NASA-CR-174149
19850005671

The Telecommunications and Data Acquisition Progress Report 42-79

July–September 1984

E.C. Posner
Editor

November 15, 1984

NASA
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

DISPLAY 85N13980/2

85N13980*# ISSUE 5 PAGE 622 CATEGORY 32 RPT#: NASA-CR-174149 JPL-TDA-PR-42-79 NAS 1.26:174149 84/11/15 193 PAGES UNCLASSIFIED DOCUMENT

UTTL: The Telecommunications and Data Acquisition Report TLSP: Progress Report, Jul. - Sep. 1984

AUTH: A/POSNER, E. C. PAT: A/ed.

CORP: Jet Propulsion Lab., California Inst. of Tech., Pasadena. AVAIL.NTIS SAP: HC A09/MF A01

MAJS: /*DATA ACQUISITION/*DEEP SPACE NETWORK/*SPACECRAFT COMMUNICATION/*VERY LONG BASE INTERFEROMETRY

MINS: / ARCHITECTURE (COMPUTERS)/ EXTRAGALACTIC RADIO SOURCES/ RANDOM NOISE/ SPACECRAFT TRAJECTORIES/ SUPERCONDUCTIVITY

ANN: Tracking and ground-based navigation; communications, spacecraft-ground; station control and system technology; capabilities for new projects; networks consolidation program; and network sustaining are described. For individual titles see N85-13981 through N85-13998.

ENTER:

The Telecommunications and Data Acquisition Progress Report 42-79

July–September 1984

E.C. Posner
Editor

November 15, 1984

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

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

Preface

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

In geodynamics, the publication reports on the application of radio interferometry at microwave frequencies for geodynamic measurements. In the search for extraterrestrial intelligence (SETI), it reports on implementation and operations for searching the microwave spectrum. The latter two programs are performed for NASA's Office of Space Science and Applications (OSSA).

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

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

OSTDS Tasks:

- DSN Advanced Systems
  - Tracking and Ground-Based Navigation
  - Communications, Spacecraft-Ground
  - Station Control and System Technology
  - Network Data Processing and Productivity
- DSN Systems Implementation
  - Capabilities for New Projects
  - Networks Consolidation Program
  - New Initiatives
  - Network Sustaining
- DSN Operations
  - Network Operations and Operations Support
  - Mission Interface and Support
  - TDA Program Management and Analysis
- GCF Implementation and Operations
- Data and Information Systems

OSSA Tasks:

- Search for Extraterrestrial Intelligence
- Geodynamics
  - Geodetic Instrument Development
  - Geodynamic Science

Discretionary Funded Tasks

Contents

OSTDS TASKS
DSN Advanced Systems
TRACKING AND GROUND-BASED NAVIGATION

COMMUNICATIONS, SPACECRAFT-GROUND

STATION CONTROL AND SYSTEM TECHNOLOGY

DSN Systems Implementation

CAPABILITIES FOR NEW PROJECTS

NETWORKS CONSOLIDATION PROGRAM

NETWORK SUSTAINING

Submilliarcsecond VLBI Using Compact Close Pairs of Radio Sources: Error Analysis

D. D. Morabito
Tracking Systems and Applications Section

The potential accuracy attainable for ΔVLBI positional measurements (submilliarcsecond level) can be reached by simultaneously observing pairs of compact radio sources whose angular separations are smaller than the beamwidth of each antenna. Simultaneous ΔVLBI (SVLBI) enables significant cancellation of measurement errors. Solar plasma is the dominant fluctuating error source in SVLBI positional measurements since there is enhanced cancellation of the troposphere and ionosphere, and complete cancellation of oscillator instabilities. Of the nonfluctuating error sources, errors due to universal time predominate.

By performing SVLBI experiments over several years with many different close pairs of radio sources, limits can be placed on reference-frame stability. Intrinsic properties of the sources, such as source structure and proper motion, will limit measurements. The SVLBI differential phase and corruptive noise sources will be discussed here along with estimated results.

I. Introduction

The advent of phase-stable atomic clocks and digital technology enabled the development of Very Long Baseline Interferometry (VLBI), which offers two orders of magnitude improvement in position determination over what was previously realized by ground-based optical systems. Conventional VLBI is used to measure crustal motion, earth rotation parameters, and relative source positions. The source positions have been measured to milliarcsecond accuracy and rely on the reference frame, defined by the extragalactic radio sources, being stable at the milliarcsecond level.

VLBI is being used by the Deep Space Network for spacecraft navigation (Refs. 1 through 4). By using a technique by which the spacecraft and an angularly nearby quasar are alternately observed, the differential position of the spacecraft can be determined to the milliarcsecond level relative to the quasar reference frame. Quasars from these reference frames have served as reference beacons in ΔVLBI experiments with the Voyager spacecraft (Ref. 2), the Pioneer Venus Orbiter (Ref. 4), and the Viking Orbiter (X X Newhall, private communication). Future applications of this technique will involve performing ΔVLBI between multiple spacecraft associated with a single planet. Such spacecraft include orbiters, landers, flybys, rovers, and balloons.

The potential accuracy attainable for ΔVLBI positional measurements can be reached by simultaneously observing source pairs with small angular separations. The angular sepa-

ration of such a pair must be smaller than the beamwidth of each antenna, so that it can be simultaneously tracked, greatly enhancing cancellation of measurement errors. By using the technique of simultaneous ΔVLBI (denoted SVLBI), there is significant cancellation of errors due to media, baseline, earth rotation parameters, and instrumentation, leaving attributes of the radio sources themselves to dominate (Ref. 5). Such intrinsic source properties include effects due to source structure and perhaps proper motion. By performing SVLBI on selected close pairs of radio sources, limits can be placed on the stability of the celestial reference frame to a precisely determined level. Figure 1 illustrates a comparison of differential position accuracies using conventional VLBI, time-multiplexed ΔVLBI, and SVLBI.

The relative proper motion of close pairs of radio sources can be determined to the several μarcsecond level using SVLBI. Repetitions of such measurements over a period of years will yield values or bounds on proper motions. Estimates of possible proper motion limits run about several μarcseconds/year, which is comparable to the estimated measurement accuracy of SVLBI. The detection of proper motion between two independent quasars would be of great scientific interest to the astronomy community and may have profound implications concerning standard interpretations of cosmological expansion theory. The absence of any detectable proper motion would support the apparent rigidity of the quasar reference frame, and would benefit the spacecraft navigation community.

Previous work using time-multiplexed ΔVLBI and SVLBI on close pairs of radio sources will be addressed here. Shapiro et al. (Ref. 6) estimated the relative position of the radio sources 3C 345 and NRAO 512 with an accuracy of 500 μarcseconds using time-multiplexed ΔVLBI. The relative proper motion of this source pair was bounded to 500 μarcseconds/year. Observations of the close pair 1038+528 have been conducted by Reid et al. (Ref. 7), Marcaide and Shapiro (Ref. 5) and Owen et al. (Ref. 8). Marcaide and Shapiro have measured the angular separation of 1038+528 using SVLBI to accuracies of 3 μarcseconds at X-band and 9 μarcseconds at S-band. The discrepancy found between their S-band and X-band measurements was significant (800 μarcseconds) and was attributed to the wavelength dependent opacity of one component. Gorenstein et al. (Ref. 9) monitored the "double" quasar 0957+56, which is thought to be multiple images caused by an intervening galaxy acting as a gravitational lens. The angular separation between the two components was measured to an accuracy of about 50 μarcseconds. Porcas et al. (Ref. 10) discussed the structure of this double quasar, which was found to be symmetrical and of similar core jet types. Morabito (Ref. 11) demonstrated 30-μarcsecond repeatability in the angular separation of GC 1342+662 and GC 1342+663 in two experiments conducted 4½ months apart using SVLBI with the DSN network.

The differential phase observable in SVLBI measurements will be defined and formulated in Section II. The nonfluctuating error sources will be discussed in Section III, and include errors due to UT1 - UTC, polar motion, baseline, and reference source position. The fluctuating or random noise sources that corrupt the differential phase observable will be discussed in Section IV. These noise sources include solar plasma, the ionosphere, the troposphere, system noise, and the interstellar medium. Source structure will be discussed separately in Section V. The expected accuracy of SVLBI measurements will be discussed in Section VI.

II. Differential Phase Observable

The VLBI instrument is used to simultaneously track two angularly close compact radio sources (see Fig. 2). The observed VLBI phase of source i as a function of time t is expressed by:

$$\phi_i(t) = (2\pi f/c)\,[B_e \cos(H_i(t)) \cos\delta_i + B_z \sin\delta_i] + \psi_i(t) + 2\pi n_i \qquad (1)$$

where

$H_i(t) = \alpha_g(t) + \lambda_b - \alpha_i$ is the hour angle between baseline and source i

f is the observing frequency in Hz

c is the speed of light in m/s

n_i is an integer

B_e is the equatorial component of the baseline in m

B_z is the spin-axis component of the baseline in m

α_i is the right ascension of source i

δ_i is the declination of source i

$\alpha_g(t)$ is the right ascension of Greenwich at time t

λ_b is the baseline longitude

ψ_i represents all dispersive and systematic noise contributions in the phase for source i

The data are recorded on magnetic tapes at each station and brought to a central processing site for cross-correlation. The VLBI software generates a model phase that is differenced from the observed phase (Eq. (1)), yielding a residual phase for source i. The residual phases of the two sources are then differ-

enced. By assuming that source 1 serves as a reference source, the resulting differential VLBI phase between sources 1 and 2 at sample time t_j is expressed by:

$$\Delta\phi(t_j) = (2\pi f/c)\,[B_e \sin H_2(t_j) \cos\delta_2\ \Delta\alpha - B_e \cos H_2(t_j)\,[\Delta\alpha^2/2 \cos\delta_2 + \sin\delta_2\ \Delta\delta] + B_z \cos\delta_2\ \Delta\delta] + F(\Delta\alpha, \Delta\delta) + \Delta\psi(t_j) + 2\pi(n_2 - n_1) \qquad (2)$$

where $H_2(t_j) = a_g(t_j) + \lambda_b - a_2$; a_2 and δ_2 are the right ascension and declination, respectively, of source 2; Δa and $\Delta\delta$ are the differential position offsets of right ascension and declination, respectively, for source 2 relative to source 1; $F(\cdot)$ denotes functional dependence on higher order terms of Δa and $\Delta\delta$. The differential noise term is expressed as:

$$\Delta\psi(t_j) = \psi_2(t_j) - \psi_1(t_j)$$

The differential phases (Eq. (2)) are input to a weighted-least-squares algorithm where $\Delta\alpha$ and $\Delta\delta$ are determined. The presence of data gaps in the pass may allow integer cycle ambiguities that must be corrected using standard phase connection techniques prior to least-squares fitting. The integer number of cycles, $n_2 - n_1$, can be determined from a preliminary fit and then subtracted from each differential phase measurement. A two-parameter fit can be performed yielding accurate estimates of $\Delta\alpha$ and $\Delta\delta$.

The noise on each measurement of differential phase, $\Delta\psi_j$, is assumed to be uncorrelated with the noise on any other measurement of differential phase $\Delta\psi_k$, $k \neq j$, with each measurement having variance σ_j^2. The noise fluctuates from measurement to measurement because of irregular fluctuations in the noncancelling part of the differential propagation delay and variations in the system noise (thermal conditions). The variance σ_j^2 includes random contributions due to the interstellar medium, solar plasma, the ionosphere, the troposphere, system noise, and perhaps source structure. Because the observations are simultaneous, any oscillator instabilities in the VLBI phase of one source will also manifest itself in the VLBI phase of the other source, and therefore cancel out completely. Most systematic noise sources are expected to significantly cancel, leaving those intrinsic to the radio sources to dominate along with the dispersive noise sources. Nonfluctuating and fluctuating noise sources will be discussed in the following sections.

III. Nonfluctuating Error Sources

This section discusses the propagation of errors in the right ascension and declination components of relative position due to errors in our knowledge of the rotation of the earth (UT1 - UTC), the position of the pole, the baseline, and the coordinates of the reference source. These error sources are considered nonfluctuating in that they do not contribute stochastically within an experiment but do contribute systematically. For this analysis, it is assumed that the observations are performed on the Goldstone-to-Madrid baseline and that the measurements are equally spaced over a diurnal period. The estimated errors presented may vary depending upon the hour angle range covered by the pass. However, they should be representative of errors incurred during actual experiments. A similar analysis is presented by Shapiro et al. (Ref. 6).

A. UT1-UTC and Polar Motion Errors

Using equation (2) and the above assumptions, errors in universal time, dt_u, and errors in polar motion parallel to the equatorial baseline component, dP_q, and perpendicular to the equatorial baseline component, dP_t, map into errors of estimated differential right ascension and differential declination as follows:

$$d\Delta\alpha \approx \Delta\alpha\, B_z/B_e\, dP_q - (\Delta\delta \tan\delta_2 + \Delta\alpha^2/2)\,(\omega_e\, dt_u - B_z/B_e\, dP_t)$$

$$d\Delta\delta \approx \Delta\delta\, B_z/B_e\, dP_q + \Delta\alpha \cot\delta_2\,(\omega_e\, dt_u - B_z/B_e\, dP_t)$$

For sources near the celestial equator, $d\Delta\alpha$ becomes negligible while $d\Delta\delta$ becomes significant. For sources near the pole, $d\Delta\alpha$ becomes significant while $d\Delta\delta$ becomes negligible. One millisecond is the typical BIH uncertainty of dt_u, but excursions of 4 milliseconds from quoted BIH values occasionally occur (Ref. 12). For BIH uncertainties of UT1 - UTC, positional errors can run up to 200 μarcseconds in right ascension at 80 degrees declination. Errors in positional measurements due to errors in UT1 - UTC can be reduced by solving for UT1 - UTC using VLBI. VLBI determined values of UT1 - UTC are accurate to 0.2 milliseconds.

The error in differential right ascension or declination due to BIH polar motion uncertainties of 0.015 arcseconds in each component is very small (≈2 μarcseconds). For the Goldstone-to-Madrid baseline, error contributions due to UT1 - UTC clearly dominate over those of polar motion and, (as shown

below), the other nonfluctuating error sources. Baselines with larger ratios of north-south to equatorial components will have larger relative source position errors due to errors in polar motion.

B. Baseline Errors

The errors in right ascension and declination due to the errors in the baseline are given by:

$$d\Delta\alpha = -\Delta\alpha\, dB_e/B_e$$

$$d\Delta\delta = -(\Delta\delta + \Delta\alpha^2/2 \cot\delta_2)\, dB_e/B_e$$

Typical DSN baseline errors of 0.1 to 1.0 m will limit errors in angular separation to <10 μarcseconds for passes involving intercontinental baselines.

C. Reference Source Position Errors

The errors in relative right ascension and declination due to errors in the reference source position ($d\alpha_2$, $d\delta_2$) can be expressed by:

$$d\Delta\alpha = \Delta\alpha \tan\delta_2\, d\delta_2 + (\Delta\delta \tan\delta_2 + \Delta\alpha^2/2)\, d\alpha_2$$

$$d\Delta\delta = (-\Delta\delta \cot\delta_2 + \Delta\alpha^2/2)\, d\delta_2 + (\Delta\alpha \cot\delta_2)\, d\alpha_2$$

For sources near the pole, $d\Delta\alpha$ becomes significant and $d\Delta\delta$ insignificant, while for sources near the celestial equator, $d\Delta\alpha$ becomes insignificant with $d\Delta\delta$ becoming significant. Errors in angular separation due to errors in reference source position (1 milliarcsecond) are limited to <10 μarcseconds for reasonable source declinations.

IV. Fluctuating Error Sources

A. Solar Plasma

The scatter in the differential phase (Eq. (2)) due to solar plasma (in radians) is shown by Callahan (Ref. 13) to be:

$$\sigma_{\Delta\psi} = 1.1\,(1/f)\,(L/v)^{0.75}\,F(\varepsilon) \qquad (3)$$

where f is the observing frequency in GHz, v is the propagation velocity of the solar plasma (400 km/s), L is the plasma scale size in km and $F(\varepsilon)$ is given by:

$$F(\varepsilon) = \begin{cases} [10 \sin\varepsilon]^{-1.3} & 0 < \varepsilon \leqslant 90^\circ \\ 10^{-1.3} & \varepsilon > 90^\circ \end{cases}$$

where ε is the elongation angle between sun and source pair.

Figure 3 illustrates the ray-path geometry between stations and sources at the solar plasma scale height. The separation of ray paths at the 1 AU (1.5×10^8 km) scale height between the two sources as seen from one station is $S = (1.5 \times 10^8)\,\Delta\theta$ km where $\Delta\theta$ is the angular separation in radians between the two sources. If B, the projected baseline length perpendicular to the source direction, is greater than S, then cancellation is between ray paths of the sources as seen from one station; hence, $L = S$ in Eq. (3). When $B < S$, then cancellation is across ray paths between stations as seen from each source; hence, $L = B$ in Eq. (3).

The effects of the solar plasma on the SVLBI phase for $B < S$ are the same as for time-multiplexed ΔVLBI or conventional VLBI. The solar plasma becomes a dominant error source for SVLBI because of the enhanced cancellation of other error sources that dominate in the other VLBI systems. The contribution of the solar plasma to the differential phase error is about 0.23 radians at S-band for the Goldstone-to-Madrid baseline under nighttime conditions with source pairs having $\Delta\theta \gtrsim 0.2$ arcminutes. For source pairs with $\Delta\theta \lesssim 0.2$ arcminutes, the solar plasma contribution to the phase error decreases as $\Delta\theta$ decreases. Solar plasma effects take place over time scales of 0.1 to 100 seconds depending upon solar elongation angle.

B. Ionosphere

For two sources separated by $\Delta\theta$ radians in the sky, we estimate the ionospheric effect on the differential phase. At the ionospheric scale height of 350 km, the separation between ray paths is 350 $\Delta\theta$ km. From Ref. 14, the relative fluctuations in electron density are given by:

$$\delta n/n = 5.9 \times 10^{-4}\,(350\Delta\theta)^{0.83} \qquad (4)$$

The ionosphere path delay I, in meters, is related to the observing frequency f in GHz, and the total electron content TEC, in 10^{17} electrons/m^2, by:

$$I = (4.03/f^2)\, TEC \qquad (5)$$

TEC is typically between 1 and 2 at night and as high as 10 during the day.

The fractional fluctuation in plasma density is assumed to be approximately equal to the fractional fluctuation in path delay:

$$\delta I/I \approx \delta n/n \qquad (6)$$

where δI is the delay fluctuation corresponding to the ray-path separation between the two sources at the ionospheric scale

height as seen by one station. By substituting Eqs. (4) and (5) into Eq. (6), we get:

$$\delta I = (0.307/f^2)\, TEC\, (\Delta\theta)^{0.83}$$

The differential phase scatter in radians due to ionosphere is expressed by:

$$\sigma_{\Delta\psi} = 6.4/f \sqrt{(TEC_1)^2 + (TEC_2)^2}\ \Delta\theta^{0.83}$$

where TEC_i is the total electron content present at station i.

The ray-path separations between sources are much smaller for SVLBI than the effective scale sizes present for time-multiplexed ΔVLBI or conventional VLBI. For these conventional VLBI systems, cancellation is over scale sizes on the order of 100 km. Typical ionospheric phase fluctuations can run about four radians over 600 s of observation time at S-band. Conventional VLBI is usually limited by the ionosphere at S-band. For SVLBI, the ionospheric contribution to the differential phase error is about 0.04 radians at S-band for a source separation of 7 arcminutes assuming nighttime conditions. Ionospheric effects are expected to take place over time scales of 1 second and above.

C. Troposphere

From power-law fits to the phase structure function at 5 GHz (Ref. 15), the tropospheric contribution to the phase uncertainty, in radians, can be approximated by:

$$\sigma_{\Delta\psi} \approx 0.44\, \Delta\theta^{0.7}$$

where $\Delta\theta$ is the source separation in radians. All parameter dependence is neglected except for source separation. The ray-path separation at the 6000-m scale height defines the appropriate scale sizes for the troposphere.

The troposphere contributes about ≈ 0.4 radians of phase noise in conventional VLBI systems. Because of the very small separation of the ray paths between the two sources, the troposphere effectively cancels out along with the ionosphere, leaving other noise sources to dominate in SVLBI experiments. The ray-path separation is ≈ 14 m for source pairs separated by 8 arcminutes using SVLBI, corresponding to a phase uncertainty of about 0.006 radians.

D. System Noise

The error on differential phase due to system noise is now discussed. Let $\Delta\tau$ be the differential delay between the correlator outputs of the two sources and W be the recording bandwidth. The signal of one component is uncorrelated with the signal of the other component when the source separation is sufficient such that $\Delta\tau > 1/W$. The signals are therefore separable in delay and reside on different bits. The differential phase uncertainty is the RSS of the individual source VLBI phase uncertainties and is given by (in radians):

$$\sigma_{\Delta\psi} = \sqrt{(SNR_1)^{-2} + (SNR_2)^{-2}}$$

where the signal-to-noise ratio for source j is estimated as follows:

$$SNR_j = 0.2\, S_j\, D_1\, D_2 \sqrt{(\varepsilon_1\, \varepsilon_2/T_1\, T_2)\, WT} \cdot \sqrt{\exp\left(-0.69\, [\Delta_{j1}^2/\theta_1^2 + \Delta_{j2}^2/\theta_2^2]\right)}$$

where S_j is the correlated flux density in Janskys of source j; D_i, ε_i, T_i, and θ_i are the diameter (m), efficiency, system noise temperature (K), and half-power beamwidth, respectively, of antenna i; W is the system bandwidth (MHz); T is the total observation time (sec); and Δ_{ji} is the angular separation between source j and the pointing center for antenna i. It is usually appropriate to use the centroid of the two sources to point each antenna in order to minimize complications ($\Delta_{1i} = \Delta_{2i}$).

The contribution to the differential phase noise due to the system noise ranges from negligible to significant depending on source strengths, system parameters, and observation time. System noise effects are characterized by rapid fluctuations as opposed to media-induced noise effects, which fluctuate more slowly. The system noise effects for reasonably small pointing offsets are the same for both conventional VLBI and SVLBI.

E. Interstellar Medium

The effects of the interstellar medium (ISM) are the least understood of all medium-related error sources affecting the differential phase observable. Scale sizes on the order of the baseline length are assumed. θ is assumed to be the scattered angular size of a source caused by diffractive interference of the ISM. A circular Gaussian source of half-power width θ in radians yields the fringe visibility given by:

$$\gamma(B, \theta, f) = \exp\left(-0.5\, (4.45\, f\, \theta\, B)^2\right) \qquad (7)$$

where B is the baseline projected against the source in meters, and f is the observing frequency in GHz. The structure function, D_ψ, of the fringe phase, ψ, can be determined from Eq. (7) using the relation:

$$\gamma(B, \theta, f) = \exp\left(-0.5\, D_\psi(B, \theta, f)\right) \qquad (8)$$

For galactic latitudes b greater than 5 degrees away from the galactic plane, the angle θ in radians has been estimated by Cordes (private communication) to be:

$$\theta \approx 0.41 \times 10^{-8} f^{-2.2} |\sin b|^{-0.6} \quad (9)$$

Assuming similar effects in the VLBI phase for both sources, the differential VLBI phase uncertainty due to the ISM is approximated using Eqs. (7) through (9) by:

$$\sigma_{\Delta\psi} \approx 1.90 \times 10^{-8} B f^{-1.2} |\sin b|^{-0.6} \qquad (b > 5 \text{ deg}) \quad (10)$$

For galactic latitudes below 5 degrees, the scatter size θ becomes more complicated and more strongly a function of galactic longitude.

Equation (10) with the Goldstone-Madrid baseline (B = 8.4 X 10^3 km) and b = 90 deg yields 0.058 radians of phase noise at S-band, and 0.012 radians of phase noise at X-band. ISM effects relate to fluctuation times of 100 to 1000 s and above, which are higher than those of the other error sources discussed. For low galactic latitudes, the ISM may dominate the differential phase noise. At higher galactic latitudes, the ISM is dominated by the solar plasma. Because the baseline defines the scale size, the effects of the ISM are the same for both conventional VLBI and SVLBI.

