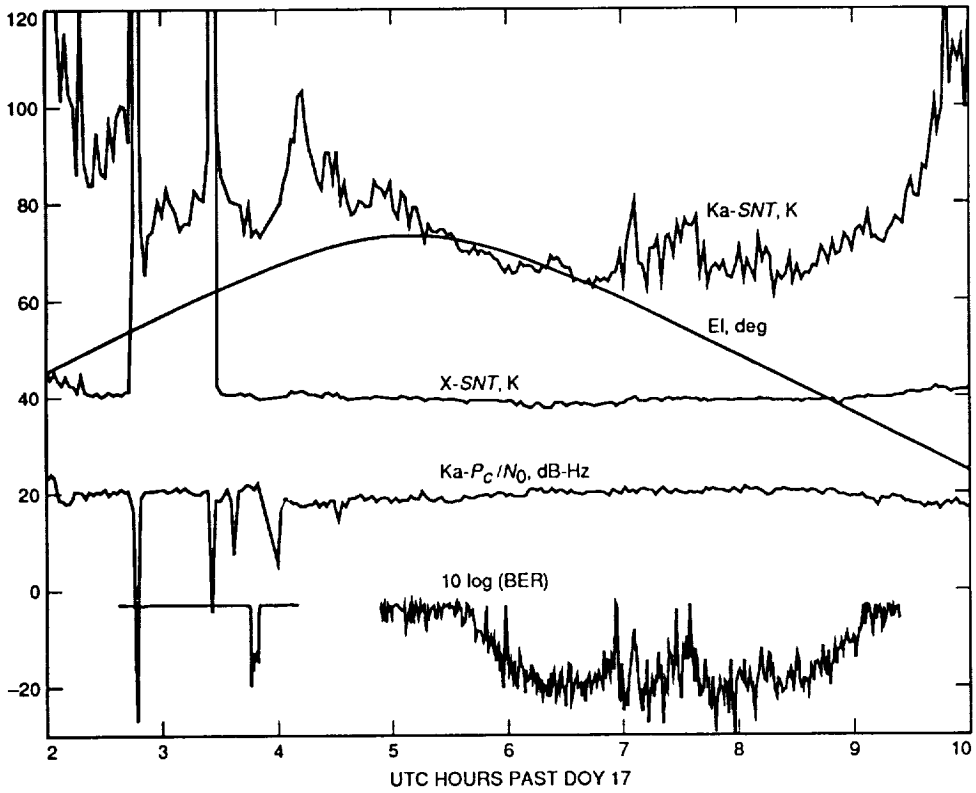


Fig. 9. Tracking statistics from DOY 17 of the telemetry demonstration showing carrier level, system noise temperature, elevation angle, and bit-error rate, indicating successful acquisition of telemetry data.

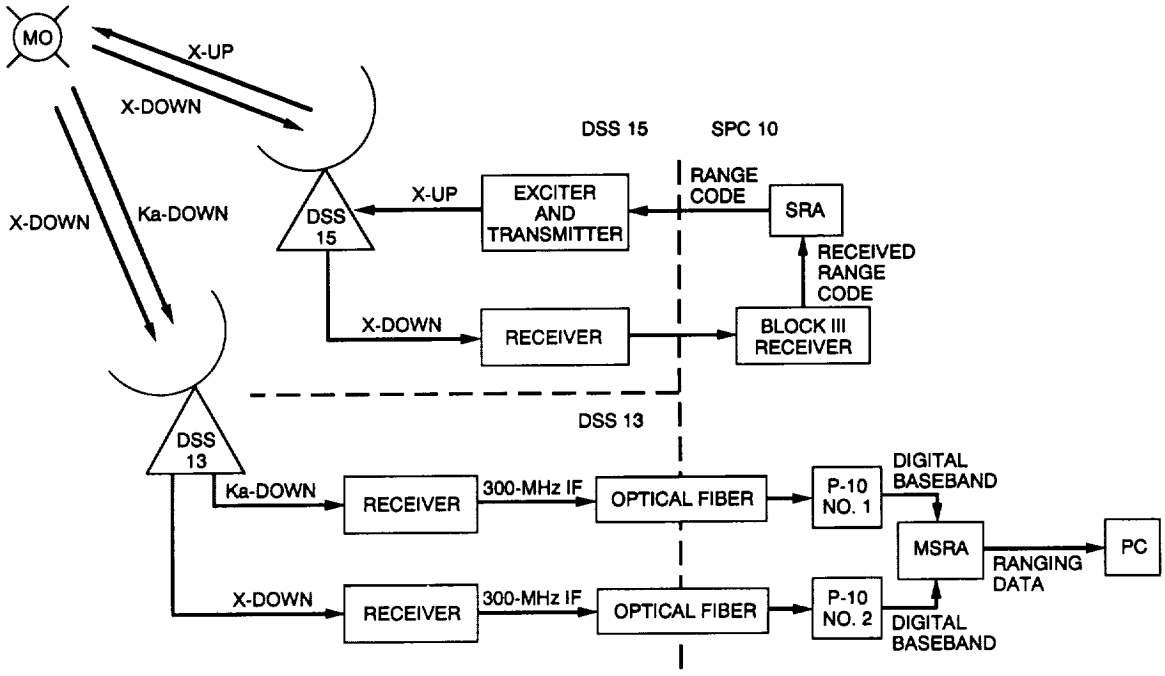


Fig. 10. Simplified diagram of test configuration used in the ranging demonstration.

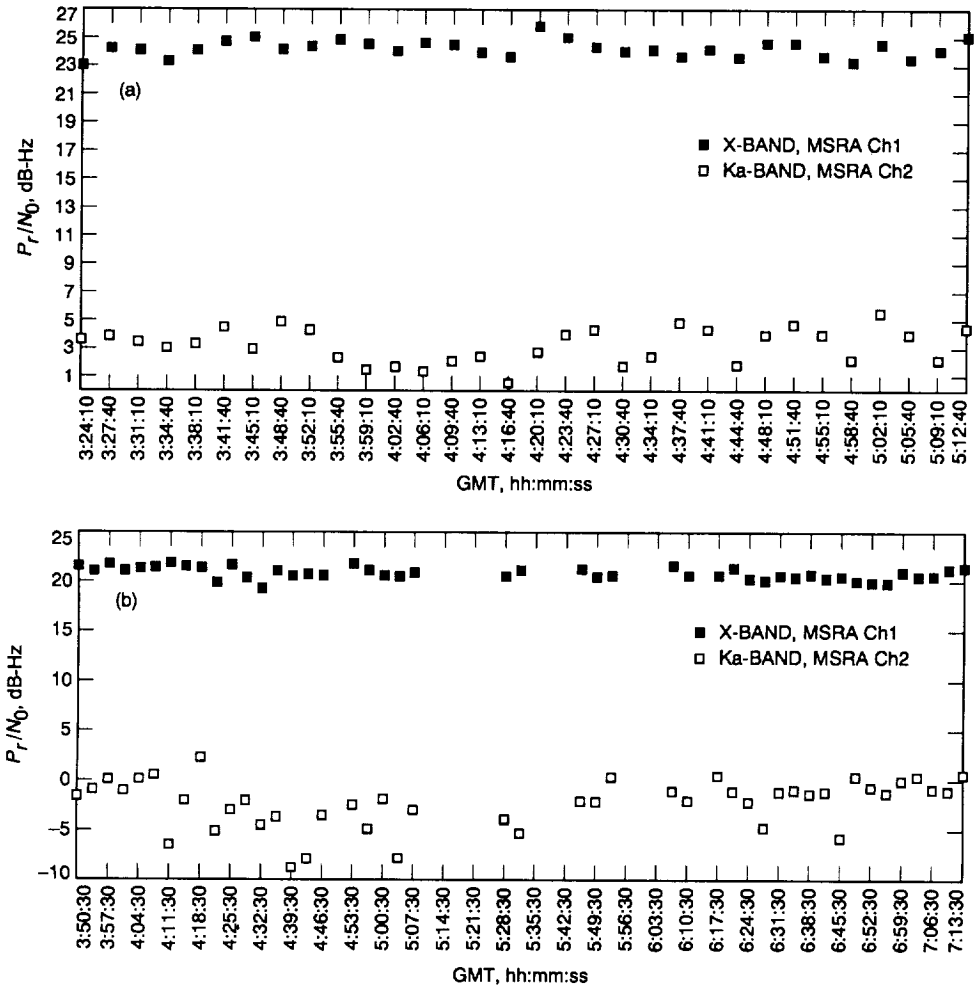


Fig. 11. Ranging SNR (P_r/N_0) received during two passes: (a) DOY 56 and (b) DOY 76.