V. Structure Effects

Complex structure in one or both components of a close pair can contribute to erroneous angular separation measurements. Structure effects typically manifest themselves as position offsets due to the apparent motion of the brightness center of a source. Brightness distributions could change detectably at different epochs and lead to different relative position measurements resulting in the erroneous detection of proper motion. Marcaide and Shapiro (Ref. 5) have demonstrated that astrometry can be perfected down to the limits placed only by the brightness distribution of the sources. The milliarcsecond dimensions of the sources to which the VLBI observations are sensitive can be contrasted to the submilliarcsecond precision of SVLBI.

The magnitude of source structure effects on SVLBI measurements are expected to: (1) be zero for two perfect point sources, (2) be bounded by the fringe spacing of the interferometer for unresolved sources, and (3) possibly exceed the fringe spacing for resolved sources.

Prospective source pairs to be used for testing reference – frame stability should be known unresolved sources. SVLBI estimates of the differential source position over a many year period using several such source pairs will yield statistical bounds on reference-frame stability. Significant movement of the differential position measurements outside of the statistical uncertainties could be interpreted as detectable proper motion over that time period. Present estimates on the bounds of proper motion assuming current cosmological theories range upwards to several μarcseconds per year.

Time scales of structure effects can be estimated by comparing measurements between segments of an experiment and between different experiments over different time epochs. By performing several experiments over many years, these time scales can be statistically determined and cross-compared for different source pairs. Structure effects and proper motion can be sorted out more accurately using this data.

VI. Expected SVLBI Accuracy

To study limits of positional measurement accuracy, a given close pair should be monitored over a continuous several-hour pass and at regular intervals. A several-hour pass will insure high accuracy measurements while continuity will provide less difficulty in solving for cycle ambiguities. By repeating measurements over time, reference-frame stability can be tested and limits inferred. Cancellation of systematic error sources is expected to be sufficiently complete, so that any mismodeling may not perceptably bias the results. Structure effects can be minimized by producing radio brightness maps of both sources using a network of several antennas.

Six known angularly close pairs of radio sources appear in Table 1 along with the 1950.0 source positions, correlated flux densities, angular separations, and galactic latitudes. In all six cases, the angular separation between components is less than 7 arcminutes. These source pairs can be simultaneously observed at S-band using DSN baselines. The VLBI sensitivity of a pair of 64-m antennas using the MARK II recording system (4 Mb/s) is ~20 mJy for integration times of 100 s. Therefore, all source pairs listed in Table 1, except 0532-05, can be observed using S-band with this configuration.

Table 2 displays estimated S-band (2.29 GHz) fluctuating phase noise contributions for the source pairs listed in Table 1. 0532-05 is not included here since it is probably self-absorbed at S-band. The Goldstone-to-Madrid baseline is assumed using a 6-hour pass at nighttime. The flux densities used to estimate the system noise effects at S-band are assumed to be those given in Table 1. The sensitivity of an interferometer consisting of two 64-m antennas is sufficient to detect both components of the source pairs listed. By performing SVLBI, the ionosphere and troposphere significantly cancel leaving the solar plasma to dominate. Note that the other noise contributions listed in the table are smaller by comparison, except

those for 2016+112, which are dominated by the ISM. The ISM is expected to dominate for source pairs with smaller galactic latitudes. Figure 4 displays estimated differential phase and source separation errors versus angular source separation for S-band using the above configuration.

Table 3 displays the corresponding estimates of fluctuating phase noise for X-band observations (8.42 GHz). 3C 66 is not included here since the source separation is larger than the X-band beamwidths of the available DSN antennas. Since GC 1342+66 has an angular separation greater than a 64-m antenna beamwidth, a pair of 34-m antennas has been assumed in order to accommodate simultaneous observations. The flux densities used to estimate the system noise effects at X-band are assumed to be those given in Table 1. From Table 3, the solar plasma is shown to be the dominating error source at X band ($\approx$ 0.064 radians), except for the cases of 0532-05, where the system noise (0.082 radians) dominates, and 2016+112, where the ISM dominates. The system noise contribution for 0532-05 can be reduced to 0.022 radians using Mark III recording. Figure 5 displays estimated differential phase and source separation errors versus angular source separations for X-band using the above configuration.

An advantage of using X-band is that the plasma media contributions are significantly reduced compared to the S-band case. It is obviously more advantageous to observe at X-band if antenna beamwidths and source strengths permit. A disadvantage of using X-band is that the antenna beamwidth may be smaller than the angular separations of some source pairs. Thus simultaneous observations are not possible without resorting to smaller antennas (larger beamwidths) at the cost of lower sensitivity.

Estimated uncertainties for observing the source pair GC 1342+66 under the conditions of Tables 2 and 3, are 0.24 radians ($\approx$22 μarcseconds) at S-band, and 0.066 radians ($\approx$5 μarcseconds) at X-band. The estimated S-band values of random phase uncertainty (0.24 radians) are in agreement with measured values (0.17 radians) (Ref. 11). Figure 6 displays the S-band error budget for observing GC 1342+66. The random errors (ie., solar plasma) clearly dominate over the systematic errors. Figure 7 is the corresponding X-band error budget for observing GC 1342+66. Here the random errors are dominated by those due to UT1 - UTC. It is therefore advisable to solve for UT1 - UTC using VLBI so that angular separation errors can be further reduced.

VII. Conclusion

By performing SVLBI on an angularly close source pair, the source separation can be estimated to the submilliarsecond level (and approach the μarcsecond level) because of the enhanced cancellation of error sources. This is a great improvement over conventional VLBI methods. The differential phase observable and the error sources affecting angular separation measurements have been discussed. The nonfluctuating error sources include errors due UT1 - UTC, polar motion, baseline, and reference source position. The fluctuating error sources include the interstellar media, solar plasma, ionosphere, troposphere, and system noise. Source structure effects have been discussed along with the expected accuracies using this technique.

Solar plasma is the dominant error source for SVLBI measurements because of the enhanced cancellation of the troposphere and ionosphere. For weaker source pairs, the system noise may start to dominate. For sources with low galactic latitudes, the ISM may start to dominate.

By performing SVLBI experiments over several years with many different close pairs of radio sources, limits can be placed on reference-frame stability. Intrinsic properties of the sources, such as source structure and proper motion which limit measurements, can then be more fully understood. SVLBI will therefore provide a useful tool for testing long-term reference-frame stability.

Acknowledgements

I would like to thank D. Trask, K. M. Liewer, G. H. Purcell, A. E. Niell, J. W. Armstrong, R. A. Preston, S. C. Wu, D. L. Jauncey, P. S. Callahan, and T. P. Yunck for many interesting discussions.

References

1. Melbourne, W. G., Curkendall, D. W., "Radio Metric Direction Finding: A New Approach to Deep Space Navigation," AAS/AIAA Astrodynamics Specialist Conference Paper, presented at Jackson Hole, Wyoming, Sept. 7-9, 1977.

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

3. Brunn, D. L., Preston, R. A., Wu, S. C., Siegel, H. L., Brown, D. S., "ΔVLBI Spacecraft Tracking System Demonstration: Part I. Design and Planning," *DSN Progress Report 42-45*, p. 111. Jet Propulsion Laboratory, Pasadena, Calif., 1978, p. 111.

4. Esposito, P. B., Donivan, F. F., Finley, S. G., Newhall, X X, Smith, C. B., Wu, S. C., "Narrowband Differential Interferometry Applied to Pioneer Venus Orbiter," paper presented at AAS/AIAA Astrodynamics Specialist Conference held at Lake Placid, New York, Aug. 22-25, 1983.

5. Marcaide, J. M., and Shapiro, I. I. "High Precision Astrometry Via Very-Long Baseline Radio Interferometry: Estimate of the Angular Separation Between the Quasars 1038+528A and B," *Astronomical Journal*, Vol. 88, 1983, p. 1133.

6. Shapiro, I. I., Wittels, J. J., Counselman, C. C. III, Robertson, D. S., Whitney, A. R., Hinteregger, H. F., Knight, C. A., Rogers, A. E. E., Clark, T. A., Hutton, L. K., Niell, A. E., "Submilliarcsecond Astrometry Via VLBI. I. Relative Position of the Radio Sources 3C 345 and NRAO 512," *Astronomical Journal*, Vol. 84, 1979, p. 1459.

7. Reid, M. J., Owen, F. N., Shaffer, D. B., Kellermann, K. I., Witzel, A., "Quasar Proper Motion Study of 1038+528," *B.A.A.S.*, Vol. 12, 1980, p. 497.

8. Owen, F. N., Wills, B. J., Wills, D., "A Close Pair of Radio-emitting Objects," *Ap. J.*, Vol. 235, No. L59, 1980.

9. Gorenstein, M. V., Shapiro, I. I., Cohen, N. L., Falco, E., Kassim, N., Rogers, A. E. E., Whitney, A. R., Preston, R. A., Rius, A., "VLBI Observations of the "Twin Quasar" 0957+56A, B," *B.A.A.S.*, Vol. 12, 1980, p. 498.

10. Porcas, R. W., Booth, R. S., Browne, I. W. A., Walsh, D., Wilkinson, P. N., "VLBI Structures of the Images of the Double QSO 0957+561," *Nature*, Vol. 289, 1981 p. 758.

11. Morabito, D. D., "Submilliarcsecond VLBI Observations of the Close Pair GC 1342+662 and GC 1342+663", TDA Progress Report 42-77, January-March, 1984.

12. Eubanks, T. M., Steppe, J. A., Spieth, M. A., "The Accuracy of Radio Interferometric Measurements of the Earth Rotation," *Nature*, (Submitted).

13. Callahan, P. S., "An Analysis of Viking S-X Doppler Measurements of Solar Wind Columnar Content Fluctuations," *The DSN Progress Report 42-44*. Jet Propulsion Laboratory, Pasadena, Calif., April 1977.

14. Phelps, A. D. R., Sagalyn, R. C., "Plasma Density Irregularities in the High-Latitude Top-Side Ionosphere," *J.G.R.*, Vol. 81, No. 515, 1976.

15. Armstrong, J. W., Sramek, R. A., "Observations of Tropospheric Phase Scintillations at 5 GHz on Vertical Paths," *Radio Science*, Vol. 17, 1982, p. 1579.

16. Morabito, D. D., Preston, R. A., Williams, J. G., Faulkner, J., Jauncey, D. L., and Nicolson, G. D., "A Full-Sky VLBI Survey at 2.29 GHz," to be published in the *TDA Progress Report.*

17. Moran, J. M., Garay, G., Reid, M. J., Genzal, R., "A Compact Radio Source Associated with the Becklin-Neugebauer Object," *B.A.A.S.* Vol. 13, 1981, p. 852.

18. Lawrence, C. R., Schneider, D. P., Schmidt, M., Bennett, C. L., Hewitt, J. N., Burke, B. F., Turner, E. L., Gunn, J. E., "Discovery of a New Gravitional Lens System," *Science*, Vol. 223, No. 4631, 1984, p. 46-49.

Table 1. Six known angularly close source pairs

Source	Right ascension, h m s	Declination, deg m s	Correlated flux density,[a] Jy	Angular separation, arcmin	Galactic latitude, deg
3C 66A	02 19 29.99	42 48 29.6	0.27 +/-0.05	6.4	-17
3C 66B	02 20 01.73	42 45 54.6	0.09 0.01		
0532-05A	05 32 46.65	-05 24 16.5	0.008[b] 0.0007	0.2	-19
0532-05B	05 32 47.43	-05 24 18.9	0.004[b] 0.0007		
0957+561A	09 57 57.3	56 08 22.9	0.026[c] 0.005	0.1	48
0957+561B	09 57 57.4	56 08 16.9	0.020[c] 0.005		
1038+528A	10 38 43.16	52 49 09.64	0.26 0.03	0.6	55
1038+528B	10 38 45.27	52 49 37.6	0.15[d] 0.01		
GC 1342+662	13 42 17.90	66 17 27.8	0.34 0.04	4.4	50
GC 1342+663	13 42 41.03	66 21 12.9	0.65 0.07		
2016+112A	20 16 55.47	11 17 46.56	0.056[e] 0.003	0.057	-14
2016+112B	20 16 55.27	11 17 45.07	0.062[e] 0.003		

[a]All measurements are at 13-cm wavelengths from Ref. 16, except where otherwise noted.
[b]Reference 17, at 2 cm.
[c]Reference 10, at 18 cm.
[d]Reference 8, at 13 cm.
[e]Reference 18, at 20 cm.

Table 2. S-band estimated phase uncertainties (radians)[a]

Source pair	Solar plasma	Ionosphere	Troposphere	System noise	ISM	RSS
3C 66	0.23	0.04	0.005	0.003	0.12	0.26
0957+561	0.14	0.0014	0.0003	0.018	0.069	0.16
1038+528	0.23	0.006	0.001	0.002	0.065	0.24
GC 1342+66	0.23	0.032	0.004	0.001	0.068	0.24
2016+112	0.094	0.00085	0.0002	0.007	0.14	0.17

[a]These estimates assume the Goldstone-to-Madrid baseline, an observing time of 6 hours, and nighttime conditions. The system noise contribution assumes two 64-m antennas with 0.5 efficiencies, 25-K system temperatures, and Mark II recording. All estimates are based on the data found in Table 1; source structure effects neglected.

Table 3. X-band estimated phase uncertainties (radians)[a]

Source pair	Solar plasma	Ionosphere	Troposphere	System noise	ISM	RSS
0532-05	0.064	0.001	0.0005	0.082	0.024	0.11
0957+561	0.039	0.0004	0.0003	0.018	0.014	0.045
1038+528	0.064	0.002	0.001	0.002	0.014	0.066
GC 1342+66	0.064	0.008	0.004	0.0035	0.014	0.066
2016+112	0.026	0.00023	0.0002	0.007	0.029	0.040

[a]These estimates assume the Goldstone-to-Madrid baseline, an observing time of 6 hours, and nighttime conditions. The system noise contribution assumes two 64-m antennas with 0.5 efficiencies, 25-K system temperatures, and Mark II recording, except for GC 1342+66 where we assume two 34-m antennas in order to accommodate simultaneous observations. All estimates are based on the data found in Table 1; source structure effects neglected.