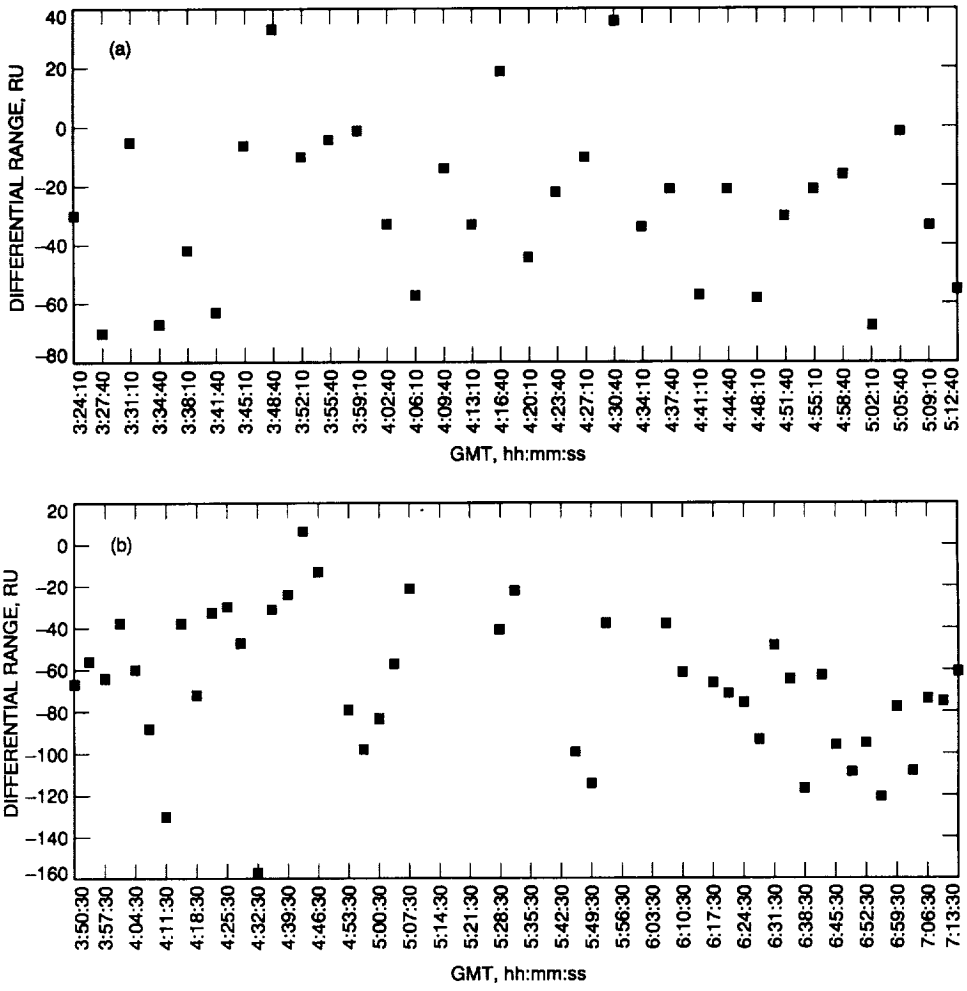


Fig. 12. X-/Ka-band differential range in RU of round-trip light time (1 RU ≈ 0.952 nsec): (a) DOY 56 and (b) DOY 76.

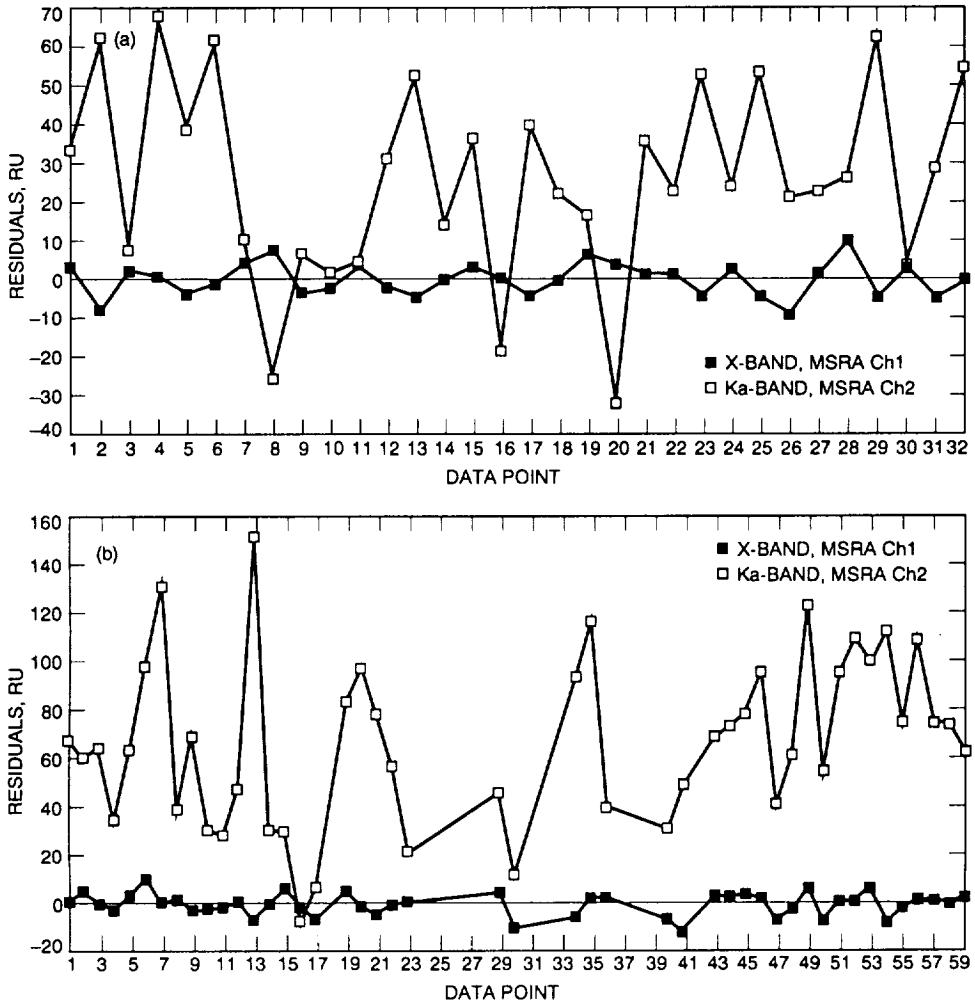


Fig. 13. X- and Ka-band residual range (after curve fitting to X-band data) showing increased scatter in the Ka-band data: (a) DOY 56 and (b) DOY 76.

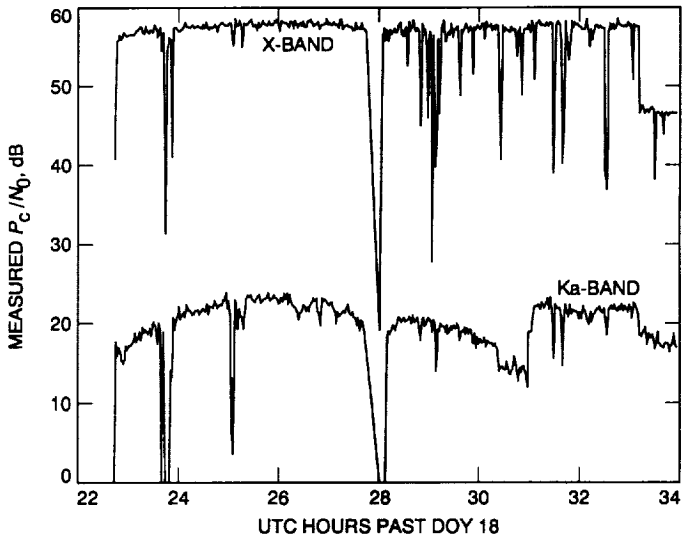


Fig. 14. Received carrier power during DOY 18 track showing dropouts in X- and Ka-band power due to uplink cycle slips and a large roll-off in Ka-band power beginning at 26:00 hr due to antenna mispointing and rain.