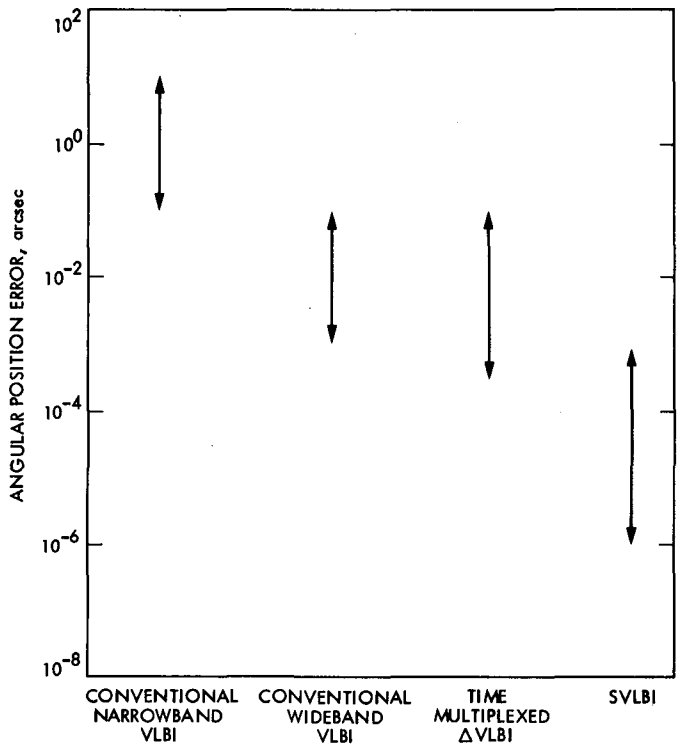


Fig. 1. VLBI system comparison

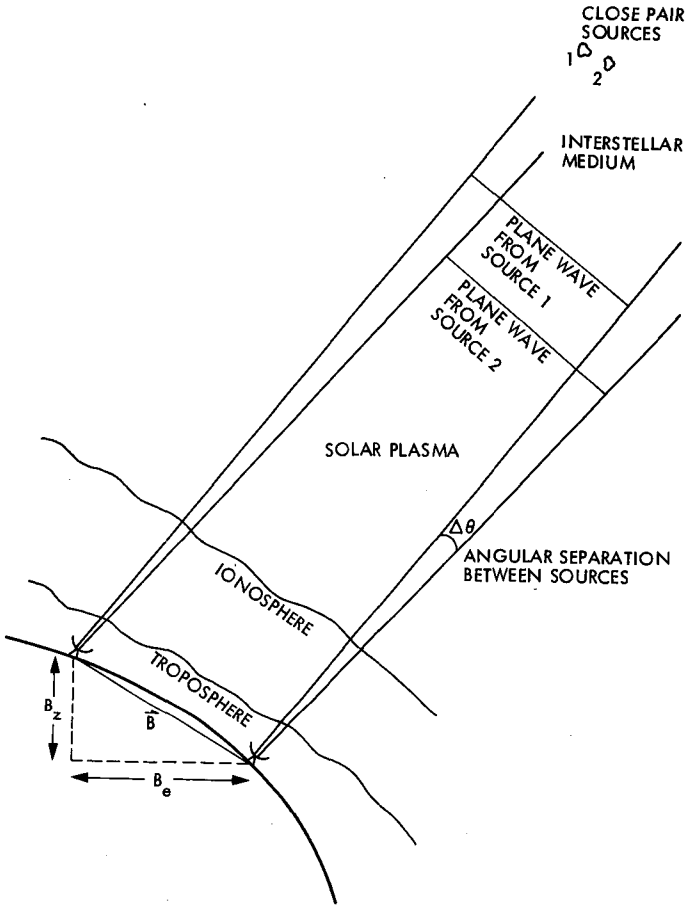


Fig. 2. Basic SVLBI geometry

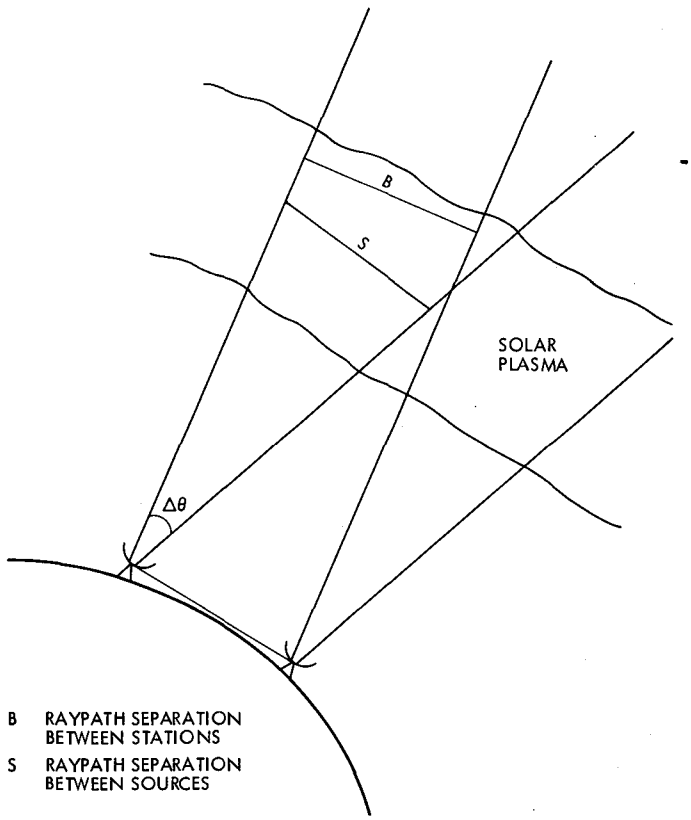


Fig. 3. Solar plasma raypath separations at 1 AU scale height

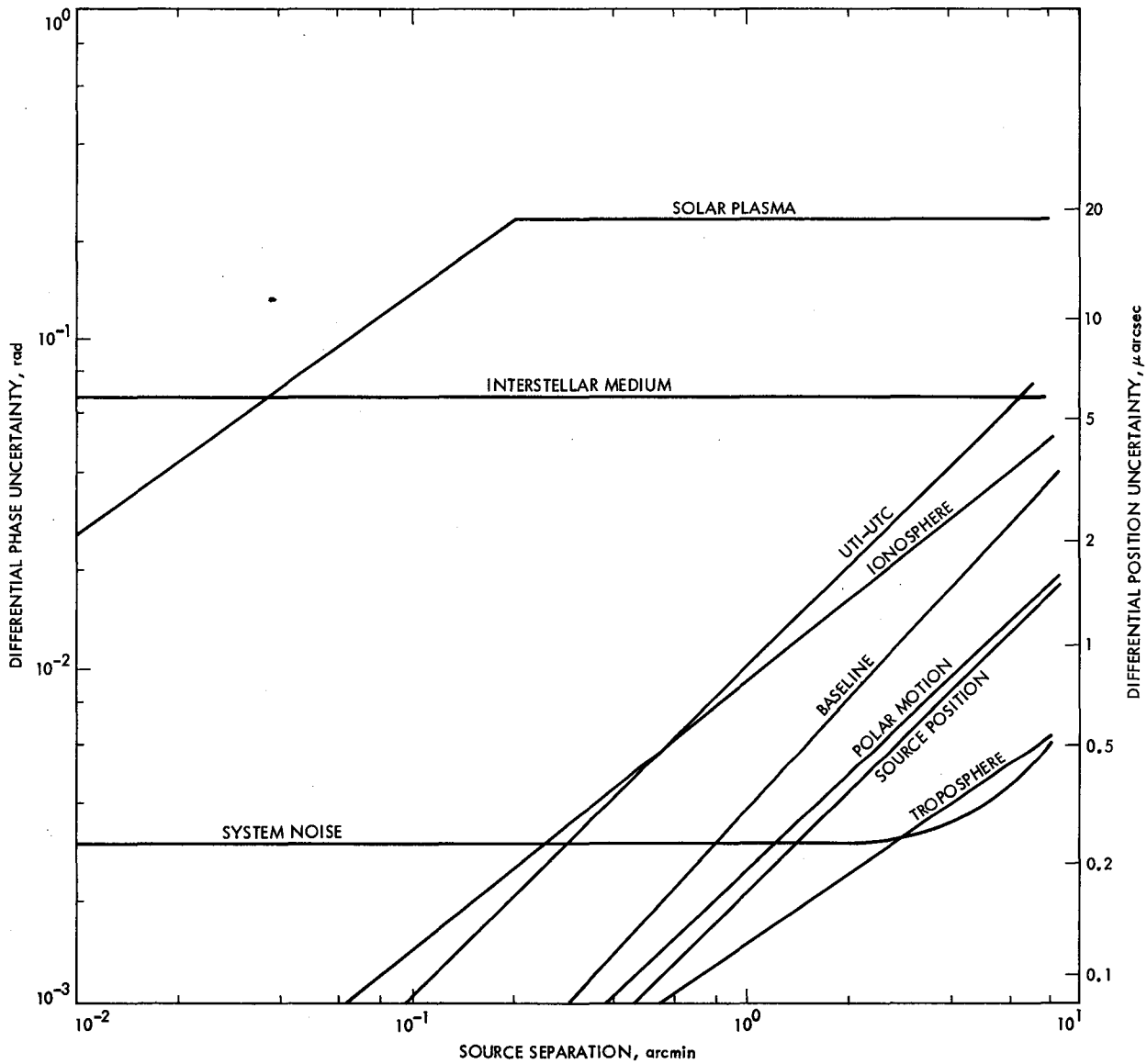


Fig. 4. S-band SVLBI errors vs angular separation between sources

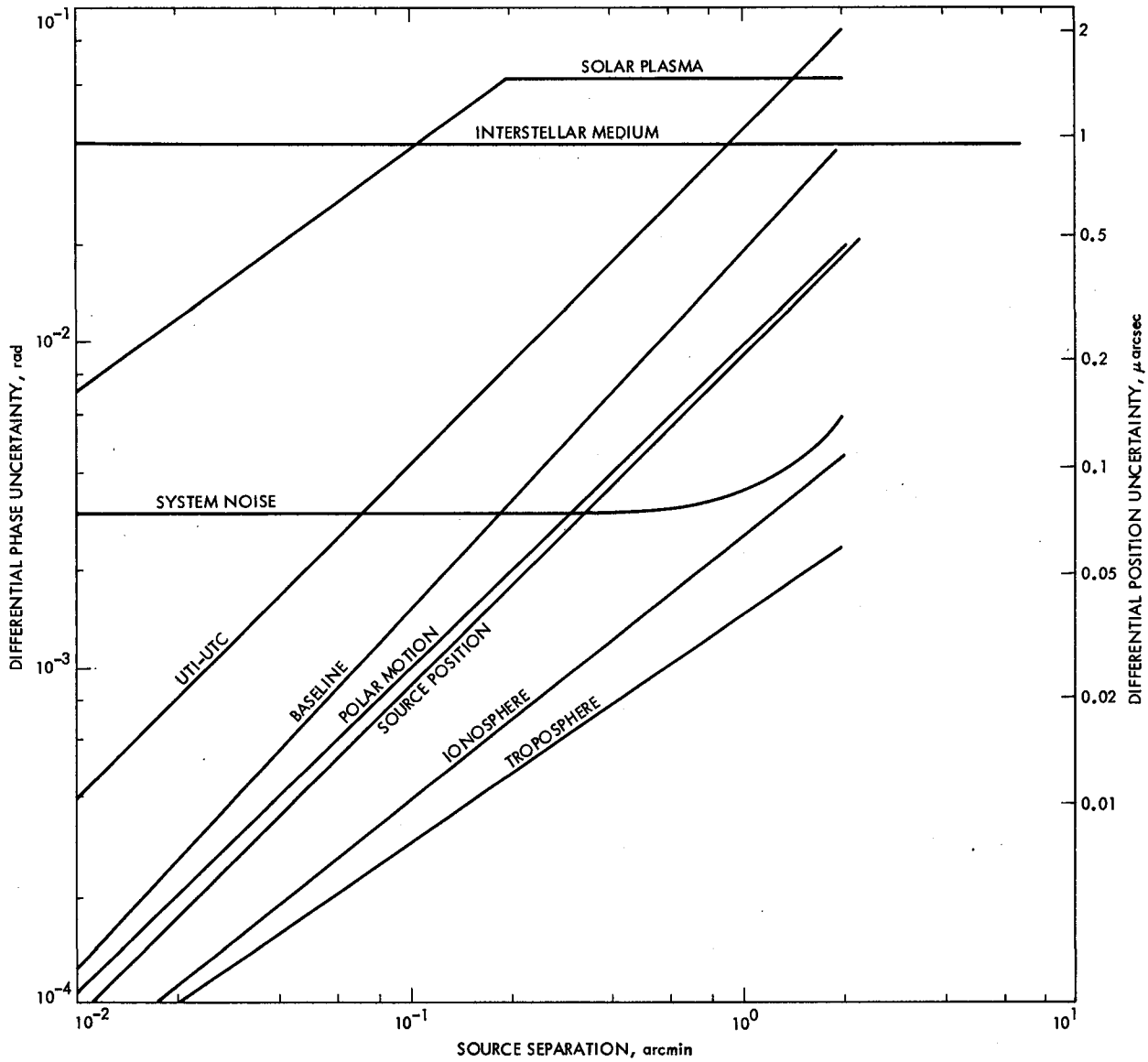


Fig. 5. X-band SVLBI errors vs angular separation between sources

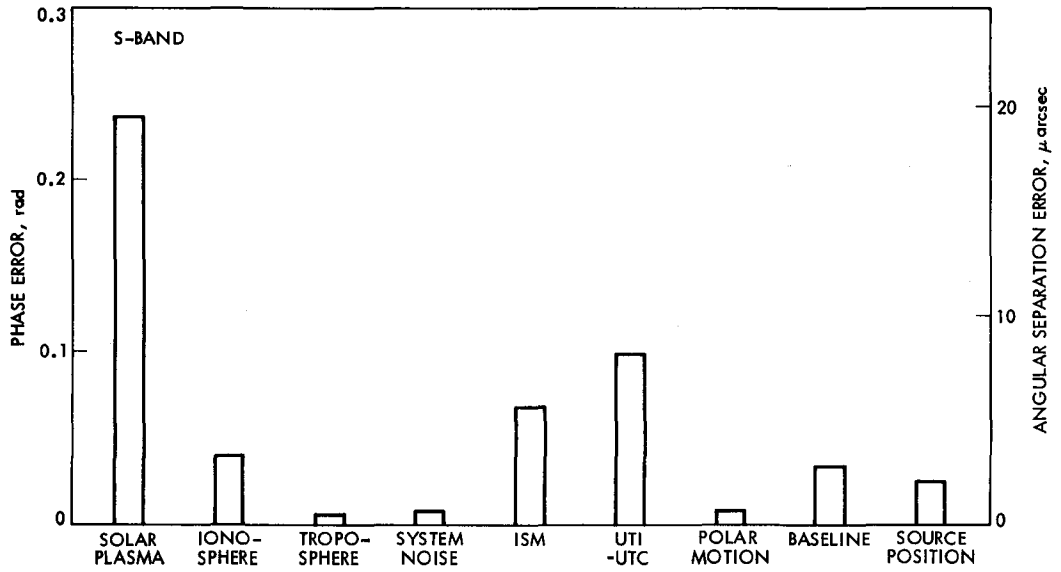


Fig. 6. S-band error bar chart for GC 1342+66

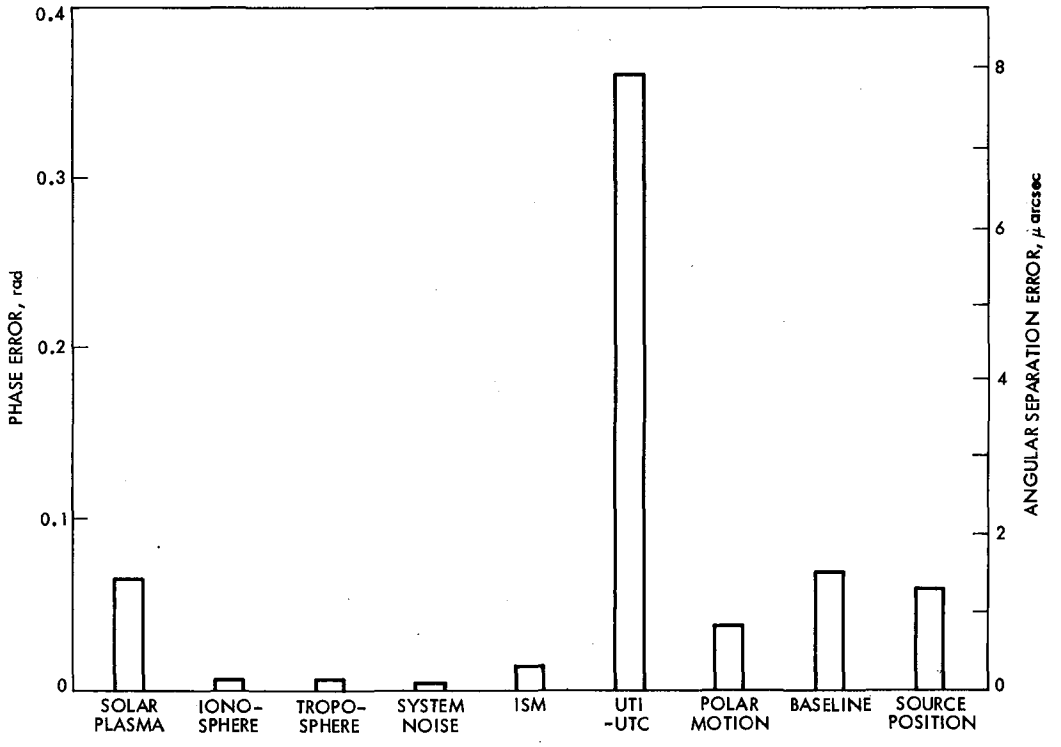


Fig. 7. X-band error bar chart for GC 1342+66

On Estimating the Phase of a Periodic Waveform in Additive Gaussian Noise – Part II

L. L. Rauch
Telecommunications Science and Engineering Division

Motivated by advances in signal processing technology that support more complex algorithms, a new look is taken at the problem of estimating the phase and other parameters of a periodic waveform in additive Gaussian noise. In Part I, the general problem was introduced and the maximum a posteriori probability criterion with signal space interpretation was used to obtain the structures of optimum and some suboptimum phase estimators for known constant frequency and unknown constant phase with an a priori distribution. In Part II, optimal algorithms are obtained for some cases where the frequency is a parameterized function of time with the unknown parameters and phase having a joint a priori distribution. In the last section, the intrinsic and extrinsic geometry of hypersurfaces is introduced to provide insight to the estimation problem for the small noise and large noise cases.

I. Introduction

The results of Part I (Ref. 1) are limited to the single-parameter estimation of phase with frequency f_c exactly known, based on an observation of duration $T = t_2 - t_1$. The maximum a posteriori probability (MAP) estimator $\hat{x}_0$ is the value of the phase x that maximizes

$$\frac{2}{N_0}\int_{t_1}^{t_2} z(\tau)\, y(f_c\,\tau + x)\, d\tau + \ln f_x(x) \qquad (17)^1$$

where N_0 is the one-sided power spectral density of the additive Gaussian noise, $z(\cdot)$ is the observed noisy signal, $y(\cdot)$ is the periodic waveform, and $f_x(\cdot)$ is the a priori probability density of the random phase. Appropriate signal space geometry is introduced and phase estimation results obtained for general periodic waveforms including, in particular, the squarewave case.

In what follows, the estimation model is extended to include phase, which is a polynomial function of time of given degree with coefficients to be estimated. The first-degree two-dimensional case (unknown frequency) is solved for small and large noise, and concepts of differential geometry are intro-

[1] Equation number from Part I.

duced to better understand and solve higher dimensional cases. The section and equation numbers continue from Part I.

VII.[2] Estimation of Phase and Frequency When the Frequency is Unknown but Constant

Using Eq. (13b)[1] we have, instead of Eq. (17),

$$\frac{2}{N_0}\int_{t_1}^{t_2} z(t)\,y(ft+x)\,dt + \ln f_f(f) + \ln f_x(x) \quad (79)^2$$

as the quantity to be maximized with respect to f and x to provide the MAP estimators $\hat{f}_0$ and $\hat{x}_0$. Then Eq. (21) is replaced by

$$2\langle \vec{z}, \vec{y}(f,x)\rangle + N_0 \ln f_f(f) + N_0 \ln f_x(x) \qquad (80)$$

where

$$\vec{z} = \vec{y}(f_0, x_0) + \vec{n} \qquad (81)$$

and

$$y(ft+x) = A_c \cos 2\pi(ft+x) \qquad (82)$$

for the present. In the previous cases of known frequency, T could be chosen to make fT an integer and therefore make $\|\vec{y}(f,x)\|$ and $\|\vec{z}\|$ not functions of x. In the present case of unknown f, this is not possible and $\vec{y}(f,x)$ lies on a closed hypersurface which is in effect the surface of a hypersphere with ripples. However, if we assume $fT \gg 1$, as is often the case of interest, then the fractional variation in $\|\vec{y}(f,x)\|$ approaches zero and

$$\lim_{fT\to\infty} \frac{1}{T}\|\vec{y}(f,x)\| = \lim_{fT\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} A_c^2 \cos^2 2\pi(ft+x)\,dt$$

$$= \frac{A_c^2}{2} \qquad (83)$$

The vector $\vec{y}(f,x)$ describes a two-dimensional curved surface parameterized by f and x. For fixed f, the cross section of the surface is a circle lying in a plane through the origin. The f,x coordinate system on the surface is locally orthogonal and the surface is embedded in an infinite-dimensional linear space. Some feeling for the curvature of the surface is given by Fig. 1. Calculation shows that

$$\left\langle \vec{y}(f,x),\ \vec{y}\left(f+\frac{m}{T}, x\right)\right\rangle = 0 \qquad (84a)$$

for n any nonzero integer and

$$\left\langle \vec{y}(f,x),\ \vec{y}\left(f, x+\frac{1}{4}\right)\right\rangle = 0 \qquad (84b)$$

Figure 1 indicates that in going from $\vec{y}(f,x)$ to $\vec{y}(f+1/T,x)$, $\vec{y}(f+1/2T,x)$ leaves the plane of $\vec{y}(f,x)$ and $\vec{y}(f+1/T,x)$ by an angle of 25.8° while making angles of 50.5° with $\vec{y}(f,x)$ and $\vec{y}(f+1/T,x)$. (In this signal space there are many directions normal to the above plane and the normal N dropped from $\vec{y}(f+1/2T,x)$ to the plane is only one of them.)

Taking the hint from Eqs. (84), a convenient orthonormal basis $\vec{u}_i$, $\vec{v}_i$ for the infinite dimensional flat (euclidean) space L_y spanned by the two-dimensional curved surface described by $\vec{y}(f,x)$ is given by

$$u_i(t) = \sqrt{\frac{2}{T}}\cos 2\pi\left(f_c + \frac{i}{T}\right)t \qquad (85a)$$

$$v_i(t) = \sqrt{\frac{2}{T}}\sin 2\pi\left(f_c + \frac{i}{T}\right)t \qquad (85b)$$

with i running over all integers, the time interval is $t_1 = -T/2$, $t_2 = T/2$, and $f_c T \gg 1$. For the case of Eq. (82),

$$\vec{y}(f,x) = \sum_{i=-\infty}^{\infty} (y_{ui}\vec{u}_i + y_{vi}\vec{v}_i) \qquad (86)$$

where

$$y_{ui}(f,x) = A_c\sqrt{\frac{T}{2}}(\cos 2\pi x)\frac{\sin\pi[(f_c-f)T+i]}{\pi[(f_c-f)T+i]} \qquad (87a)$$

$$y_{vi}(f,x) = A_c\sqrt{\frac{T}{2}}(\sin 2\pi x)\frac{\sin\pi[(f_c-f)T+i]}{\pi[(f_c-f)T+i]} \qquad (87b)$$

[2]Section and equation numbers continue from Part I.

If we define

$$I_i = \int_{-T/2}^{T/2} z(t) \cos 2\pi \left(f_c + \frac{i}{T}\right) t \, dt \tag{88a}$$

$$Q_i = \int_{-T/2}^{T/2} z(t) \sin 2\pi \left(f_c + \frac{i}{T}\right) t \, dt \tag{88b}$$

then the coordinates z_{ui}, z_{vi} of $\vec{z}$ are given by

$$z_{ui} = \langle \vec{z}, \vec{u}_i \rangle = \sqrt{\frac{2}{T}} I_i \tag{89a}$$

$$z_{vi} = \langle \vec{z}, \vec{v}_i \rangle = \sqrt{\frac{2}{T}} Q_i \tag{89b}$$

Now

$$\langle \vec{z}, \vec{y}(f, x) \rangle = \sum_{i=-\infty}^{\infty} (z_{ui} y_{ui} + z_{vi} y_{vi}) \tag{90}$$

Substituting from Eqs. (87) and (89) gives

$$\langle \vec{z}, \vec{y}(f, x) \rangle = \sum_{i=-\infty}^{\infty} A_c (I_i \cos 2\pi x + Q_i \sin 2\pi x) \times \frac{\sin \pi[(f_c - f)T + i]}{\pi[(f_c - f)T + i]} \tag{91}$$

for substitution in Eq. (80). This is then maximized to find the MAP estimate of $\hat{x}_0$ and $\hat{f}_0$.

For the case of small noise, the a priori probability terms in Eq. (80) generally become negligible and the MAP estimator $\hat{f}_0$, $\hat{x}_0$ is obtained by maximizing $\langle \vec{z}, \vec{y}(f, x) \rangle$ in Eq. (91). Following the argument on page 159 of Part I, Eq. (54) is replaced by

$$\sigma^2(\hat{x}_0) = \frac{\frac{N_0}{2}}{\left(\frac{\partial s}{\partial x}\right)^2} \tag{92}$$

$$\sigma^2(\hat{f}_0) = \frac{\frac{N_0}{2}}{\left(\frac{\partial s}{\partial f}\right)^2} \tag{93}$$

Now

$$\left(\frac{\partial s}{\partial x}\right)^2 = \left\| \frac{\partial}{\partial x} \vec{y}(f, x) \right\|^2 = \int_{-T/2}^{T/2} \left[\frac{\partial}{\partial x} y(ft + x)\right]^2 dt \tag{94}$$

and by Eq. (82),

$$\left(\frac{\partial s}{\partial x}\right)^2 = 4\pi^2 A_c^2 \int_{-T/2}^{T/2} \sin^2 2\pi(ft + x) \, dt = 2\pi^2 A_c^2 T \tag{95}$$

since $T \gg 1/f$. Then by Eq. (92),

$$\sigma^2(\hat{x}_0) = \frac{\frac{N_0}{2}}{2\pi^2 A_c^2 T} = \frac{1}{4\pi^2} \frac{N_0}{A_c^2 T} \tag{96}$$

which is the same as the result in Eq. (54) for a sinusoid of known frequency. Thus for the small noise case, the phase error is not increased if the frequency is unknown (but constant).

Proceeding with the estimate $\hat{f}_0$ of the frequency f_0 we have

$$\left(\frac{\partial s}{\partial f}\right)^2 = \left\| \frac{\partial}{\partial f} \vec{y}(f, x) \right\|^2 = \int_{-T/2}^{T/2} \left[\frac{\partial}{\partial f} y(ft + x)\right]^2 dt \tag{97}$$

and by Eq. (82),

$$\left(\frac{\partial s}{\partial f}\right)^2 = 4\pi^2 A_c^2 \int_{-T/2}^{T/2} t^2 \sin^2 2\pi(ft + x) \, dt = \frac{\pi^2}{6} A_c^2 T^3 \tag{98}$$

since $T \gg 1/f$. Then by Eq. (93),

$$\sigma^2(\hat{f}_0) = \frac{3}{\pi^2} \frac{N_0}{A_c^2 T^3} \tag{99}$$

From Eqs. (96) and (99),

$$T = 2\sqrt{3} \frac{\sigma(\hat{x}_0)}{\sigma(\hat{f}_0)} \tag{100}$$

$$\frac{P}{N_0} = \frac{A_c^2/2}{N_0} = \frac{1}{16\sqrt{3}\,\pi^2} \cdot \frac{\sigma(\hat{f}_0)}{\sigma^3(\hat{x}_0)} \tag{101}$$

For example, with $\sigma(\hat{x}_0) = 0.01$ cycles (3.6°) and $\sigma(\hat{f}_0) = 1$ Hz we have $T = 0.0346$ s and $P/N_0 = 3656$ (35.6 dB). The frequency spacing $1/T$ of the correlators in Eq. (88) would then be 28.87 Hz. Both the frequency spacing and the required P/N_0 are directly proportional to $\sigma(\hat{f}_0)$.

Returning now to the general case, let us consider the maximization of Eq. (80). For the initial estimate, we generally have $f_x(\cdot) = 1$ and $f_f(\cdot)$ constant over some interval and zero elsewhere. Thus let us consider the maximization of Eq. (91). Define

$$\alpha = (f_c - f)T \tag{102}$$

and

$$I(\alpha) = \sum_{i=-\infty}^{\infty} I_i \frac{\sin \pi(\alpha - i)}{\pi(\alpha - i)} \tag{103a}$$

$$Q(\alpha) = \sum_{i=-\infty}^{\infty} Q_i \frac{\sin \pi(\alpha + i)}{\pi(\alpha + i)} \tag{103b}$$

so that Eq. (91) becomes

$$\langle \vec{z}, \vec{y}(f, x) \rangle = A_c [I(\alpha) \cos 2\pi x + Q(\alpha) \sin 2\pi x] \tag{104}$$

$I(\alpha)$ and $Q(\alpha)$ are simply the cardinal series (minimum bandwidth) interpolations of the samples I_i and Q_i respectively, with a sampling rate of unity in the α coordinate. Now by Eqs. (81), (82), (86), (89), and (68) the I_i and Q_i are simply $\sqrt{T/2}\, y_{0ui}$ and $\sqrt{T/2}\, y_{0vi}$, respectively, plus independent Gaussian random variables of mean zero and variance $N_0 T/4$. By Eq. (87)

$$\sqrt{T/2}\, y_{0ui} = \frac{A_c T}{2} (\cos 2\pi x_0) \frac{\sin \pi(\alpha_0 + i)}{\pi(\alpha_0 + i)} \tag{105a}$$

$$\sqrt{T/2}\, y_{0vi} = \frac{A_c T}{2} (\sin 2\pi x_0) \frac{\sin \pi(\alpha_0 + i)}{\pi(\alpha_0 + i)} \tag{105b}$$

Thus, I_i and Q_i are samples of

$$\frac{A_c T}{2} (\cos 2\pi x_0) \frac{\sin \pi(\alpha - \alpha_0)}{\pi(\alpha - \alpha_0)} \tag{106a}$$

and

$$\frac{A_c T}{2} (\sin 2\pi x_0) \frac{\sin \pi(\alpha - \alpha_0)}{\pi(\alpha - \alpha_0)} \tag{106b}$$

respectively, taken at integer values of α plus the random variables described above. Now the function of α in Eq. (106) is band limited at half the sampling rate and so the interpolated functions of Eq. (103) are simply the functions of Eq. (106) plus Gaussian band-limited white noise of bandwidth 1/2 and variance $N_0 T/4$. Thus

$$I(\alpha) = \frac{T}{2} \left[A_c (\cos 2\pi x_0) \frac{\sin \pi(\alpha - \alpha_0)}{\pi(\alpha - \alpha_0)} + n_I(\alpha) \right] \tag{107a}$$

$$Q(\alpha) = \frac{T}{2} \left[A_c (\sin 2\pi x_0) \frac{\sin \pi(\alpha - \alpha_0)}{\pi(\alpha - \alpha_0)} + n_Q(\alpha) \right] \tag{107b}$$

where $n_I(\alpha)$ and $n_Q(\alpha)$ are independent Gaussian band-limited white random processes with

$$\sigma^2 [n_I(\alpha)] = \sigma^2 [n_Q(\alpha)] = \frac{N_0}{2} \tag{108}$$

Returning now to Eq. (104) we observe that $\langle \vec{z}, \vec{y}(f, x) \rangle$ can be maximized in two steps. In the first step we choose $x(\alpha)$ to maximize the expression for each α. A necessary condition is

$$\frac{\partial}{\partial x} \langle \vec{z}, \vec{y}(f, x) \rangle = 2\pi A_c [-I(\alpha) \sin 2\pi x + Q(\alpha) \cos 2\pi x] = 0 \tag{109}$$

which gives

$$\tan 2\pi x = \frac{Q(\alpha)}{I(\alpha)} \tag{110}$$

This has two solutions for x, one giving the absolute maximum and the other the absolute minimum of Eq. (104) with respect to x. The value of x in Eq. (110) giving the maximum is that which causes $\cos 2\pi x$ to have the same sign as $I(\alpha)$ and $\sin 2\pi x$ to have the same sign as $Q(\alpha)$. From Eq. (110) and the above,

$$\sin 2\pi \hat{x}_0 = \frac{Q(\alpha)/|I(\alpha)|}{\sqrt{1 + \left(\frac{Q(\alpha)}{I(\alpha)}\right)^2}} \tag{111a}$$

$$\cos 2\pi \hat{x}_0 = \frac{\operatorname{sgn}[I(\alpha)]}{\sqrt{1 + \left(\frac{Q(\alpha)}{I(\alpha)}\right)^2}} \tag{111b}$$

Substitution in Eq. (104) gives

$$\langle \vec{z}, \vec{y}(f, \hat{x}_0) \rangle = A_c \sqrt{I^2(\alpha) + Q^2(\alpha)} \tag{112}$$

Thus, by Eq. (102) $\hat{f}_0 = f_c - (\hat{\alpha}_0/T)$ where $\hat{\alpha}_0$ is the value of α which maximizes

$$\frac{(\vec{z}, \vec{y}(f, \hat{x}_0))^2}{A_c^2} = I^2(\alpha) + Q^2(\alpha) \tag{113}$$

The large noise behavior of $\hat{\alpha}_0$ may be studied by using Eq. (107) to substitute for $I(\alpha)$ and $Q(\alpha)$ in Eq. (113). The actual estimator must use Eqs. (103), (113), and (111). Even in the absence of noise, Eq. (113) has many stationary points (relative maxima and minima) and so an indirect estimator is a problem. Since the maxima and minima of Eq. (113) are always stationary points ($n_I(\alpha)$ and $n_Q(\alpha)$ have all derivatives), one approach would be to solve

$$\frac{d}{d\alpha}(I^2(\alpha) + Q^2(\alpha)) = 0 \tag{114}$$

for all of the stationary points in the a priori interval determined by $f_f(\alpha)$ and then evaluate Eq. (113) at these points to locate the absolute maximum. For a particular noise realization, Eq. (112) will generally not be an even function about its maximum and so the MAP estimator is generally not the minimum mean square error (MSE) or mean absolute error (MAE) estimator. In the case of a general a priori distribution $f_f(\cdot)$ on f and noise which is not small, it is necessary to substitute Eq. (112) in Eq. (80) and seek the maximum of

$$2A_c \sqrt{I^2(\alpha) + Q^2(\alpha)} + N_0 \ln f_\alpha(\alpha) \tag{115}$$

where $f_\alpha(\cdot)$ is easily obtained from $f_f(\cdot)$ in the usual way.

VIII. Some Geometrical Aspects

In the more general case where Eq. (82) is replaced by

$$y\left(x + ft + \dot{f}\frac{t^2}{2} + \cdots + f^{(N-2)}\frac{t^{N-1}}{(N-1)!}\right) \tag{116}$$

then the terminal points of

$$\vec{y}(x, f, \dot{f}, \cdots, f^{(N-2)}) \tag{117}$$

form an N-dimensional subspace (generally nonlinear) that is coordinatized by the N parameters to be estimated and that is imbedded in the infinite dimensional linear space[3] in which the observed signal plus noise is represented. As we have seen, the estimation problem involves such questions as the point in the parameter subspace that is at minimum distance from an arbitrary point in the imbedding linear signal (plus noise) space and the effect of the noise on the location of the point in the parameter subspace. The solution of the estimation problem depends on both the intrinsic (Ref. 3) and extrinsic geometrical properties of the nonlinear parameter space with the small noise results depending only on the intrinsic geometry. To gain insight, we examine several cases.

First case: Phase estimation only with known frequency f_c ($N = 1$). This was treated analytically in Section VI (Part I) with the periodic waveform given by Eq. (55). The geometrical situation is illustrated in Fig. 2. The one-dimensional parameter space (path) lies on the surface of a $2n$ dimensional hypersphere (and is therefore nonlinear). For this one-dimensional case, the intrinsic geometry of the path is completely described by the single scalar ds/dx, given by Eq. (56), with the small noise result given by the second term of Eq. (54). If $y(\cdot)$ has one or more discontinuities, such as a square wave, then $ds/dx \to \infty$ and the small-noise performance becomes arbitrarily good (the closed path, while continuous, has no derivative and is infinite in length).

For large noise, the extrinsic geometry of the path becomes important. The radius of curvature (*ROC*) (Ref. 4) is given by (see Eq. 55, Part I)

$$ROC = \sqrt{\frac{T}{2}} \frac{\sum_{j=1}^{n} j^2\left(a_j^2 + b_j^2\right)}{\sqrt{\sum_{j=1}^{n} j^4\left(a_j^2 + b_j^2\right)}} \tag{118}$$

which vanishes if there is a discontinuity in $y(\cdot)$ or its derivatives. The distance between any two points on the path, $\vec{y}(x_1)$ and $\vec{y}(x_2)$, is given by

$$\|\vec{y}(x_1) - \vec{y}(x_2)\|^2 = T\sum_{j=1}^{n}(a_j^2 + b_j^2)\,[1 - \cos 2\pi j(x_1 - x_2)] \tag{119}$$

which, of course, never exceeds twice the radius of the hypersphere in whose surface the path is imbedded. Note that the above extrinsic properties are invariant over the path.

Second case: Phase and frequency estimation with unknown constant frequency ($N = 2$) and sinusoidal waveform. This was treated analytically in Section VII. The two-dimensional parameter space (surface) is nonlinear as indicated by Fig. 1. Even for the special case of $y(\cdot)$ sinusoidal, the two-

[3] An N-dimensional noneuclidean space can always be imbedded in euclidean space of not more than $N(N+1)/2$ dimensions (Ref. 2).

dimensional surface spans (must be imbedded in) an infinite dimensional signal space. It is something like a tube extending through all of the dimensions of the signal space as f runs over its range. The intrinsic geometry of the surface is completely described by (Ref. 5)

$$ds^2 = \sum_{i,j=1}^{2} g_{ij} du^i du^j \tag{120}$$

where s is the distance in the surface, $u^1 = x$ and $u^2 = f$ are the contravariant coordinate values in the surface, and g_{ij} is the second-order covariant metric tensor of the surface. Calculation gives

$$\begin{aligned} g_{11} &= 2\pi^2 A_c^2 T \\ g_{22} &= \frac{\pi^2 A_c^2 T^3}{6} \\ g_{12} &= g_{21} = 0 \end{aligned} \tag{121}$$

Thus the f, x coordinate system in the surface is everywhere orthogonal ($g_{12} = g_{21} = 0$) and uniform (g_{11} and g_{22} are constant). Also the surface is developable, that is, it has no intrinsic curvature. Now

$$\left(\frac{ds}{dx}\right)^2 = g_{11}, \quad \left(\frac{ds}{df}\right)^2 = g_{22} \tag{122}$$

and the small noise result has been given in Eqs. (96) and (99).

The extrinsic geometry of the surface is given by Eqs. (86) and (87) and the large noise estimation case has been considered begining with Eq. (102).

Third case: Phase, frequency, and frequency-rate estimation with unknown constant frequency rate ($N = 3$). The surface of the second case now becomes a three-dimensional volume by the coordinate $\dot{f}$, which turns out not to be orthogonal to the surface coordinatized by x and f. For small noise, the intrinsic geometry is determined by

$$ds^2 = \sum_{i,j=1}^{3} g_{ij} du^i du^j \tag{123}$$

with $u^1 = x$, $u^2 = f$, $u^3 = \dot{f}$. g_{11}, g_{22}, g_{12}, and g_{21} have the same values as in the second case. Calculation gives

$$g_{33} = \frac{\pi^2 A_c^2 T^5}{160}$$

$$\begin{aligned} g_{23} &= g_{32} = 0 \\ g_{13} &= g_{31} = \frac{\pi^2 A_c^2 T^3}{12} \end{aligned} \tag{124}$$

Thus the $\dot{f}$ coordinate is orthogonal to the f coordinate and the angle θ between the $\dot{f}$ and x coordinates is given by

$$\cos\theta = \frac{g_{13}}{\sqrt{g_{11} g_{33}}} = \frac{\sqrt{5}}{3} = 0.745$$

$$\theta = 42^\circ \qquad \sin\theta = \frac{2}{3} \tag{125}$$

So in the small noise case, the errors in the estimates $\hat{\dot{f}}_0$ and $\hat{x}_0$ have a correlation of 0.745 and the errors in the other two pairs of estimates are uncorrelated. As in the two-dimensional case, the intrinsic geometry of the three-dimensional manifold has no curvature (g_{ij} components are constants).

For the small noise case

$$\sigma^2(\hat{\dot{f}}_0) = \frac{N_0/2}{\left(\frac{ds}{df}\right)^2} = \frac{N_0/2}{g_{33}} \tag{126}$$

giving

$$\sigma^2(\hat{\dot{f}}_0) = \frac{80}{\pi^2} \frac{N_0}{A_c^2 T^5} \tag{127}$$

For the example following Eq. (101), where it was found that $\sigma(\hat{x}_0) = 0.01$ cycles and $\sigma(\hat{f}_0) = 1$ Hz, with $T = 0.0346$ s and $P/N_0 = 3656$, the result for $\hat{\dot{f}}_0$ is

$$\sigma(\hat{\dot{f}}_0) = 149.5 \text{ Hz/s}$$

If T is increased by a factor of 10 to 0.346 s, the above becomes

$$\sigma(\hat{\dot{f}}_0) = 0.47 \text{ Hz/s}$$

The above results can be extended in a straightforward way to the estimation of the higher derivatives of frequency ($N > 3$) if a priori knowledge indicates they are significant.

IX. Algorithms

Computational algorithms based on the preceding results such as Eqs. (88), (103), (113), (111), and (102) are feasible with current digital signal processing technology and would

provide a sequence of MAP estimates at a rate $1/T$. Such algorithms are quite different from phase-locked loops, which are frequently used for phase and frequency estimation. The phase-locked loop is an invented causal dynamical system that does not provide an optimum estimate based on a criterion. The phase-locked loop, like any real feedback system, has its own dynamics and stability considerations, which are not really a part of the estimation problem. The stability considerations of third- and higher-order phase-locked loops are not present in algorithms based on the results of this paper.

References

1. Rauch, L. L., "On Estimating the Phase of a Periodic Waveform in Additive Gaussian Noise–Part I," *DSN Progress Report 42-45*, pp. 152-164, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1978.
2. Eisenhart, L. P., *Riemannian Geometry*, Second Edition, Section 55, Princeton University Press, Princeton, New Jersey, 1949.
3. Eisenhart, L. P., *An Introduction to Differential Geometry with Use of the Tensor Calculus*, Chapter III, Princeton University Press, Princeton, New Jersey, 1940.
4. Ibid., Section 4.
5. Ibid., Section 24.

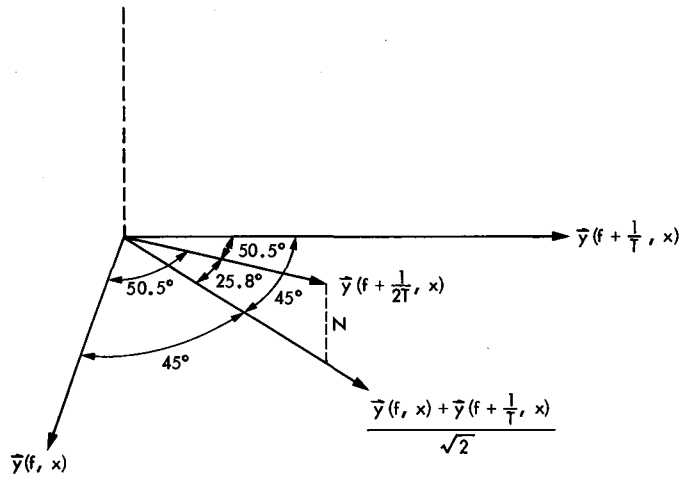
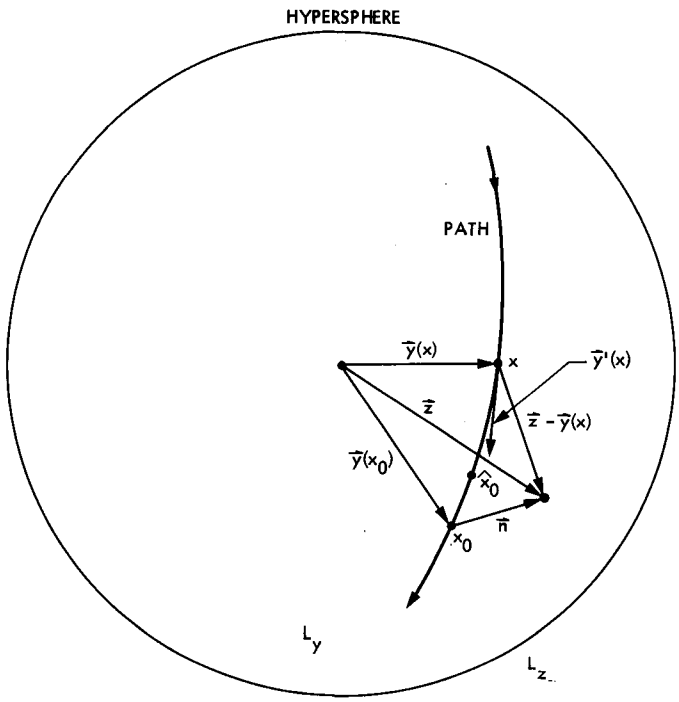


Fig. 1. Surface curvature



$\hat{x}_0$ IS POINT ON PATH NEAREST TO $\vec{z}$

$\hat{x}_0$ MINIMIZES ANGLE BETWEEN $\vec{y}(\hat{x}_0)$ AND $\vec{z}$

Fig. 2. Paths on hypersphere

VLBI Solutions for the Time Variation of DSN Baselines: 1978 to 1983

R. N. Treuhaft, J. L. Fanselow, K. M. Liewer, A. E. Niell, O. J. Sovers, J. B. Thomas, and K. S. Wallace
Tracking Systems and Applications Section

Very Long Baseline Interferometry (VLBI) results are presented for the two baseline vectors between the Goldstone DSN antenna complex and the overseas sites at Canberra, Australia and Madrid, Spain. Results from solutions using data taken between 1978 September and 1983 May show an apparent California-Spain baseline length increase of 21 cm during this time span, while the California-Australia length has remained constant. Statistical investigations of the integrity of the data are discussed along with dominant systematic error sources and their effect on baseline length determination. Results and interpretation of the time behavior of the angle between DSN baselines are also described.

I. Introduction

Three principal concerns motivate the study of changes in intercontinental baseline vectors. Navigational demands, which require the location of the DSN stations in an inertial radio reference frame, necessitate modeling relative motion of points on the Earth. Secondly, because DSN baselines span the North American, Pacific, Indian, and Eurasian plates, detection of their motion will aid in the understanding of continental drift, which is of geophysical interest. Finally, since VLBI solutions entail the adjustment of hundreds of correlated parameters, the plausibility of any changes in the estimated baselines supports the integrity of the entire solution.

This article focuses on the two baseline lengths and the single angle (inner angle) between the baselines because, for any two vectors, these three quantities are invariant under rotations of the coordinate system in which they are presented. Rotational invariance is an important criterion for parameters studied in baseline time-behavior analyses because, in VLBI, the orientation of a baseline is determined with respect to an inertial reference frame defined by extragalactic radio sources (EGRS), and we wish to measure the baselines in an Earth-fixed reference frame. Since translations relative to the EGRS can be ignored, transforming a baseline from the EGRS frame to the Earth-fixed frame requires only a rotation. The necessary rotations are monitored regularly by services such as the Bureau International de l'Heure (BIH), TEMPO at the Jet Propulsion Laboratory, and POLARIS at the National Geodetic Survey (Ref. 1).

To study the baseline variations in an Earth-fixed frame, we could (1) use the Earth orientation measurements of the above services to transform the EGRS-referenced VLBI results to the Earth-fixed frame, and/or (2) consider the variations of baseline parameters which are unchanged in the transformation between the two reference frames. Since the errors in the Earth orientation measurements are often larger than the intrinsic VLBI errors for the data set considered[1] (Ref. 2), we choose

[1] Treuhaft, R. N., "The Time Variation of Intercontinental Baselines using VLBI," IOM 335.1-113, June 1983, private communication to Tracking Systems and Applications Section, Jet Propulsion Laboratory.

to study the rotationally invariant lengths and inner angle. The inner angle, as measured, is actually quasi-rotationally invariant as will be discussed later. Future studies of this type will also include rotationally sensitive quantities as the accuracy of Earth orientation measurements improves.

The parameter estimates which are discussed in this article were obtained from multiparameter adjustments using dual-frequency VLBI data. The data were taken between 1978 September and 1983 May on the California-Australia (**C-A**) and California-Spain (**C-S**) baselines. 3090 delay and delay rate observations of 132 extragalactic sources comprise the data set analyzed. A general description of the 1971 to 1980 VLBI observation and fitting procedure is given in Ref. 3. The baseline results in this article are estimated independently for each experiment, as opposed to estimating a single baseline for the entire span of data; in Ref. 3, results from both methods of estimation are reported. Three requirements placed on the data currently analyzed force the exclusion of experiments reported in Ref. 3 which were performed before 1978 September: (1) All experiments considered here contain delay and delay rate data (as opposed to delay rate only); (2) All the data were taken at S- and X-band simultaneously for charged-particle calibration; and (3) Hydrogen maser time standards were used at both stations of each baseline. Comparison of the estimated global parameters (e.g., source positions) of Ref. 3 with those derived here yields agreement within quoted errors. In addition to solving for baseline components and source positions, the fit to the VLBI data includes solutions for station clock parameters and total (dry + wet) zenith troposphere delays for each session.

In section II of this article, we present baseline length results and interpretation. Section III contains a discussion of dominant systematic errors affecting the length measurements; troposphere mismodeling is given special attention. Section IV contains inner angle results, and section V, the conclusions.

II. Baseline Length Results and Interpretation

The solutions for baseline components and lengths obtained from each experiment are given in Table 1. As mentioned above, for single-baseline VLBI solutions, Earth orientation transformations must be applied to the VLBI results. The BIH transformations were used for the components presented in Table 1. This Table also gives the DSS 13-DSS 14 short-baseline tie used to refer all measurements to the 64-meter antenna at Goldstone (DSS 14). The results for the baseline lengths plotted as a function of time are shown in Figs. 1(a) and (b) for **C-A** and **C-S**, respectively. The linear fit to the data was performed by including all the correlations between the baseline coordinate parameters in the global fit from which the lengths were derived. From Fig. 1, the baseline time rate of change for **C-A** is statistically insignificant and, from the chi-square per degree of freedom (χ^2/DF) of 0.97, the time variation of the **C-A** baseline is consistent with a linear fit, or indeed with no motion at all. For **C-S**, the baseline rate is six times its formal error, and is therefore significant with regard to that measure of error. However the large χ^2/DF means that there may be problems with the linear parameterization of the data. The probability of getting a χ^2/DF of 1.98 or greater with twelve degrees of freedom is about 2%, assuming that the measured length variation results from a pure linear drift degraded only by Gaussian noise. Below, we assess the validity of this assumption and interpret the result of the chi-square test.

A linear model for intercontinental baseline motion is potentially applicable because points in the interior of continental plates are not expected to respond to high frequency driving forces applied at the boundaries (Ref. 4). Of the DSN stations, the Goldstone complex is the only one very close to a known plate boundary – the San Andreas Fault. In addition to any tectonic motion, local, small-scale phenomena far from plate boundaries near either station of a baseline could produce abrupt changes in the long baseline results. Therefore, the quality of linear fits over the five-year data span may not be a good indicator of the integrity of the data, and we must explore the possibility of abrupt changes in baseline length. Table 2 shows the lengths of Fig. 1 averaged for the 1978 to 79, 1979 to 80, and 1982 to 83 observation clusters, or epochs. Note that the Table shows a statistically significant change in baseline length between the second and third epochs for **C-S**, while the **C-A** lengths are remarkably constant over the three epochs.

Concentrating on the more controversial **C-S** results, it is possible that abrupt motion near Goldstone or Madrid produced the 21-cm change in the **C-S** baseline length between the last two epochs in the Table. However, the Crustal Dynamics Project (CDP) regional VLBI data show no such significant motion on any of the baselines containing Goldstone over the same time span (Ref. 5). It should be noted that since the CDP data do not preclude large (20-cm) changes in the Goldstone vertical, and since the Goldstone vertical has a 66% projection onto the **C-S** length, the CDP data alone do not exclude the possibility that much of the 21-cm jump was due to Goldstone local vertical motion. However, the Goldstone vertical also has an 83% projection onto the **C-A** length, which is relatively constant over the same three-year span. Considering both the CDP and the **C-A** data, it is unlikely that Goldstone local motion could be responsible for the 21-cm shift observed in the **C-S** data. We do not have a firm case for excluding the local motion possibility at Madrid, since we do not have available

any measurements to local reference points for the Spain complex. A geophysical phenomenon as large as a 21-cm change over three years would probably call attention to itself via ground survey techniques. The frequency of past measurements by survey as well as VLBI techniques must be investigated. The appropriateness of the linear model for the **C-S** baseline variation is therefore contingent upon evidence excluding the possibility of local, abrupt motion near the Madrid complex.

If such evidence finally does exclude the possibility of abrupt motion, the 2% probability quoted above for the chi-square test is an indication of a problem in the error analysis: namely that it is highly improbable that the actual length errors are Gaussian with standard deviations given by the error estimates of Table 1. The error estimates could be incorrect due to unmodeled systematic errors, or due to underestimation of the statistical errors. If the apparent change in the **C-S** baseline is due to faulty error analysis, the step-function pattern in Fig. 1(b) suggests that unmodeled systematic effects are part of the problem. Further, a comparison of Figs. 1(a) and 1(b) indicates systematic problems may affect different parameters in different ways; for example, in this case, the **C-S** linear fit is suspect, and the **C-A** fit is acceptable.

III. Systematic Errors in Baseline Length

The baseline length is principally derived from observations of the interferometric time delay, between two ends of a baseline, of the arrival of an electromagnetic signal from a distant radio source. Uncalibrated phenomena affecting the measured delay can cause errors in baseline length estimates. If, for a group of measurements, the observation-to-observation delay errors are correlated with the sequence of partial derivatives of time delay with respect to baseline length, the error in the estimated length may be large. Since the partial derivative of delay with respect to baseline length parameters will be largest for low elevation angle observations, and since unmodeled troposphere delay errors are also likely to be large at low elevation angles, the errors will track the partials for many observation strategies.

Studies of the effect of troposphere mismodeling show that dry troposphere elevation mapping function fluctuations can cause 12-cm errors in **C-S** baseline and up to 25-cm errors in the **C-A** baseline. These mapping fluctuations are caused by changes in local meteorology at the two ends of the baseline and should be seasonally variable (Ref. 6). Wet troposphere fluctuations can also cause 5 to 10 cm of baseline length error and would be expected to vary from experiment to experiment (Ref. 7). Neither the wet nor the dry effect should exhibit the qualitative temporal behavior of the **C-S** baseline lengths shown in Table 2. The magnitude of these systematic effects is, however, within a factor of two of the **C-S** length changes observed, and, therefore, more extensive calibrations will be attempted before conclusions are drawn concerning the statistical significance of length changes. Acquiring surface meteorology and water vapor radiometer data should help to reduce the errors contributed by both of these tropospheric effects.

Another potentially important systematic effect which could contribute to baseline length errors can arise from instrumental effects which depend on time or elevation angle. Phase calibration of the data (Ref. 8) can help to remove such effects, but the results shown in Fig. 1 do not include phase calibration. For some experiments, phase calibration was impossible due to instrumentation failures, but calibration may be possible for a subset of the data. Preliminary tests show that instrumental miscalibration might also contribute as much as 10 cm to the baseline length error. The effects of changes in the data acquisition hardware between 1980 and 1982 have not been thoroughly investigated, and instrumental effects cannot be discounted.

IV. Inner Angle Results

Along with the baseline lengths, the inner angle between simultaneously measured **C-A** and **C-S** baseline vectors (approximately 82 degrees) is a rotational invariant, and its time behavior could be another good measure of station movement. In Fig. 2 the inner angle (θ) between the two baseline vectors is shown in the standard Earth-fixed coordinate system, with the x-axis passing near the Prime Meridian. Since the **C-S** and **C-A** baselines are not measured simultaneously, but are typically observed within one to three days of each other, the Earth's short-term orientation changes must be calibrated in order to avoid interpreting global rotations as relative station motion. Because the inner angle measurement is sensitive to short-term Earth rotations, but is insensitive to rotations with periods longer than the time between experiments, the inner angle is quasi-rotationally invariant and its time variation might yield information on continential motion in one direction perpendicular to the baseline vectors. Fig. 3 shows the inner angle as a function of time after applying corrections based on the BIH Earth orientation data. The large χ^2/DF for the linear fit is mostly due to the problematic second to last point on the plot. This point represents data taken on 23 January 1983 and 25 January 1983 for **C-A** and **C-S**, respectively.