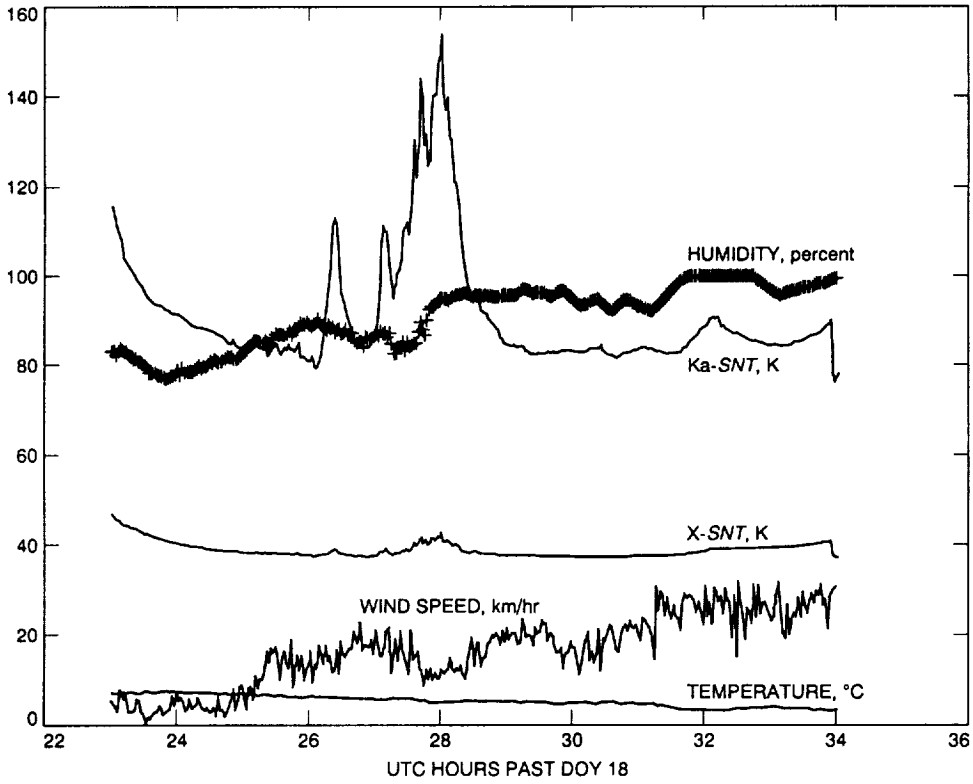


Fig. 15. System noise temperature and weather statistics for DOY 18 track showing rain signature in Ka-band noise temperature.

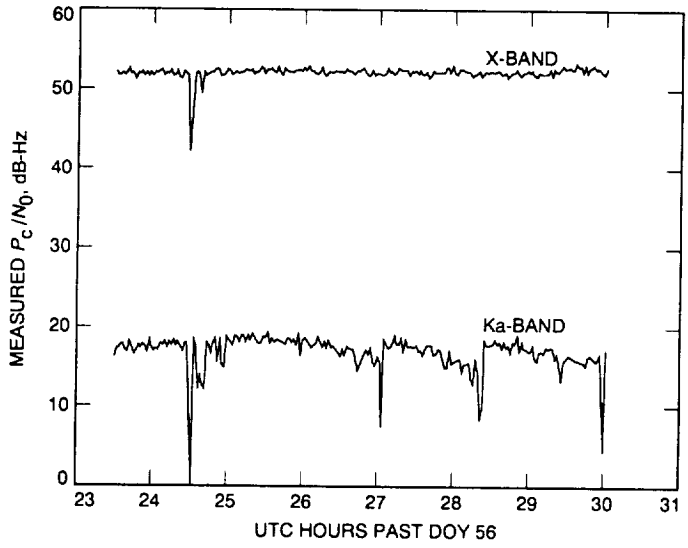


Fig. 16. Received carrier power during DOY 56 track showing accumulation of Ka-band pointing loss between boresights.