To investigate whether that point's severe departure from the line could be due to high-frequency orientation errors, Fig. 4, reproduced from Ref. 9, shows the Kalman filtered BIH

Length of Day (LOD) data, which is a measure of the difference between the rate of change of the actual global rotation about the pole (UT1) and a universally accepted standard rate used in VLBI analysis (IAT). Predictable tidal changes in UT1 were removed before estimating the LOD. The peak of this plot is near the dates of the experiments determining the inner angle for the questionable point in 1983 January. An arrow points to the date of the problematic inner angle measurement. The peak in Fig. 4 is believed to result from the El Niño phenomenon (Ref. 9). To see if better high-frequency (UT1-IAT) data would track this peak more faithfully, the inner angle was recalculated with Kalman filtered POLARIS data for 1982 to 83 and lunar laser ranging data for the rest of the time span by T. M. Eubanks (personal communication), and the results are plotted in Fig. 5(a). The decreased χ^2/DF suggests that the fit is improving, but the best fit by far is that in Fig. 5(b) where the 1983 January point is eliminated completely. It is possible that the actual angular rate peak of Fig. 4 was beyond the high-frequency detection capability of any external service, since the typical sampling interval of the external services, such as BIH or POLARIS, is one week. For the 1983 January point our two experiments were 2.3 days apart, and it is possible that our measurements were more sensitive to high-frequency global orientation changes than any others. To investigate this possibility, we are currently studying the apparent angular rate of the baselines within each experiment of the suspect inner-angle pair. If Earth orientation changes are large between the experiments, they may also be large within a single experiment. A complete explanation of the behavior of the 1983 January point will require further analysis, but for now, the data suggest that poor high-frequency Earth orientation determination may plague the use of the inner angle to measure real continental motion.

It should be pointed out that the inner angle errors as well as the linear fit parameters were calculated including all the relevant correlations among the baseline coordinates in the global fit from which the inner angles were extracted. If correlations are ignored, the parameter estimates and errors change only at the 10 to 20% level, but the χ^2/DF values change enough to substantially alter the conclusions of hypothesis tests. This comment applies to the baseline length analyses as well.

V. Conclusions and Discussion

The baseline length results can be summarized as follows: based on formal baseline errors, there is a statistically significant change in the **C-S** length over the time spanned by the data. The change in the baseline is only twice the possible contribution of either of our two biggest known systematic error sources, the troposphere and instrumental calibration. Although the time variation of troposphere errors is not expected to induce the sort of time variation observed on the **C-S** baseline, the magnitude of the possible length error is sufficiently close to the observed length change that skepticism about the significance of that change seems appropriate for now. Instrumental effects, on the other hand, could conceivably cause the step-function behavior of the observed **C-S** baseline length, although we know of no major changes in the instrumentation between 1980 and 1982. The large χ^2/DF of the linear fit indicates the presence of unmodeled error sources or the possibility that the actual motion is not linear. Verification of the possibility of a nonlinear length change would require external evidence of local motion near the Spain complex.

The **C-A** length results are in good statistical agreement with no motion. It should be noted that until the above-mentioned systematic problems are resolved, the **C-A** baseline length results should also be suspected. The formal errors of the **C-A** baseline are, however, on the average larger than those of the **C-S** baseline and may dwarf the relevant systematics, perhaps explaining why the χ^2/DF for the **C-A** length linear fit is so close to 1.

The next step in the analysis of the length results is to examine further the consequences of unmodeled effects in the data. Current efforts include estimating the effect of errors in spin-axis orientation (precession and nutation) on baseline length estimates. We have also investigated the possibility of allowing the VLBI data to determine new troposphere mapping function parameters, which include station-to-station variations. Results from such studies show that indeed there is enough strength in the data to determine statistically significant changes to both the standard Chao (Ref. 10) and the new Lanyi (Ref. 6) mapping functions, both developed at JPL. The details of the new mapping function parameters, as well as the results, will be reported in the near future. Also to be reported are the effects on baseline lengths and source positions of the time-varying wet component of the troposphere.

Length rate analyses will soon depend critically on our ability to determine short baseline intracomplex ties to 1 cm, because the availability of the DSN stations changes from year to year. For example, DSS 14, the 64-meter antenna at Goldstone used for most of the data of Fig. 1, has been unavailable from 1983 July to 1984 August. In the interim, we are using DSS 12. Also, when DSS 14 becomes operational again, it may be displaced by a few centimeters, and its position relative to the other antennas will have to be redetermined. That short-baseline ties may be a problem is evident from the fact that the first two points of the **C-S** 1982 to 83 epoch were measured with DSS 13 and a DSS 13-DSS 14 tie. They, as well as the single **C-A** DSS 13 point in 1982 July, are systematically above the line.

The local vertical at Goldstone, which is poorly determined by short baseline experiments without accurate dry and wet troposphere calibration, has large projections on both the **C-A** and **C-S** baseline vectors, as noted in section II. Consequently, improved troposphere calibration might be necessary. We have therefore performed experiments, currently being analyzed, involving long and short baselines simultaneously, to check our procedure for determining the intracomplex ties. To improve the precision of the short-baseline ties, phase-delay results, which are now being analyzed, will supplement the bandwidth-synthesis result shown in Table 1. In addition, ground survey results will be investigated. In the absence of accurate short-baseline ties, the baseline length time behavior could still be parameterized with one slope and several intercepts, one for each Goldstone antenna. This least desirable alternative would weaken the determination of slope by introducing more parameters, but is always available as a last resort.

From the inner angle results in the most plausible case of Fig. 5(b), there is a statistically insignificant slope over the time span of the data. Studying the individual baseline component time behavior may explain the problem with the inconsistent point of 1983 January. If we can explain this point's behavior without resorting to the inadequacy of external high-frequency Earth rotation measurements, then there is hope that inner-angle studies may be useful. There is only a remote possibility of making regular, nearly-simultaneous measurements of the **C-A** and **C-S** baselines, in order to exclude any contribution from the short-term Earth orientation error. One solution to this dilemma is to measure a third baseline to monitor rapid changes in earth orientation. The time behavior of the inner angle would then complement the station movement results obtained from the length measurements. The possibility of using facilities in Japan to form triangles with both **C-A** and **C-S** is being explored.

Acknowledgment

The collection and analyses of the data presented in this article resulted from the efforts of many people over the years. For their design and implementation of the VLBI hardware, we would like to acknowledge D. Rogstad, L. Skjerve, D. Spitzmesser, and E. Cohen. G. Purcell and F. Bletzacker performed much of the analysis of the data taken before 1982. We would like to thank J. Gunckel for her diligence in the initial processing of the data. Finally, we thank the personnel at the DSS complexes in Goldstone, Madrid, and Canberra for taking the data and accurately documenting the conditions of the experiments.

References

1. Carter, W. E., Robertson, D. S., Pettey, J. E., Tapley, B. D., Schutz, B. E., Eanes, R. J., and Lufeng, M., "Variations in the Rotation of the Earth," *Science*, 224, pp. 957-961, June 1984.

2. Treuhaft, R. N., et. al., "The Time Variation of Intercontinental Baselines Using VLBI," *EOS 64*, p. 210, 1983.

3. Sovers, O. J., Thomas, J. B., Fanselow, J. L., Cohen, E. J., Purcell, Jr., G. H., Rogstad, D. H., Skjerve, L. J., and Spitzmesser, D. J., "Radio Interferometric Determination of Intercontinental Baselines and Earth Orientation Utilizing Deep Space Network Antennas: 1971 to 1980," *Journal of Geophysical Research*, 89, pp. 7597-7607, Sept. 10, 1984.

4. Bott, M. H., and Dean, D. S., "Stress Diffusion from Plate Boundaries," *Nature*, 243, pp. 339-341 (1973).

5. Allen, S. L., et. al., Current Mobile VLBI Data Base, submitted to the NASA Crustal Dynamics Information System, November 9, 1983.

6. Lanyi, G. E., "Tropospheric Delay Effects In Radio Interferometry." *TDA Progress Report*, 42-78, p. 152.

7. Treuhaft, R. N., et. al., "Empirical Troposphere Modeling From DSN Intercontinental VLBI Data," *EOS 65*, p. 191, 1984.

8. Thomas, J. B., "The Tone Generator and Phase Calibration in VLBI Measurements," *DSN Progress Report*, 42-44, pp. 63-74, 1978.

9. Rosen, R. D., et. al., "An El Niño Signal in Atmospheric Angular Momentum and Earth Rotation," *Science*, 225, p. 411, 1984.

10. Chao, C. C., "The Tropospheric Calibration Model for Mariner Mars 1971," Technical Report 32-1587, p. 75, Jet Propulsion Laboratory, Pasadena, Calif., 1974.

Table 1. DSS 14-DSS 43 (California–Australia)

Date yr/mo/day	x +2107200, m	Error	y −7323700, m	Error	z +7351800, m	Error	L −10588900, m	Error
78/9/3	−72.23	0.26	3.35	0.24	−1.56	0.27	66.58	0.35
78/10/28	−72.98	0.11	3.41	0.08	−1.14	0.12	66.48	0.12
78/11/4	−73.27	0.11	3.31	0.10	−0.99	0.14	66.37	0.15
78/12/31	−73.87	0.10	2.39	0.07	−1.70	0.12	66.34	0.12
79/11/23	−73.78	0.27	3.06	0.23	−1.14	0.24	66.39	0.31
79/12/20	−73.07	0.10	2.93	0.07	−1.44	0.11	66.37	0.11
79/12/29	−73.50	0.14	2.56	0.10	−1.30	0.13	66.10	0.14
80/1/12	−73.58	0.10	2.68	0.09	−1.60	0.13	66.41	0.13
80/1/27	−72.90	0.15	2.94	0.12	−1.79	0.14	66.59	0.16
80/2/14	−75.05	0.12	2.51	0.11	−1.60	0.14	66.59	0.16
80/2/23	−74.12	0.10	2.73	0.08	−1.50	0.12	66.48	0.12
82/7/1[a]	−71.31	0.09	3.76	0.06	−1.26	0.11	66.47	0.10
82/11/30	−72.79	0.14	3.30	0.10	−1.24	0.14	66.44	0.15
83/1/23	−71.84	0.10	3.66	0.06	−1.19	0.10	66.46	0.10
83/5/20	−73.69	0.08	3.29	0.04	−0.91	0.09	66.38	0.07

[a]Experiment was performed with DSS 13 referred to DSS 14 with the tie given below.

DSS 14-DSS 63 (California–Spain)

Date yr/mo/day	x −7202700, m	Error	y −4281100, m	Error	z −438000, m	Error	L −8390400, m	Error
78/10/30	13.73	0.22	60.24	0.32	57.05	0.32	29.79	0.20
78/11/6	13.65	0.18	60.62	0.19	56.89	0.25	29.91	0.19
79/11/26	13.00	0.08	61.81	0.10	56.57	0.11	29.94	0.06
79/12/21	13.46	0.16	60.71	0.29	57.12	0.25	29.80	0.13
79/12/27	13.02	0.09	61.66	0.15	56.83	0.17	29.89	0.09
80/1/26	13.32	0.16	61.22	0.17	57.06	0.22	29.93	0.14
80/2/14	12.30	0.07	62.86	0.08	56.58	0.11	29.87	0.05
80/2/24	12.98	0.08	61.60	0.14	56.46	0.15	29.80	0.08
82/7/4[b]	14.87	0.11	59.10	0.11	56.51	0.13	30.16	0.09
82/8/17[b]	14.42	0.10	59.98	0.11	56.40	0.13	30.22	0.08
82/11/28	13.50	0.05	61.06	0.06	56.96	0.08	30.01	0.03
83/1/25	14.85	0.05	59.05	0.06	56.89	0.08	30.14	0.04
83/4/8	13.81	0.05	60.83	0.06	56.66	0.08	30.14	0.04
83/5/22	13.27	0.04	61.55	0.06	56.99	0.08	30.06	0.03

[b]Experiment was performed with DSS 13 referred to DSS 14 with the tie given below.

DSS 14-DSS 13

Date yr/mo/day	x(m)	Error	y(m)	Error	z(m)	Error	L(m)	Error
79/8/15	2492.10	0.01	−14135.49	0.02	−16095.48	0.02	21565.88	0.01

Table 2. Baseline lengths per epoch.

Epoch (year)	C-A (m)	C-S (m)
78.84	66.40 ± 0.07	29.85 ± 0.14
80.03	66.41 ± 0.05	29.88 ± 0.03
82.97	66.42 ± 0.05	30.09 ± 0.02

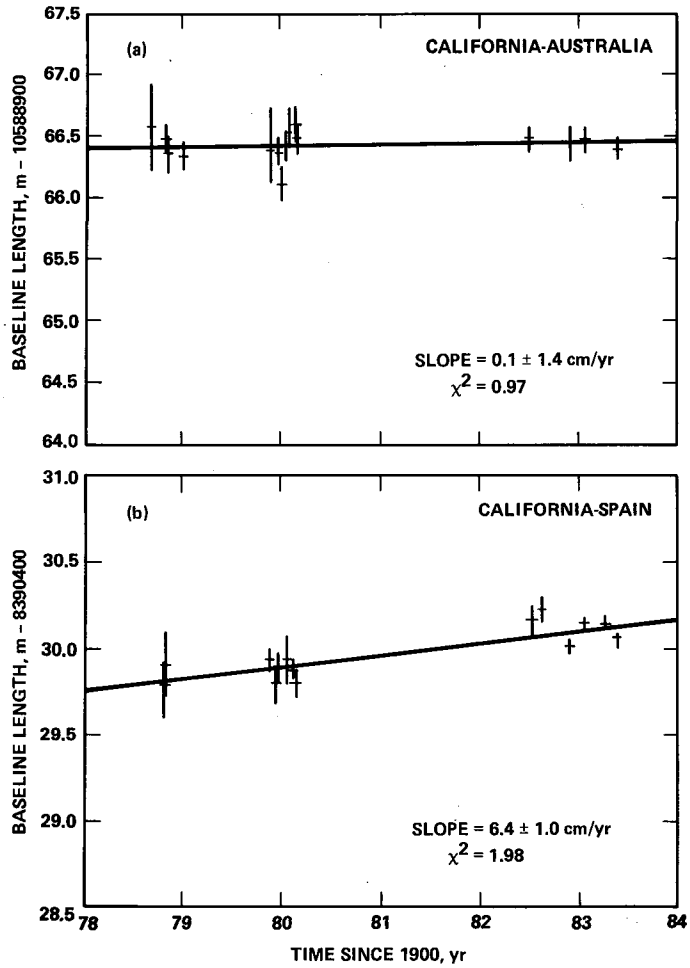


Fig. 1. The baseline length as a function of time: (a) California–Australia; (b) California–Spain

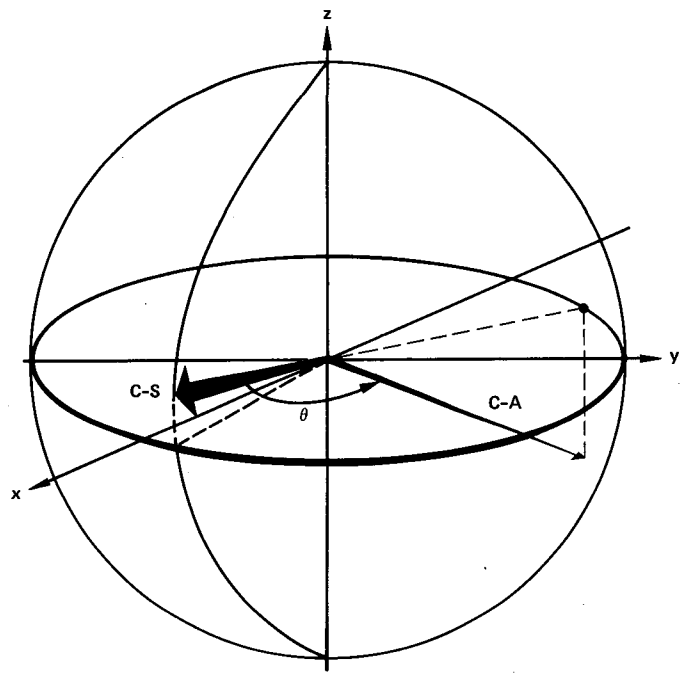


Fig. 2. The California–Australia and California–Spain vectors shown in the standard, Earth-fixed frame. θ is the angle between the two vectors, called the "inner angle" in the text

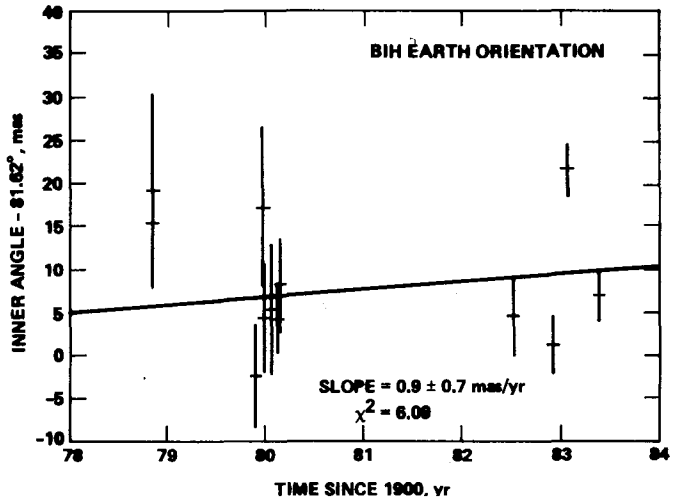


Fig. 3. The inner angle as a function of time using BIH Earth orientation values

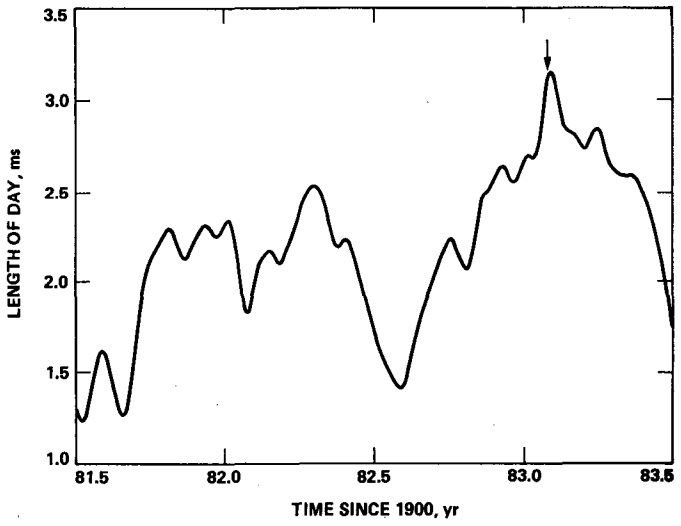


Fig. 4. The length of day, which is $d(UT1-IAT)/dt$ in milliseconds per day, versus time. The arrow points to the time of the second to last experiment pair, corresponding to the second to last inner angle of Fig. 3.

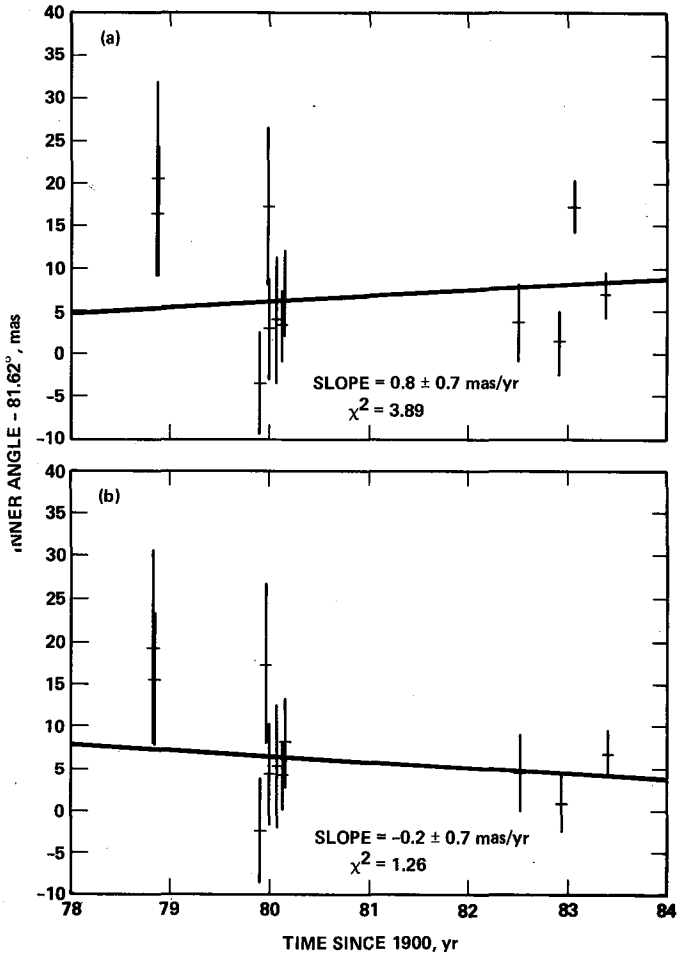


Fig. 5. The inner angle as a function of time using: (a) a combination of Kalman filtered POLARIS and lunar laser ranging Earth orientation values; (b) BIH Earth orientation values, excluding the second to last point

The Performance of Differential VLBI Delay During Interplanetary Cruise

B. Moultrie, P.J. Wolff, and T.H. Taylor
Navigation Systems Section

Project Voyager radio metric data are used to evaluate the orbit determination abilities of several data strategies during spacecraft interplanetary cruise. Benchmark performance is established with an operational data strategy of conventional coherent Doppler, coherent range, and explicitly differenced range data from two intercontinental baselines to ameliorate the low declination singularity of the Doppler data. Employing a Voyager operations trajectory as a reference, the performance of the operational data strategy is compared to the performances of data strategies using differential VLBI delay data (spacecraft delay minus quasar delay) in combinations with the aforementioned conventional data types. The comparison of strategy performances indicates that high-accuracy cruise orbit determination can be achieved with a data strategy employing differential VLBI delay data, where the quantity of coherent radio metric data has been reduced by over 95% with a concurrent 90% reduction in the DSN time allocated to radio metric data acquisition.

I. Introduction

Interplanetary missions of the future will impose increasingly stringent requirements on navigational systems for precision, accuracy, and overall economy. Chief amongst the radio metric systems developed to meet these demands is NASA's Deep Space Network (DSN) operational Very Long Baseline Interferometry (VLBI) system. A primary product of this system is the differential VLBI (DVLBI) delay data type, which is used in spacecraft orbit determination.

DVLBI delay offers performance that is currently unexcelled by any other radio metric data type in the determination of spacecraft angular position. (Data precision has been demonstrated at 70 nanoradians – approximately 10.5 km/AU, where 1.0 AU $\approx 149.6 \times 10^6$km.) An (implicit) accuracy of 50 nanoradians is anticipated by 1986 in support of Project Galileo. An extensive discussion of the data type and the associated VLBI system accuracies and attributes is provided by Border et. al. (Ref. 1).

The applicability of DVLBI delay to differental navigational situations has been a topic of great interest (Refs. 1 through 9). This article serves to solidify knowledge in this area by validating the significant contributions that DVLBI delay can make to spacecraft orbit determination during interplanetary cruise. Results are given that demonstrate that the addition of DVLBI delay to a cruise data strategy simplifies and economizes the determination of spacecraft state with retained or enhanced precision and accuracy.

The main text of this article begins with a discussion of the salient facts concerning the Voyager 2 Jupiter-to-Saturn cruise geometry. This discussion is followed by a description of the radio metric data and spacecraft tracking strategies used in the study. Then the study's assumptions and data filter structure are detailed. Thereafter, an assessment is made of the abilities of the competing data strategies to determine spacecraft state. This assessment is followed by a summary and presentation of conclusions.

II. Geometry

A segment of Voyager 2's 1981 Jupiter-to-Saturn cruise is employed for the study. Radio metric data coverage is from March 26th, the trajectory's epoch, to June 13th.

Voyager 2 maintained an absolute declination (DEC) of less than 1.4 degrees during the time of data coverage with a right ascension (RA) of approximately 183 degrees. A low DEC environment greatly reduces the spacecraft DEC sensitivity of coherent Doppler data, which are the traditional mainstay of radio metric tracking.

The sun-earth-spacecraft angle was greater than 100 degrees during the time of data coverage. A large angular separation between the sun and the spacecraft minimizes the dispersive effect that space plasma has on radio metric data.

III. Data Arc and Data Strategies

The investigated tracking data strategies are developed from eleven weeks of Voyager Navigation S-band radio metric data consisting of three coherent data types (Doppler, range, and differenced range) and DVLBI delay data acquired in a ground-station receive-only spacecraft tracking mode. Data from Deep Space Stations where the spacecraft's elevation angles are less than fifteen degrees are not used in the study. All radio metric data are calibrated with operationally employed seasonal models to account for radio signal delay due to the troposphere, and models developed from Faraday rotation data to account for signal delay due to the ionosphere.

Four tracks of Doppler data, each containing one range point, are included in each week of data. These data directly measure the spacecraft's line-of-sight velocity and distance, and, over a period of weeks, the combination of Doppler and range data determine spacecraft angular velocities. Hamilton and Melbourne (Ref. 10) have shown that each track of Doppler indirectly determines the spacecraft's RA and DEC. However, the accuracy of the DEC determination deteriorates rapidly as the spacecraft's DEC approaches zero.

Differenced range data measure spacecraft angular position and provide a means for amelioration of the low DEC singularity of Doppler data (Refs. 2, 11, and 12). This differenced data type is obtained by explicitly differencing two range points from two Deep Space Stations that define a "baseline." The Deep Space Station combination of Goldstone-Madrid (G-M) defines an east-west baseline; Goldstone-Canberra (G-C) defines a north-south baseline. (Data from the G-C baseline measure spacecraft DEC, while data from the G-M baseline measure spacecraft RA.) This differencing of range points reduces sensitivity to errors that are common to baseline stations, e.g., unmodeled spacecraft accelerations. Measurement accuracy is limited by errors that are not common to both stations such as station location errors, transmission media errors, and station instrumental effects.

The study data arc includes twenty G-M differenced range measurements and sixteen G-C differenced range measurements. Due to geometry constraints and coherency requirements, range points differenced in this study are offset in time from eleven minutes to over three hours. The resulting differenced range points are clustered at the beginning, middle, and end of the data arc.

Navigational DVLBI delay is a data type developed by differencing the VLBI delay measurements from a spacecraft and an angularly nearby natural radio source (typically a quasar), where each delay is determined as differential range obtained from simultaneous observation of each radio source by two widely separated Deep Space Stations. The result is a data type that measures the angular offset of the spacecraft from a known position in the sky (the quasar's angular position), whereas differenced range measures the spacecraft's total angular position in the sky. During the time frame of the study, the angular separation between Voyager 2 and the reference quasar (3C 273) varied between zero and three degrees. For the study, DVLBI delay data are included at the rate of one point from each of the G-M and G-C baselines per week.

For DVLBI delay data, the simultaneous observation of a radio source by two stations and the subsequent differencing over a baseline as well as the differencing between sources that are angularly close results in a data type that is highly self-calibrating. (The double differencing greatly reduces the effects of station common errors as well as the effects of errors that are common to each delay.) Inclusion of the quasar delay in the differencing process introduces an error in the DVLBI delay due to imperfect knowledge of the quasar's location (as determined from radio interferometry) with respect to the optically determined planetary ephemerides. This error is referred to as a "frame tie" error.

In a previous analysis of these data (Ref. 2), biases between differenced range and DVLBI delay were observed (-8 m on the G-C baseline and -4 m on the G-M baseline). Substantial contributions to the magnitudes of these biases can be credited to a frame tie error. In the current study, G-C and G-M biases have been reduced to -1.19 m and -0.85 m, respectively. The reductions in the biases are achieved by the determination of quasar positions with respect to Voyager models and values for Deep Space Station locations, earth orientation parameters, precession, and nutation (Border, J.S., and Sovers, O.J., "Radio Source Position Catalog for Delta DOR," Tracking Systems and Applications Section internal document, October 6, 1982).

The frame tie error is expected to be limited to 100 nanoradians in the near future. For the purpose of future applicability, a 1-sigma (one standard deviation) uncertainty or 100 nanoradians is assumed for each component of quasar position in this study.

The data strategies examined are outlined in Table 1. The nominal operational data mix is reflected in strategy A. This strategy involves least squares fitting to all available tracks of Doppler data where these data have been augmented by occasional range points, and differenced range data (particularly from the G-C baseline for DEC sensitivity) under low DEC conditions. This approach typically results in a data set that contains hundreds if not thousands of points. In strategy B, all of the available Doppler, range, and DVLBI delay data are combined. The total number of points in this strategy is comparable to the total number in strategy A. In strategy C, the DVLBI data are combined with one Doppler point and one range point per week. Strategy D combines one Doppler point and one range point per week with one DVLBI delay point every two weeks from each baseline. Strategy E combines the delay data with one range point per week.

Examination of Fig. 1 reveals that strategies C, D, and E, when compared to strategy A, achieved reductions in data volume greater than 95% and reductions in station tracking time that approach 90%.

IV. Filter Structure and Assumptions

Covariances for each data strategy are generated and propagated by a (least squares) Square Root Information Consider Filter (Ref. 13). These covariances are derived from the estimation of the spacecraft's state, impulsive maneuvers, and nongravitational accelerations due to gas leaks in the spacecraft's attitude control system. These accelerations are treated as biases and as white process noise in a four-day batch structure. The effects of errors in quasar RA and DEC are considered as are the effects of errors in station longitudes, distances off the earth spin axis, and distances above the equatorial plane. Nominal a priori sigmas and assumed correlations between station coordinates are listed in Table 2.

For each data strategy, 1-sigma a priori uncertainties for the data types are 1 mm/s for Doppler (for a 60-s count time), 10 km for range, 5 m for differenced range, and 2 m for DVLBI delay. A relatively large uncertainty is assumed for range data to allow for the range's noisier quality and vulnerability to unknown accelerations in the spacecraft's line-of-sight direction with respect to DSN stations (Ref. 14).

V. Performance of Data Strategies

The nominal spacecraft trajectory employed in this study is developed from Voyager force and ephemeris models, and Doppler, range, and differenced range data that were operationally acquired prior to the start of the study data arc. Each data strategy is used to correct the initial spacecraft state and to derive associated error covariances. These covariances and corrected states are examined in a plane-of-sky frame at the end of the data arc.

The magnitudes of postfit data residuals provide indications of how effective data strategies are in eliminating systematic signatures in the data. Figure 2 displays the reductions in root-mean-square residual values realized for each data type in the respective strategies. It can be observed that strategies B, C, and D, which involve DVLBI delay data, achieved reductions in data scatters varying from 40% for delay and differenced range data to 96% for range data. These reductions in scatters are very close to the reductions obtained with strategy A, which utilized only the coherent data types. Reductions in scatters obtained with strategy E, which relies on infrequent range points and DVLBI delay data, are less dramatic by only a factor or two. Thus, the DVLBI strategies are nearly as effective as the conventional strategy in removing signatures from the data.

As summarized in Table 3, the DVLBI strategies provide estimates of state that are in close agreement with the conventional state estimate at the end of the data arc. In particular, RA and DEC estimates differ by maximums of 36 km and 125 km (27 nanoradians and 92 nanoradians), respectively, at 9.035 AU.

Figures 3 and 4 show the geocentric plane-of-sky uncertainties at the end of the data arc. The effects that the considered parameters have on state uncertainty are shown in Figs. 5 through 10.

For conventional tracking, station longitudes and spin axis distances are the dominant error sources. It can be seen from

Figs. 5 and 6 that, for strategy A, 1-sigma errors in these station parameters contribute uncertainties of nearly 600 km to RA and 200 km to DEC compared to total uncertainties of roughly 650 km and 420 km, respectively (Figs. 3 and 4).

In the DVLBI strategies, total uncertainties in RA and DEC are under 230 km and 375 km, respectively. Inspections of Figs. 5 and 6 reveal that the DVLBI strategies reduce the effects of errors in longitudes and spin axis distances by more than an order of magnitude. In particular, for strategies C, D, and E, the components of quasar position are the dominant error sources in the determination of spacecraft RA and DEC. A 1-sigma error in quasar RA contributes 125 km to uncertainty in spacecraft RA. Similarly, a 1-sigma error in quasar DEC contributes 125 km to uncertainty in spacecraft DEC. Note that Fig. 4 shows that the DVLBI strategies also reduce the uncertainties in the rates for RA and DEC by 10% to 15%.

In the DVLBI strategies, the effects of errors in the considered parameters are subkilometer for radial distance (Fig. 7), and less than 0.7 mm/s for radial velocity (Fig. 10). The uncertainties in these components of state increase in strategies C and E, for which the quantities of Doppler and range data are reduced. The increase in these uncertainties is to be expected since Doppler and range data are strongest in determining the radial components of state.

Comparable uncertainties in the radial components are obtained with the strategies that include Doppler data. Uncertainties in the position component are approximately 1.7 km for strategies A and B, and 3.2 km for strategies C and D. The position uncertainty increases to 13 km in strategy E, which does not include Doppler. Radial velocity uncertainty is 4.5 mm/s for strategy A, and less than 2 mm/s for strategies B, C, and D. The lower uncertainty in radial velocity for DVLBI strategies B, C, and D results primarily from the decoupling of radial velocity from station longitudes. In strategy E, velocity uncertainty increases to 9 mm/s. This increase in uncertainty is due to the lack of a direct measure of radial velocity and the high correlation between radial velocity and a loosely determined radial component of position.

The performances of the above DVLBI strategies demonstrate that improved orbit determination accuracy can be achieved, with a concurrent reduction in the requirement for navigation and tracking system resources, if DVLBI delay data are added to a cruise tracking strategy.

VI. Summary and Conclusions

Results have been presented that are believed to be realistic indications of the cruise orbit determination enhancements achievable when DVLBI data are included in radio metric data strategies. These enhancements are reflected in the ability of DVLBI strategies to satisfy the increasing demands placed on navigation systems to develop methods for the determination of spacecraft state with greater economy without compromising state accuracy requirements.

DVLBI strategies were shown to provide estimates of spacecraft state that closely agree with the conventional state estimate and are generally less uncertain than the conventional state estimate. The most accurate estimates of state were obtained with the DVLBI strategies, which included infrequent coherent Doppler and range data. The high level of accuracy achieved with the DVLBI strategies is partially due to the relative immunity of DVLBI strategies to conditions and error sources that are significant accuracy constraints in conventional tracking. In DVLBI strategies, sensitivity to spacecraft declination is retained in low declination environments and quasar RA and DEC replace station-location parameters as dominant error sources, although at substantially lower levels. As the uncertainty in the tie between the planetary ephemeris frame and quasar reference frame decreases below 100 nanoradians (a level expected to be a achieved in the near future), the effects of errors in quasar RA and DEC will be diminished.

DVLBI strategies were shown to economize orbit determination. The maximum reduction in data volume exceeded 95% of the conventional benchmark, while the maximum reduction in station tracking time approached 90%. These savings imply reduced requirements for data management and processing resources.

The above contributions of high-level accuracy and economy cast DVLBI strategies as compelling alternatives to the conventional coherent data strategy for cruise orbit determination.

Acknowledgments

The authors wish to express their thanks to Voyager Navigation for its valuable assistance in the acquisition of spacecraft trajectory models and radio metric navigation data, and in helping to establish an orbit determination procedure. Also, thanks are offered to C.S. Christensen and J. Ellis of the Navigation Systems Section for their rendering of several helpful suggestions.

References

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

2. Taylor, T.H., Campbell, J.K., Jacobson, R.A., Moultrie, B., Nichols, R.A. Jr., and Riedel, J.E., "Performance of Differenced Range Data Types in Voyager Navigation," *TDA Progress Report 42-71*, Jet Propulsion Laboratory, Pasadena, Calif., November 1982, pp. 40-52.

3. Melbourne, W.G., and Curkendall, D.W., "Radio Metric Direction Finding: A New Approach to Deep Space Navigation," presented at AAS/AIAA Astrodynamics Specialist Conference, Jackson Hole, Wyoming, September 1977.

4. Brunn, D.L., Preston, R.A., Wu, S.C., Siegel, H.L., Brown, D.S., Christensen, C.S., and Hilt, D.E., "DVLBI Spacecraft Tracking System Demonstration: Part I. Design and Planning," *DSN Progress Report 42-45*, Jet Propulsion Laboratory, Pasadena, Calif., March-April 1978, pp. 111-132.

5. Christensen, C.S., Callahan, P.S., Moultrie, B., Donivan, F.F., and Wu, S.C., "Results of a Demonstration of the Use of DVLBI Data for Precise Navigation of Interplanetary Spacecraft," paper 80-1648, presented at AIAA/AAS Astrodynamics Conference, Danvers, Mass., August 1980.

6. Brown, D.S., Hildebrand, C.E., and Skjerve, L.J., "Wideband DVLBI for Deep Space Navigation," Proceedings of the IEEE Position Location and Navigation Symposium, Atlantic City, N.J., December 1980, pp. 389-396.

7. Wood, L.J., and Jordan, J.F., "Interplanetary Navigation in the 1980's and 1990's," paper 81-113, presented at AAS/AIAA Astrodynamics Specialist Conference, Lake Tahoe, Nev., August 1981.

8. Ellis, J., "Deep Space Navigation with Noncoherent Tracking Data," *TDA Progress Report 42-74*. Jet Propulsion Laboratory, Pasadena, Calif., April-June 1983, pp. 1-12.

9. Mohan, S.N., and Ananda, M.P., "Simultaneous Dual Baseline Differential VLBI," *The Journal of the Astronautical Sciences*, Vol. XXXI, No. 1, pp. 117-134, Jan-Mar. 1983.

10. Hamilton, T.W., and Melbourne, W.G., "Information Content of a Single Pass of Doppler Data from a Distant Spacecraft," *Space Program Summary No. 37-39*, Vol. III. Jet Propulsion Laboratory, Pasadena, Calif., May 1966, pp. 18-23.

11. Rourke, K.H., and Ondrasik, V.J., *Application of Differenced Tracking Data Types to the Zero Declination and Process Noise Problems*, Technical Report 32-1526, Vol. IV. Jet Propulsion Laboratory, Pasadena, Calif., August 1971, pp. 49-60.

12. Siegel, H.L., Christensen, C.S., Green, D.W., and Winn, F.G., "On Achieving Sufficient Dual Station Range Accuracy for Deep Space Navigation at Zero Declination," presented at AAS/AIAA Astrodynamics Specialist Conference, Jackson Hole, Wyoming, September 7-9, 1977.

13. Moyer, T.D., *Mathematical Formulation of the Double-Precision Orbit Determination Program (DPODP)*, Technical Report 32-1527. Jet Propulsion Laboratory, Pasadena, Calif., May 15, 1971.

14. Ondrasik, V.J., and Rourke, K.H., *An Analytical Study of the Advantages Which Differenced Tracking Data May Offer for Ameliorating the Effects of Unknown Spacecraft Accelerations*, Technical Report 32-1526, Vol. IV. Jet Propulsion Laboratory, Pasadena, Calif., August 1971, pp. 61-70.

Table 1. Data strategies

Data Type	A (Conventional)	B	C	D	E
Doppler	Four 5-h passes per week	Four 5-h passes per week	One observation per week	One observation per week	-
Range	Four observations per week	Four observations per week	One observation per week	One observation per week	One observation per week
Differenced Range	20 E-W observations 16 N-S observations	–	–	–	–
DVLBI	–	One E-W observation per week One N-S observation per week	One E-W observation per week One N-S observation per week	1/2 E-W observation per week 1/2 N-S observation per week	One E-W observation per week One N-S observation per week

Table 2. 1-sigma a priori uncertainties for estimated and considered parameters

Parameters	Uncertainties	Intercontinental station-pair correlations
Spacecraft position	10^6 km[a]	–
Spacecraft velocity	1 km/s[a]	–
Maneuvers	10^{-5} km/s[a]	–
Nongravitational accelerations	5×10^{-12} km/s^2[a]	–
Station spin axis distance	1.5 m	0.778
Station longitude	3×10^{-5} deg	0.955
Station equatorial height	15 m	0.998
Quasar position	100 nrad[a]	

[a]Each component.

Table 3. Maximum difference between the conventional estimate of state and the DVLBI estimates of state and the DVLBI estimates

Parameter	Position components, km	Velocity components, mm/s
RA	36	39
DEC	125	25
Radial	1	1

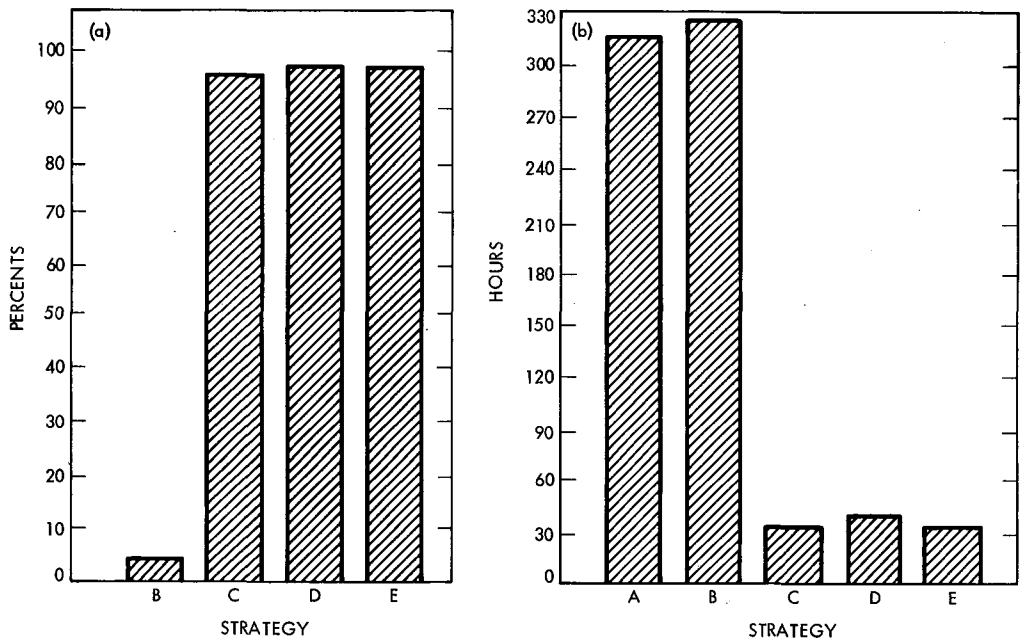


Fig. 1. Resource requirements: (a) reductions in the number of data points with respect to the strategy A; (b) station nominal tracking time requirements

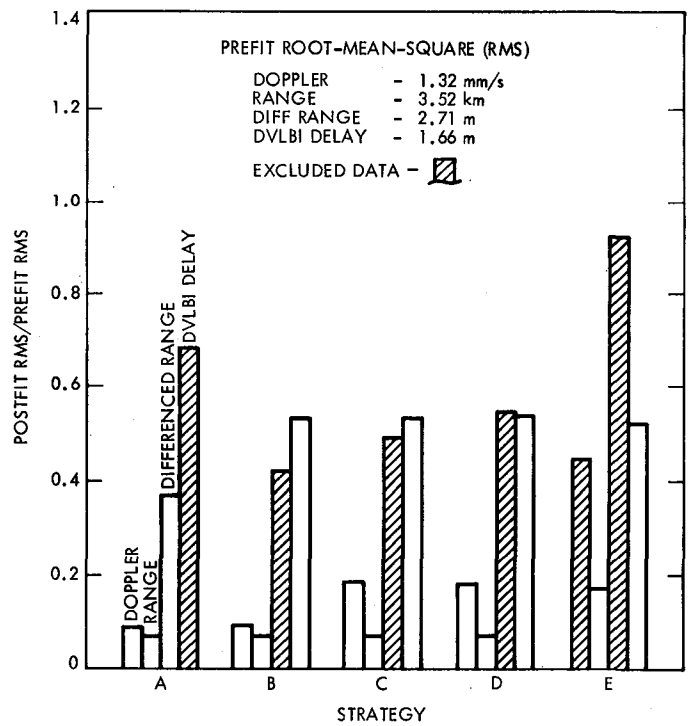


Fig. 2. Reductions in data scatters achieved with each data strategy

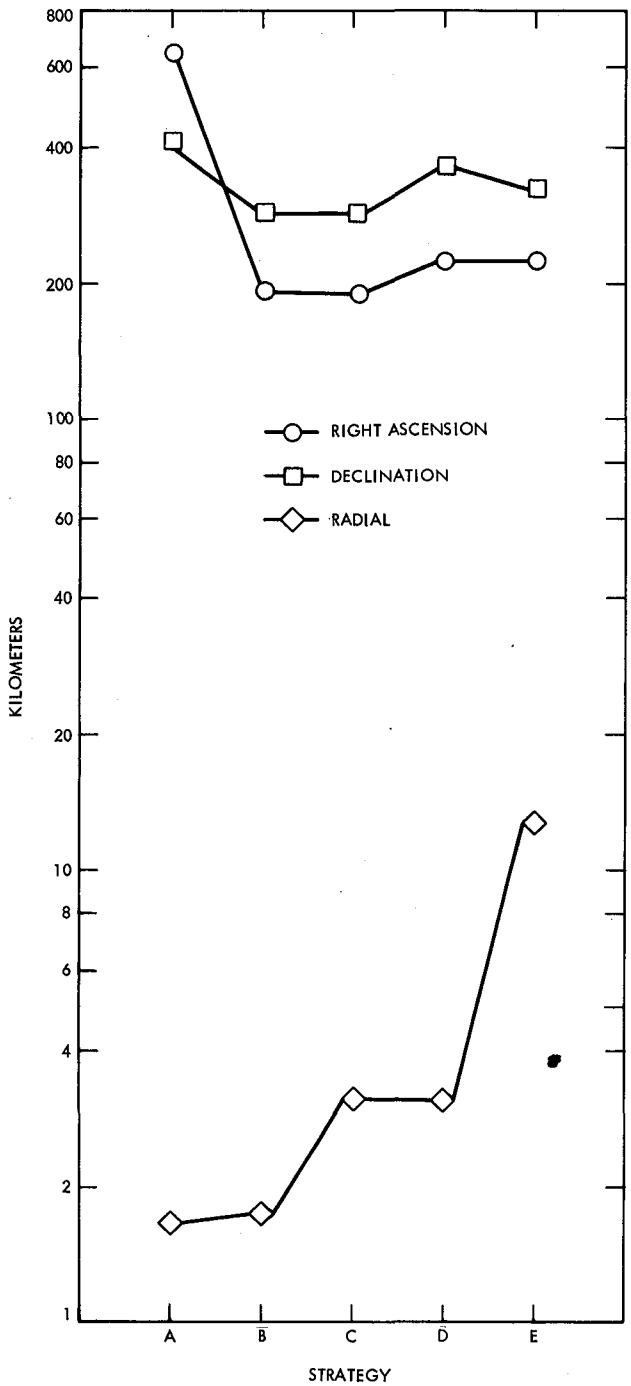


Fig. 3. 1-sigma spacecraft geocentric position component uncertainties at 9.035 AU