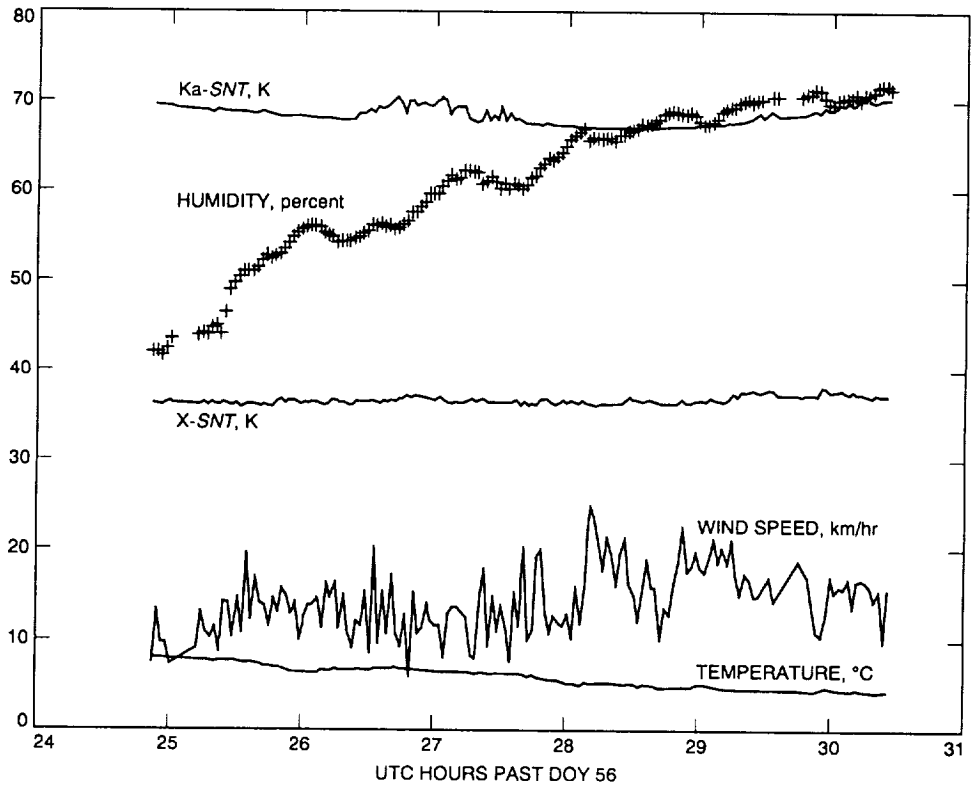


Fig. 17. System noise temperature and weather statistics for DOY 56 track, showing high Ka-band SNT due to HEMT.

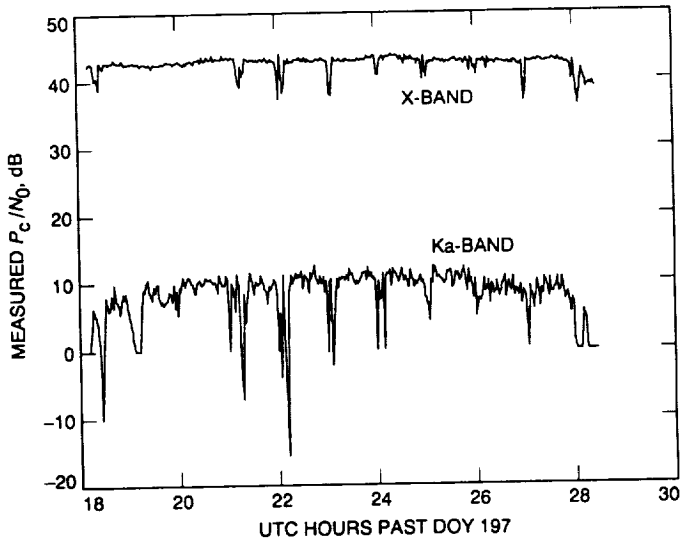


Fig. 18. Received carrier power for DOY 197 track with downward spikes showing hourly minicalibrations and boresights.