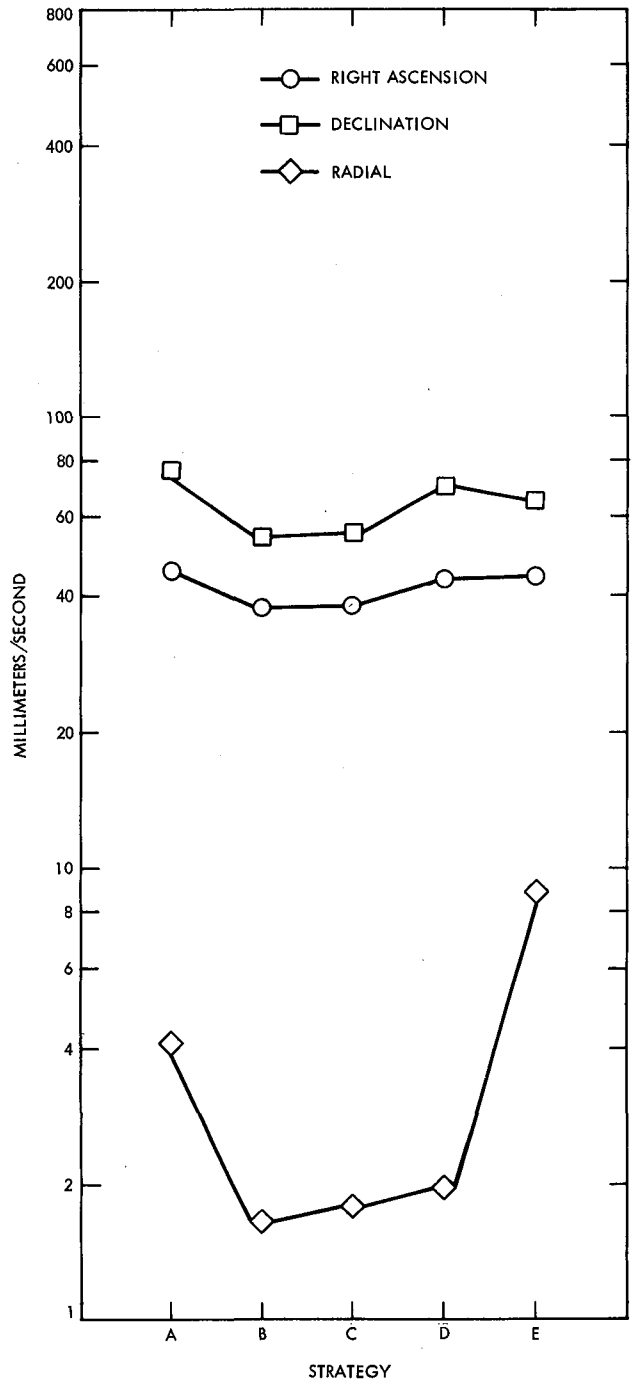


Fig. 4. 1-sigma spacecraft geocentric velocity component uncertainties at 9.035 AU

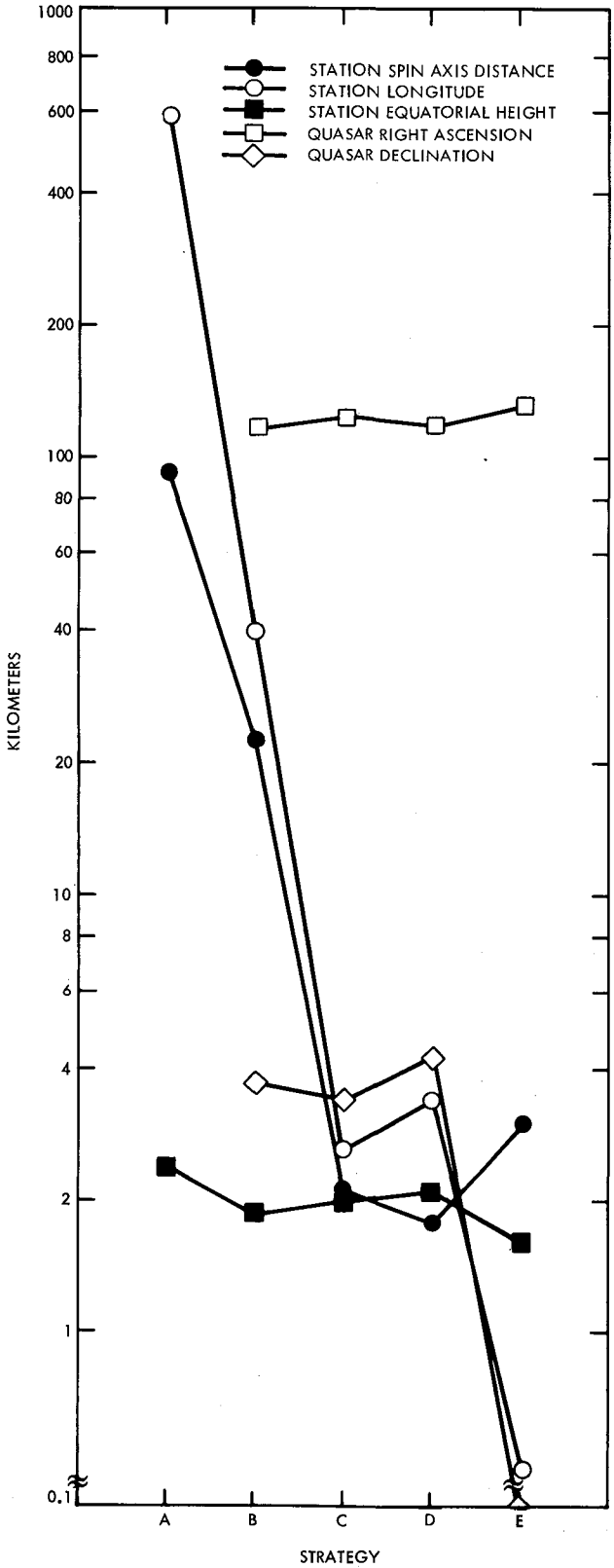


Fig. 5. Contributions of considered parameters to uncertainty in spacecraft geocentric right ascension at 9.035 AU

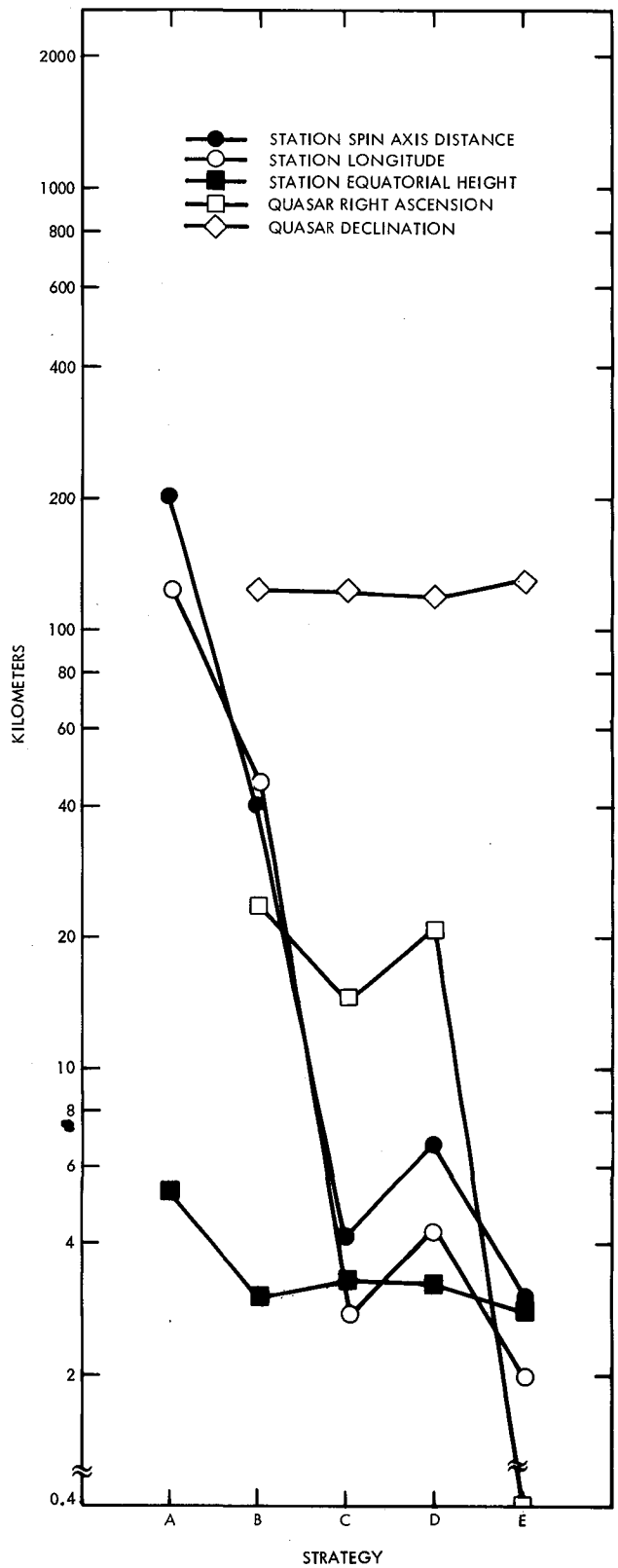


Fig. 6. Contributions of considered parameters to uncertainty in spacecraft geocentric declination at 9.035 AU

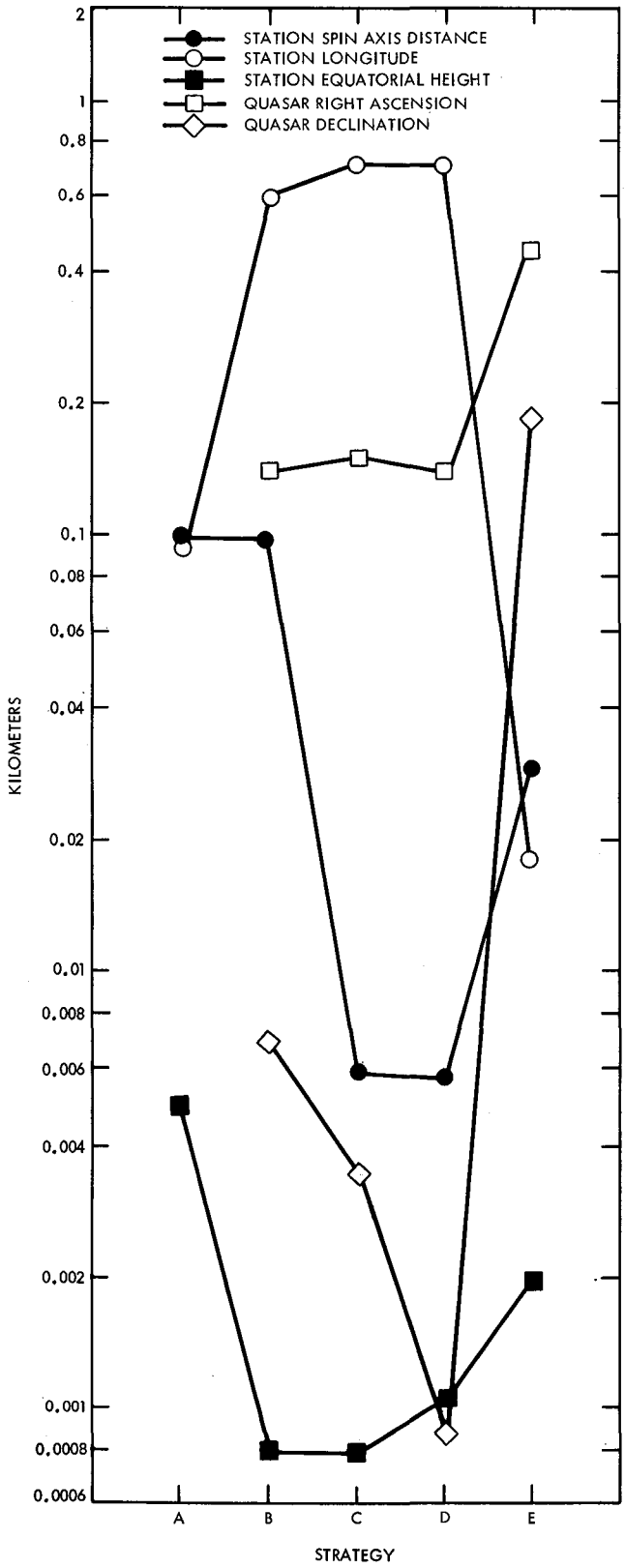


Fig. 7. Contributions of considered parameters to uncertainty in spacecraft geocentric radial distance at 9.035 AU

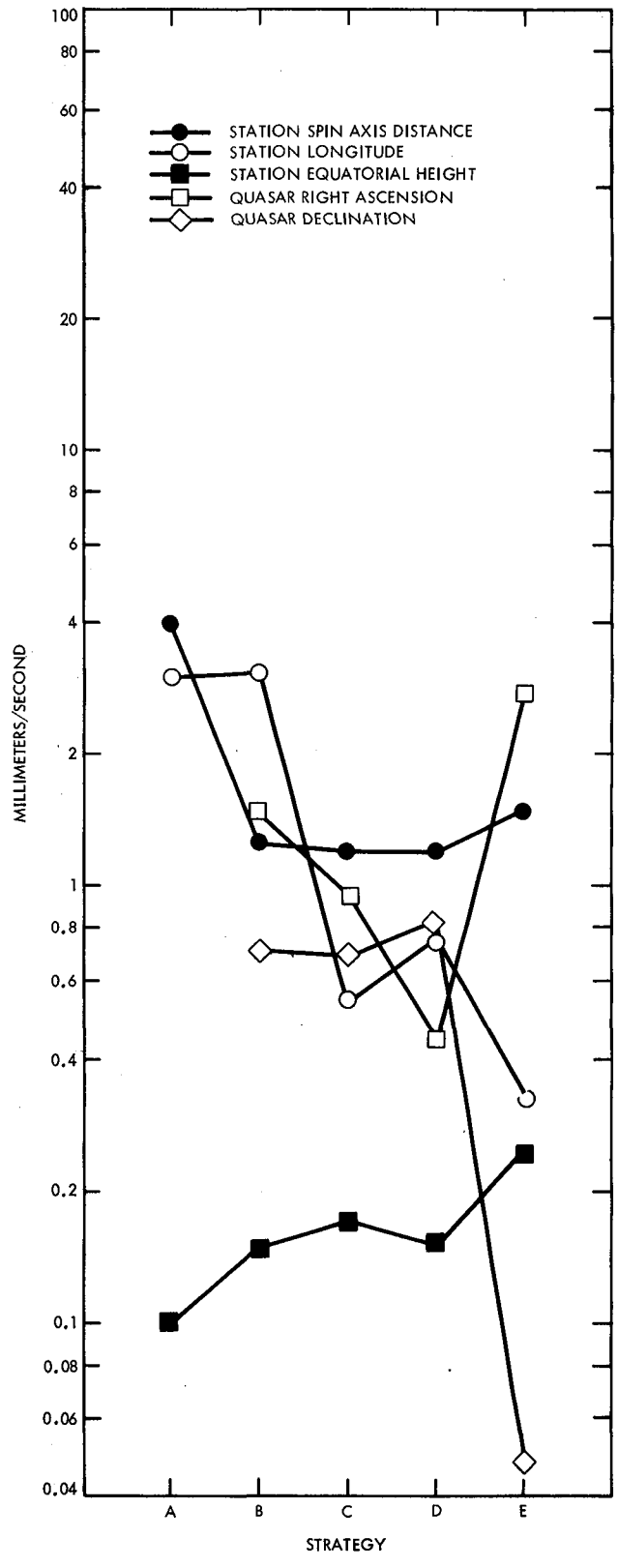


Fig. 8. Contributions of considered parameters to uncertainty in spacecraft geocentric right ascension rate at 9.035 AU

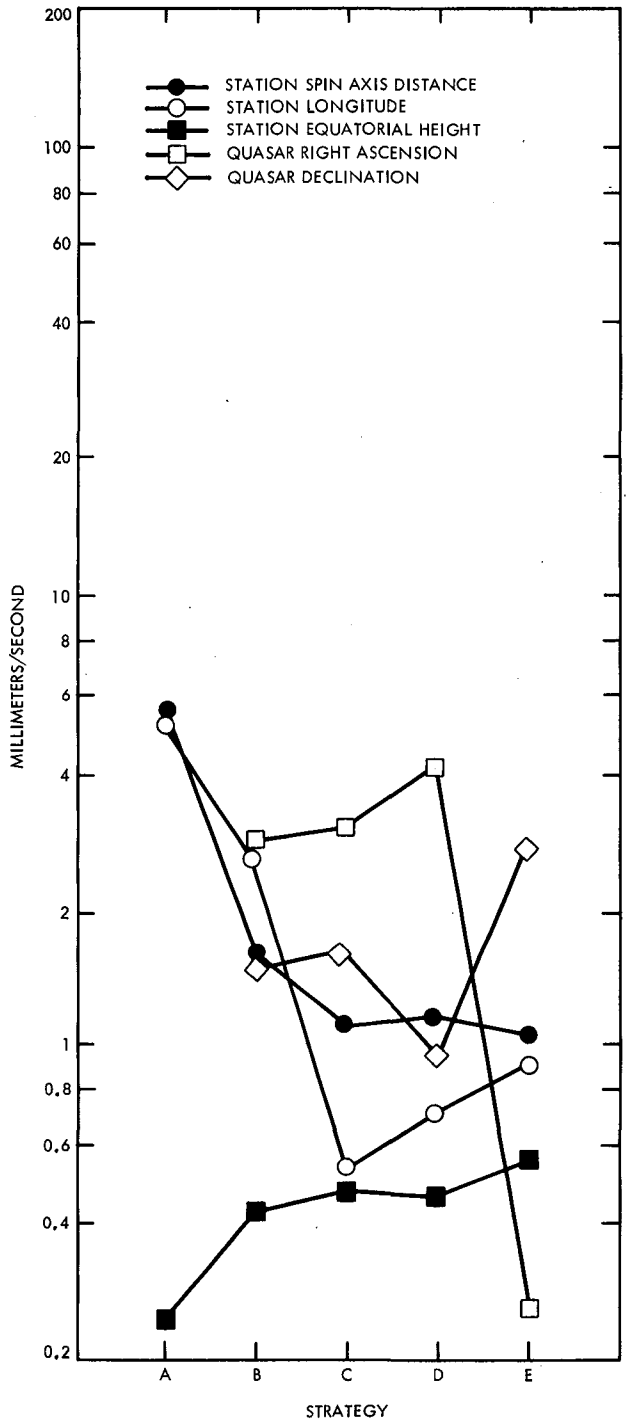


Fig. 9. Contributions of considered parameters to uncertainty in spacecraft geocentric declination rate at 9.035 AU

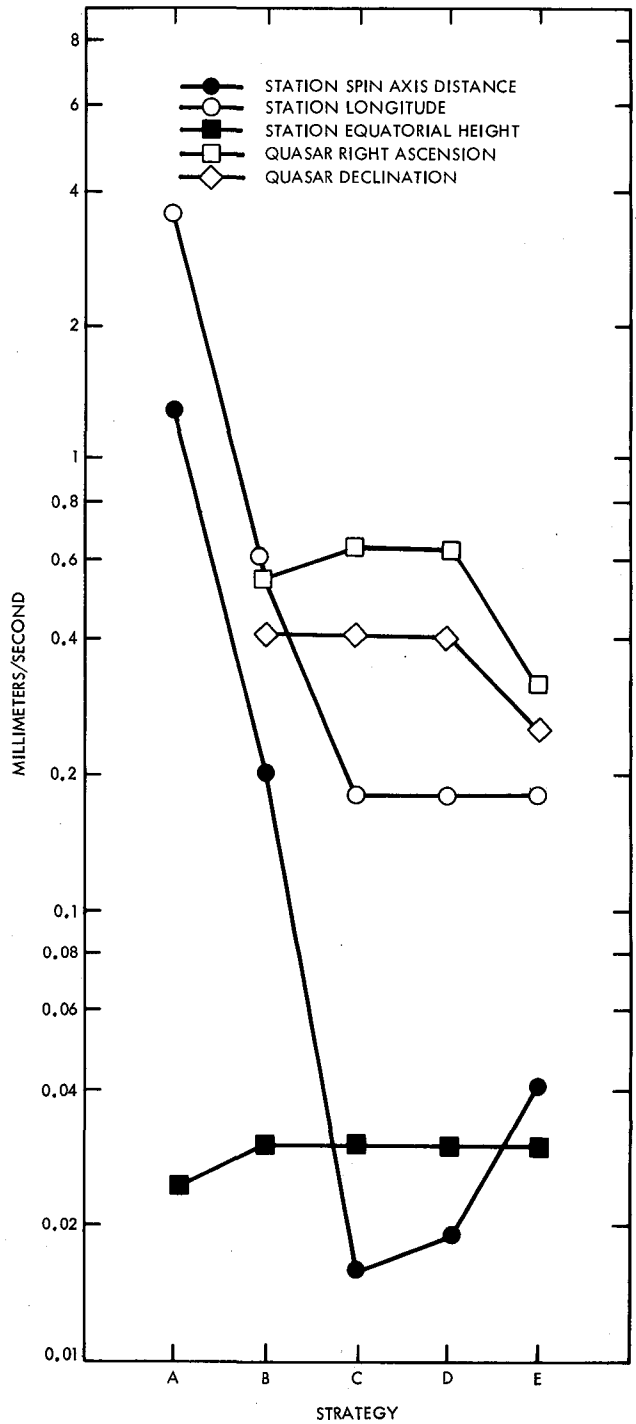


Fig. 10. Contributions of considered parameters to uncertainty in spacecraft geocentric radial velocity

Development and Testing of a 20-kW X-Band Transmitter With High Phase Stability

R. Cormier and T. Tesarek
Radio Frequency and Microwave Subsystems Section

A new 20-kW X-band transmitter has been built and installed on the antenna at DSS 13. A description of the hardware and operation are presented. This transmitter is the prototype for the new X-band uplink to be added to the DSN to support Galileo and future missions.

I. Introduction

The design for an X-band uplink transmitter began in June 1979 to provide uplink capability in the frequency range of 7145 to 7235 MHz. By June 1982, a model of this transmitter was installed at DSS 13. Since then it has been operated intermittently for the purpose of evaluation before incorporating subsequent units into the Deep Space Network.

Some of the significant results obtained were the testing of the S-X feedcone assembly, distributed control, and phase stability. The S-X Feed test results have been reported (Ref. 1).

The klystron tube selected is a Varian Associates Model VA-876. Extensive testing and evaluation were made and the results reported earlier (Ref. 2).

The Exciter and Control systems have been reported on by R. Hartop, et al. (Ref. 3). The Control system has been changed since and the modifications will be covered in this report.

This report contains a description of the transmitter as configured at DSS 13. Operational results, modifications, and possible changes or alternatives for the transmitters to be placed into the Deep Space Network will be presented, as well as early results of its phase stability.

II. Hardware

The transmitter hardware at DSS 13 is distributed with parts or assemblies located on the ground, and other parts located in and around the antenna deckhouse and feedcone assembly. The motor-generator set and its controls, the beam power supply, and the controller are located at ground level, while the klystron assembly along with its input and output RF assemblies, the associated low voltage supplies, and some signal processing equipment are located in the feedcone. The heat exchanger is located just outside the deckhouse on the antenna. Fig. 1 is a block diagram of this transmitter. A brief description of the most important features of each of the major assemblies along with possible alternatives is presented. The operating parameters are summarized in Table 1.

A. Power Amplifier Assembly

The power amplifier (PA) enclosure is a compact unit measuring 61 cm wide, 53 cm deep, and 97 cm in height (Fig. 2) with all the joints and removable panels made to have metal contact for best RF shielding integrity. Three of the enclosure's sides and the top are removable for ease of maintenance and klystron removal. The fourth side serves as a mounting surface for the output waveguide structure and buffer amplifier assembly (Fig. 3). Klystron coolant connections are located on the top panel and consist of AN-type flare bulkhead fittings.