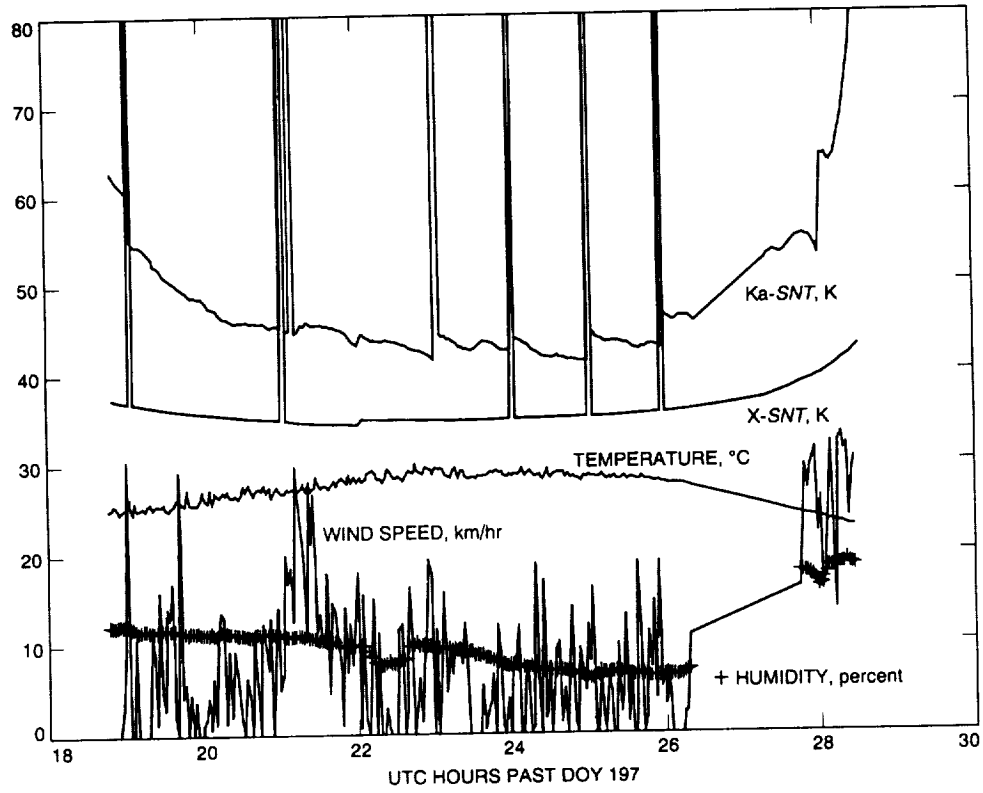


Fig. 19. System noise temperature and tracking statistics for DOY 197 track, showing fluctuations in Ka-band noise temperature due to gain instability.

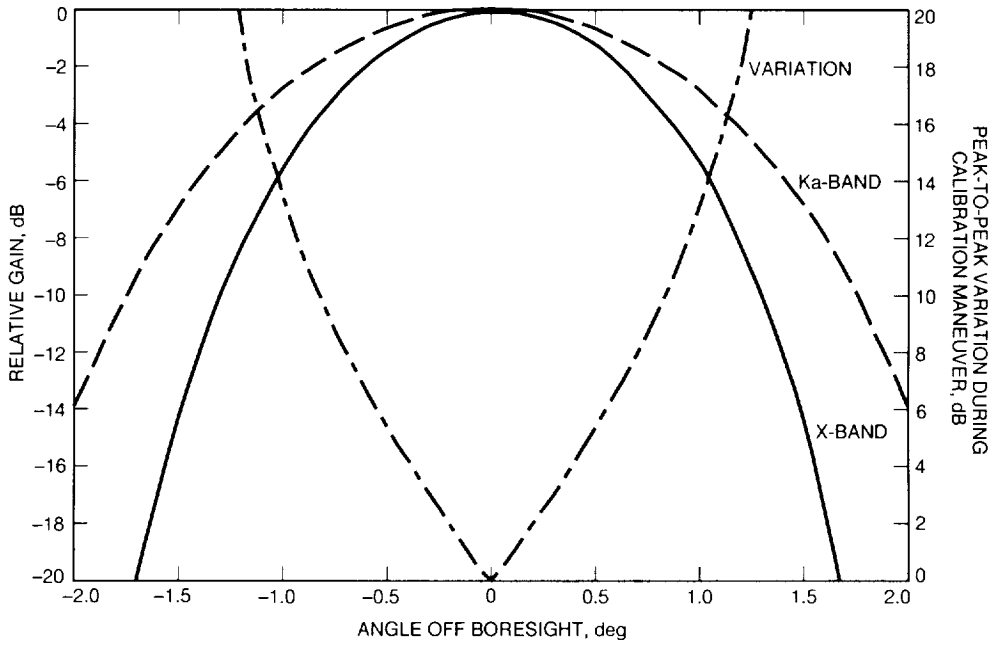


Fig. 20. X- and Ka-band HGA gain with expected variation in X-band power during calibration maneuver, as a function of pointing error.

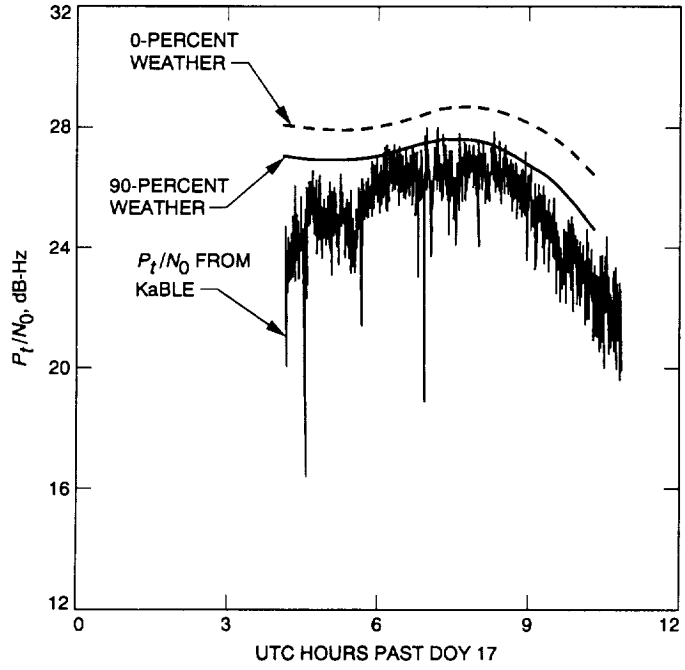


Fig. 21. Ka-band received SNR (P_t/N_0) compared with expectation based on modeled DSS-13 antenna performance.

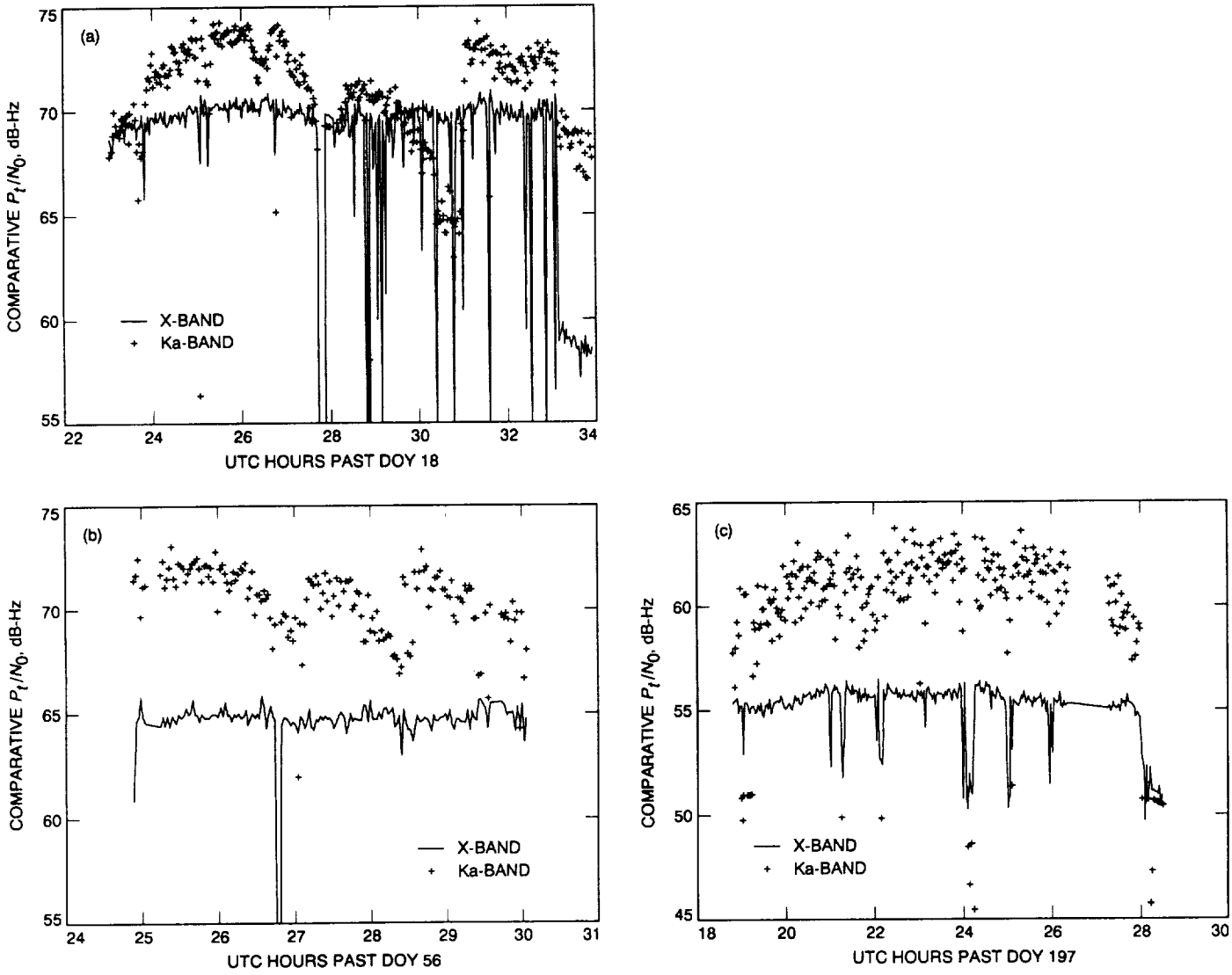


Fig. 22. Normalized X- and Ka-band SNR (P_t/N_0) for (a) DOY 18; (b) DOY 56; and (c) DOY 197, showing performance advantage of Ka-band over X-band at DSS 13, assuming identical spacecraft transmitter power and antenna size, as well as optimized X- and Ka-band receiver noise temperature (see VI.D.3). Scatter in Ka-band data is high due to weak (7 to 12 dB-Hz) signal strength.