The enclosure is divided into three compartments. The upper compartment contains the Varian VA-876 klystron, the focus magnet, and the RF power monitoring circuits. One of the lower compartments contains the filament transformer, cathode cooling fan, and high voltage connection. The other lower compartment contains filament control circuits and interfacing terminal boards. The filament control circuit includes a "step start" sequence. During the first 5 minutes of filament "warm-up," a series resistor is inserted into the filament primary circuit to limit inrush current. When the initial 5 minute time delay is complete, the series resistor is switched out and a second 5 minute time delay begins to complete filament warm-up. Final filament voltage is set by an adjustable autotransformer in the primary circuit.

The output waveguide structure (Fig. 4) consists of a harmonic filter, arc detector, directional coupler, waveguide switch, and water load. The entire assembly is water cooled and was fabricated as one unit. This has the advantage that all waveguide flanges have been eliminated except for the input and output. It has the disadvantage that it must be completely replaced if a problem develops with any part of this assembly.

The buffer amplifier, mounted on the back panel of the PA cabinet, receives a low level drive signal (7145 to 7235 MHz) which is amplified by a two-watt solid-state intermediate amplifier. Drive on/off, drive level, and RF-path switching commands are received from the buffer control chassis located in the instrumentation rack. Drive level and switch position indications are also sent to the buffer control chassis. Another feature of this unit is a local oscillator and RF test port for independent transmitter testing, or for use when the receiver/exciter drive signal is not available.

B. Instrumentation Cabinet

The four-foot instrumentation rack, mounted next to the PA enclosure in the feedcone, contains a control/monitor interface chassis (Fig. 5), buffer control chassis (Fig. 6), and focus magnet and system DC power supplies (Fig. 7). The front panel has indicators and controls associated with the operation of the power amplifier only. In the original design an Intel ICS-80 microprocessor assembly did the signal conditioning and control functions for the power amplifier and heat exchanger. The distributed control approach was abandoned in October 1982 for the use of a central processor control located at ground level. A multiconductor cable was added for communication between the antenna and the ground equipment.

The control/interface chassis provides signal conditioning, interlock summing logic, and command control interface. Also included are optical isolator PC boards to isolate antenna mounted equipment from ground-based units when multiconductor interconnect cables are used. Signal conditioning is required for turbine flow meter and analog indications since the microprocessor accepts only 0 to 5 VDC full scale analog signals.

All protective interlocks associated with antenna-mounted equipment are logically summed at this point and a single "PA ready" indication is sent to the high voltage power supply via a "hard wire" interlock (I/L) cable. At the high voltage power supply, the PA ready is summed with ground-based interlocks and an absence of this indication will prevent the beam voltage from being activated.

The command control interface PC board serves the purpose of isolating commands from different points of origin, such as microprocessor, local control, or remote control. The buffer control chassis contains control/monitoring circuits and DC power supplies required for the operation of the buffer amplifier.

C. Heat Exchanger

Since the azimuth wrap-up at DSS 13 could not accommodate an extra set of coolant lines required for a ground-based unit, the heat exchanger (HE) was installed (Fig. 8) on one of the platforms above the azimuth bearing. Individual components of the heat exchanger were mounted on the platform in such a way as to provide easy access for service and maintenance. The radiator core and direct drive fan/motor are mounted vertically on the edge of the platform providing a side-discharge air flow away from the antenna structure. The radiator has been successfully tested with a 100-kW heat load. Coolant temperature during this July weather test was maintained within 15°C above ambient. Since coolant temperature fluctuations affect the phase stability of the klystron, the following features were incorporated into the heat exchanger:

(1) A 200-gallon surge tank reduces rapid temperature fluctuations.

(2) Two 18-kW tank heaters provide for cold weather operation.

(3) A three-way modulating valve bypasses the radiator to maintain a set temperature.

(4) A temperature controlled louver on output side of the radiator core is provided.

(5) The direct drive fan/motor is temperature controlled.

The cycle of items 2 through 5 provide coolant stability about a preset, fixed temperature setting.

D. High Voltage Power Supply

The high voltage power supply assembly (Fig. 9) consists of a metal enclosure with two compartments. One compartment contains the high voltage unit, the other contains the crowbar unit and beam voltage monitoring circuits. Attached to one end of the enclosure is a six-foot instrumentation rack which houses control panels, a high voltage control chassis, a computer, and motor-generator field power supplies. Separation of the control and monitoring circuits from the main high voltage section provides necessary shielding of EMI radiation generated by the high voltage section and the crowbar assembly. In addition, all wiring within the crowbar and high voltage sections is run through conduit to further minimize false triggering of control logic and interlock circuits. The high voltage section (Fig. 10) of the power supply contains a high voltage transformer, choke, rectifier stacks, and filtering networks. Input power of three phase 0 to 480 Vac, 400 Hz is applied to the high voltage transformer and stepped up, rectified, and filtered with an output of 20.0 kV at 3 amperes DC. A standard 20-kW S-band high voltage supply was used with only the step-up transformer being changed because an inductrol was not used.

The present S-band 20-kW transmitters in the DSN have fixed output voltage 400-Hz generators. The beam voltage is adjusted by the use of an inductrol, but it is not regulated. The inductrol is set by the operator and no adjustment of beam voltage is made automatically. The regulation is better than 1%. For the developmental X-band transmitter, a field variable generator is used which is part of a closed-loop regulator. The regulation is maintained to 0.1%. A similar means of adjusting the beam voltage is also used by Spacecraft Tracking Data Network's version of the S-band transmitters, except that the regulator controlling the field current of the generator is operated open loop.

An alternative is the use of thyristor phase control at the input to the high voltage transformer. The advantages are: (1) a closed-loop regulator can be implemented with better regulation than can be obtained with a field control generator, (2) the use of thyristors is more reliable and requires less maintenance than a generator, and (3) the fast turn-off capability of thyristors would make it possible to eliminate the crowbar. This alternative is being investigated for the X-band transmitters for implementation into the DSN.

The crowbar section (Fig. 11) contains a dual ignitron crowbar assembly, body current detector, and high voltage dividers for beam voltage monitoring. In the event of a klystron-beam intercept, the crowbar will remove beam voltage from the klystron within 10 microseconds.

The instrumentation rack (Fig. 12) contains all necessary manual transmitter controls and indications. The high voltage control chassis monitors beam voltage levels and provides drive to the motor-generator field power supplies.

E. Motor Generator Set

The motor-generator (Fig. 13) set is a 75-kW unit with a separate start/stop and protection circuit J-box. The generator field is controlled by the high voltage control chassis. This provides a closed-loop system and a regulated high voltage power supply.

III. Software Development

The control system planned for the X-band uplink transmitter was to consist of a microprocessor controller for each major subassembly, namely, one for the power amplifier, another for the heat exchanger and a third for the power supply. In addition, a central master controller was to interface the transmitter with either local operator (or maintenance terminal) or with an interface controller for automatic operation. This arrangement is described in Ref. 3.

However, the use of a controller at the heat exchanger would have been difficult because the heat exchanger is unsheltered and no controller capable of operating in the outside environment was readily available. The heat exchanger was located in close proximity to the power amplifier. A multi-conductor control cable from the heat exchanger to the power amplifier was therefore used to control the heat exchanger directly by the power amplifier microprocessor.

Development of a distributed microprocessor control system with four controllers, (1) the interface controller for station control, (2) the master controller to coordinate the total transmitter operation, (3) the power supply controller to provide high voltage supply signal and fault processing, and (4) the power amplifier to provide the klystron's signal and fault processing, proved to be unwieldy. Three factors led to the abandonment of this arrangement.

Firstly, this system is over-complicated. The program and signal information were duplicated in three microprocessors. In order for the power amplifier controller to perform its function it was necessary to obtain real time status information from the power supply controller.

Second, communication between the various processors caused unacceptable delays. This problem was further aggravated by failures in the optic cable utilized for communication

between the power supply and the power amplifier controllers. After multiple breaks and splicing of the optic cable the losses required the addition of amplifiers.

The third factor was compatibility with the S-band transmitter, namely, the requirement to include both S- and X-band transmit capability. The S-band transmitter power supply, heat exchanger, and cables were to be utilized with two power amplifiers, one at X-band and one at S-band. In the S-band a central controller had been developed with all controls hand-wired between the various subassemblies.

Because of these factors and because of schedule and budget constraints a central processor control was therefore implemented. The central processor control had been operating the high-power S-band transmitter at DSS 13 and it was relatively easy to adapt to this transmitter.

An X-band transmitter simulator was designed and assembled. The simulator is a hardware type and provides the same functions as the X-band transmitter when connected to the controller. This allows the testing of software without any risks of damaging equipment or the expense of operating a full transmitter subsystem.

IV. Software Description

The Local Transmitter Controller (LTC) operates the Transmitter Subsystem (TXR) (see Fig. 14) in an unattended sequence of operations, beginning with system turnon through calibration and culminating in a tracking pass. The LTC monitors the operations of the TXR and submits progress, status, and operational data to display terminals at the Local Control Console (LCC) and the Remote Control Console (RCC). The LTC does not override the internal protection circuitry of the TXR, nor does it prohibit manual intervention by maintenance personnel. If difficulties arise during operation the LTC will alert the operator; then, if the problem is corrected within the operational parameters of the TXR, the LTC will continue to provide automatic control. Otherwise, the operator elects to use the LTC to return the TXR to the off state or to control the TXR manually via the LCC. A brief description of the monitor and control functions of the LTC follows.

Transmitter performance and parameters reported to the operator includes, but is not limited to:

(1) Power output: XX.X kW.

(2) Drive status: on/off.

(3) RF load: antenna/waterload.

(4) Interlocks: reports I/L name if active.

(5) Configuration: transmitter or simulator.

(6) Beam voltage and current: XX.X kV and XX.X A.

(7) Drive power: XX.X MW.

(8) Filament voltage and current: XX.X V and XX.X A.

(9) Magnet current: XX.X A.

(10) Reflected power: XX.X W.

(11) Vacuum current: XX.X μA.

(12) Body current: XX.X MA.

These parameters are sampled at approximately 1 second intervals and displayed on both the local and remote terminals.

Either terminal accepts the following high level command and performs the associated functions without operator intervention.

(1) WAR.

(2) CAL.

(3) PWR XX.X.

(4) STB.

(5) OPR.

(6) OFF.

The WAR command causes the LTC to: turn on the control power supply, MG set, magnet, and filament supplies; wait for the filament time delay to be completed; and, inform the operator when the TXR is ready for calibration.

The CAL command causes the LTC to perform a crowbar test; perveance test; and to saturate the transmitter at the default or requested power level. Following completion of the calibration phase, the LTC removes beam voltage, enters the standby state, and announces that it is ready to operate.

The OPR command causes the LTC to raise the beam voltage to the value settled upon during the calibration phase. Then, following a short delay to allow beam voltage to raise, the LTC informs the operator the status of the following operating parameters:

(1) Output power level.

(2) Saturated or unsaturated mode.

(3) Antenna or waterload.

The STB command causes the LTC to go from the operate to the standby state. The beam voltage is lowered and the LTC announces that the TXR is in the standby state.

The OFF command causes the LTC to begin a shutdown sequence that returns the TXR to the pre-warm-up state. The beam voltage, then drive power, filament voltage and magnet voltage are removed in that order. Finally, the MG set and control power are turned off and the displays are re-drawn.

The PWR XX.X command causes the LTC to use the value XX.X in kW, as the target value for output power during the calibration phase; or, following a request to "SEND CAL," to recalibrate the output power at the new value provided by XX.X while in the operate state.

The LTC interfaces with the TXR via a junction panel as shown in Fig. 15. The firmware needed to perform the functions discussed above is resident on the BLC 8432 prom board. It consists of about 20K bytes of code that was developed using an Intel MDS 800 system, with ISIS operating software, and Intel's PLM80 compiler. Digital interfacing with the many interlock, etc., indicators was provided via the I/O Expansion Board BLC 519. Analog interfaces were provided via a 12-bit analog to a digital, differential input, converter. Including spares, a total of 24-analog and 72-digital input signals are accommodated. Local storage for the status of these signals, along with other variables, is provided by a core memory MCM 8080 board. Twenty-four output signals required to activate relays and motors involved in the control of the TXR are provided on the system CPU board. Signal conditioning and isolation for the interface signals is provided by the computer interface chassis. An RC pi-filter is used in each pair of analog input lines and opto-isolation is provided for all digital input and output lines.

V. Tests

The X-band transmitter has provided the means to perform system tests with associated equipment and, as a consequence, make modifications to various uplink subsystems. The transmitter contributed to the development of S-X band feed assembly, of the diplexer, of the phase correction loop, and of the software. The S-X feed assembly and diplexer development was reported in Ref. 1. Since that report the two-cavity band stop filter has been reworked and the design goal for the receive band isolation/rejection ratio of 73 dB has been met. Simultaneous S- and X-band has been operationally tested. A future report will discuss system noise temperature effects and intermodulation product investigations.

A problem encountered with the phase correction loop forced a decrease in loop gain from 100 to 40. (This has since been analyzed and a new design is now possible with a gain of 100.) The phase loop, consisting of a single pole transfer function, tracks the carrier and has two bandwidths. The 120-Hz wide bandwidth is utilized to correct for 60-Hz modulation. The narrow bandwidth of 3 Hz is utilized to correct for drifts. This phase correction loop does not affect the phase modulation.

The Allan Variance stability of the uplink carrier was measured at the output of the transmitter. Measurement values varied from 0.5 to 1.4 parts in 10^{15}. This represents 6 measurements performed in April and May of this year, and at various times of the day or night with commercial and with diesel prime power. The worst data was obtained between 5 a.m. and 7 a.m. with commercial power. It might be advisable to use diesel power during this time period for the Galileo gravity wave experiments.

There is a phase correction loop from the output of the transmitter to the input of the exciter. Both the transmitter and the exciter have been tested individually with this loop opened and closed. The exciter open-loop measurement values varied from 0.6 to 1.35×10^{-15}, while the closed-loop values were consistent at about 0.7×10^{-15}. The test instrumentation did not allow the transmitter to be measured separately. The uplink chain was measured with the transmitter and the phase loop opened. With the transmitter operating in a saturated power mode, values of 1.13 to 2.05×10^{-15} were obtained. In an unsaturated mode values of 1.06 to 1.85×10^{-15} were obtained. A small improvement may be obtained by operating the transmitter unsaturated. The effect of the transmitter on the uplink carrier can be seen by comparing the exciter data and the exciter and transmitter combination, both with the phase loop open and closed. The variability of the data is due to prime power stability and ambient temperature variations. Improvement can be obtained by tighter supply regulation and stabilization of coolant temperature and flow rate. Phase stability tests are continuing and will be reported for all the applicable subsystems in the future.

VI. Conclusion

The successful operation of the X-band transmitter at DSS 13 has provided a prototype for the development of the hardware necessary for X-band uplink. This technology is now in the process of being transferred to the implementation of X-band transmitters in the DSN which will provide support to Galileo and future spacecraft. In addition, the improved stability will provide new capability in the performance of precision measurements such as the possible detection of gravitational waves.

References

1. Withington, J.R., "Second-Generation X/S Feedcone: Capabilities, Layout and Components," *TDA Progress Report 42-63*, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1981.

2. Kolbly, R.B., "Evaluation of the VA-876P Klystron for the 20 kW X-Band Uplink Transmitter," *DSN Progress Report 42-54*, Jet Propulsion Laboratory, Pasadena, Calif., December 15, 1979.

3. Hartop, R., Johns, C., and Kolbly, R., "X-Band Uplink Ground Systems Development," *DSN Progress Report 42-56*, Jet Propulsion Laboratory, Pasadena, Calif., April 15, 1980.

Table 1. Transmitter characteristics

Parameter	Value
Power output, kW CW	20
Frequency band, MHz	7145 to 7190
Instantaneous band, MHz	45
Track duration, h	12
Transmitter uncompensated, ($\Delta F/F$)	3×10^{-15} over 1000 s

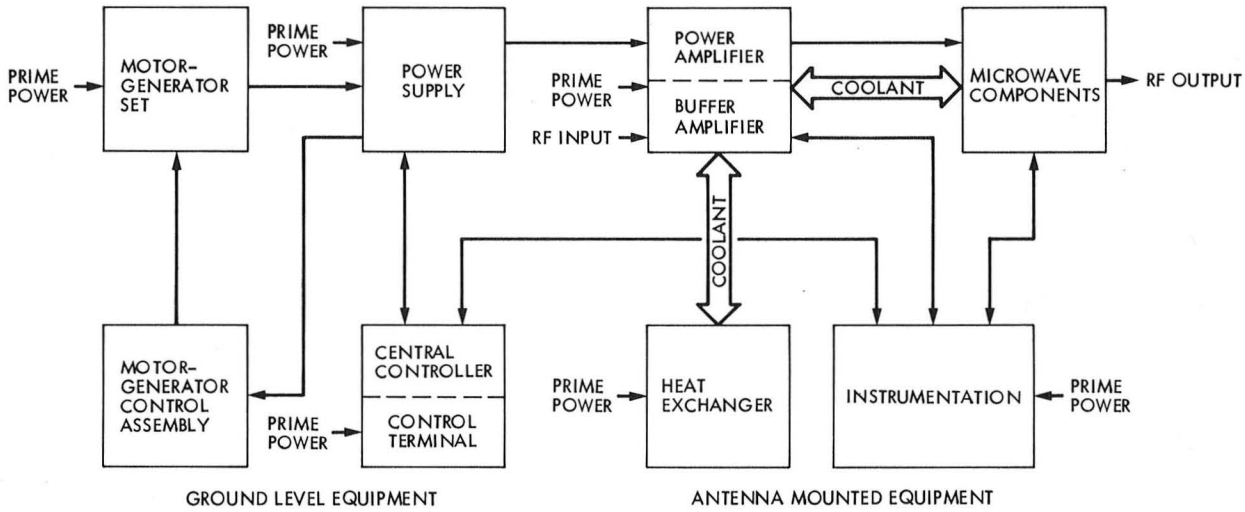


Fig. 1. Equipment block diagram 20-kW X-band transmitter prototype at DSS 13

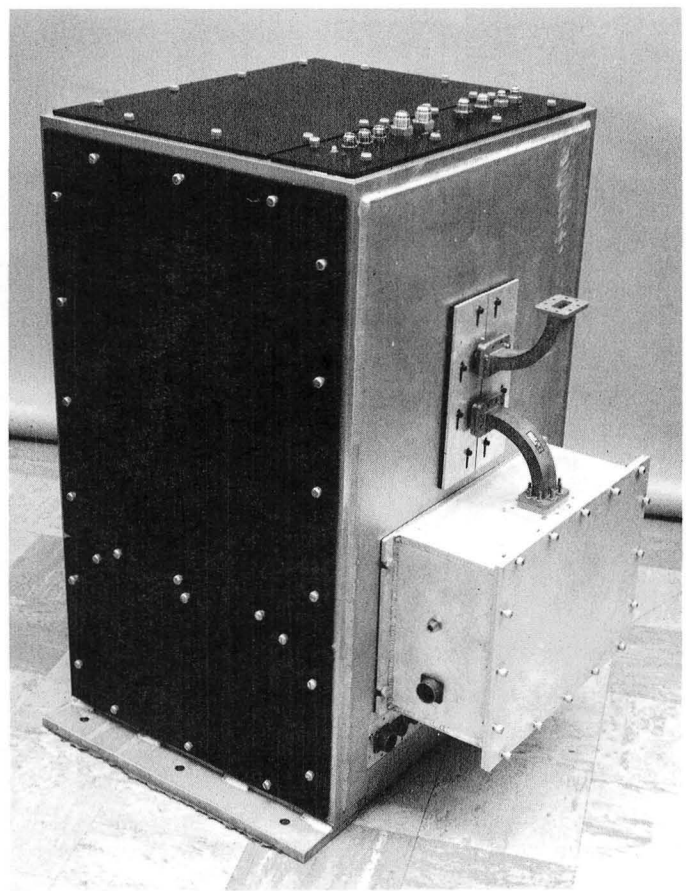


Fig. 2. Power amplifier

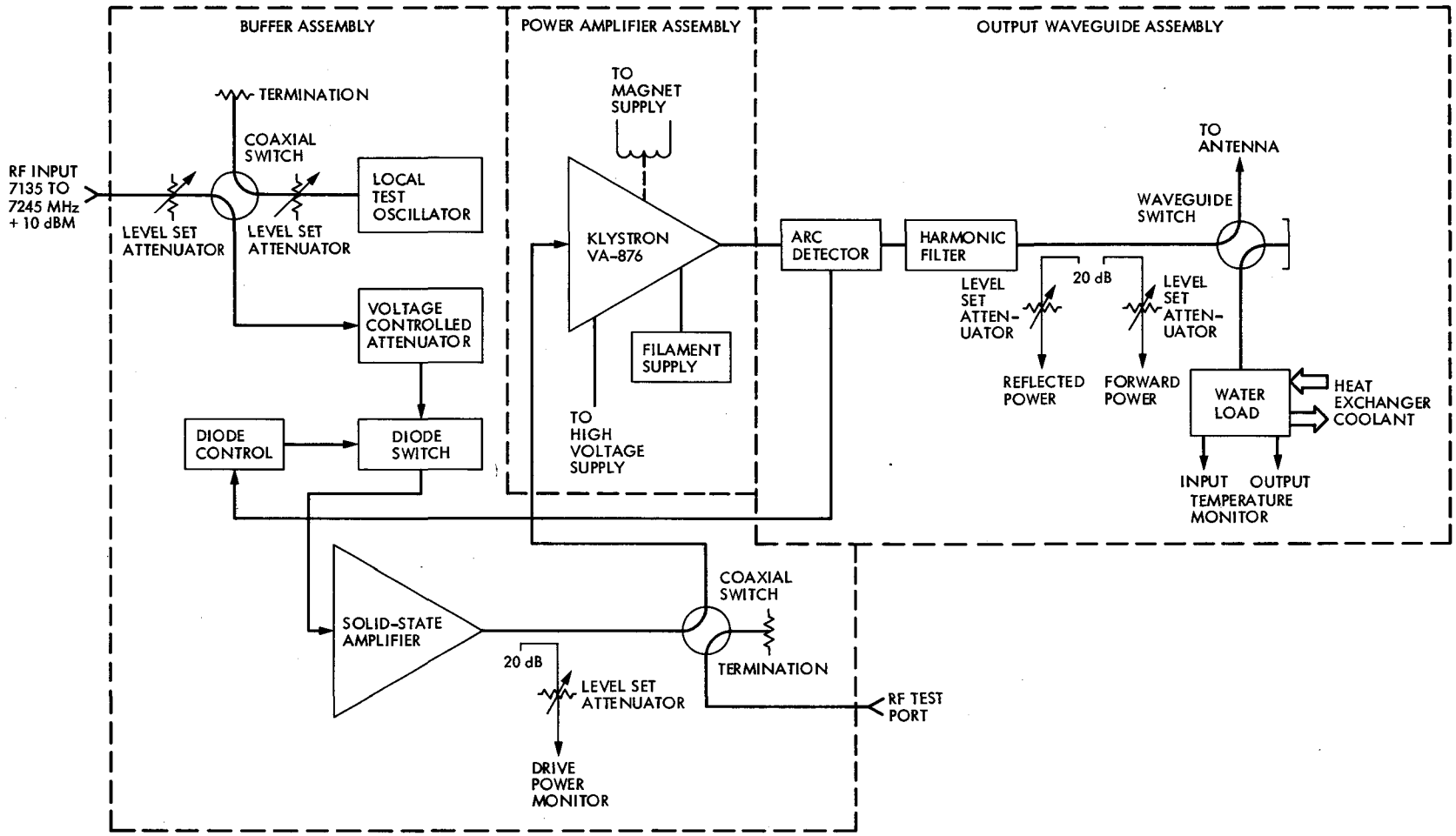


Fig. 3. Block diagram of power amplifier

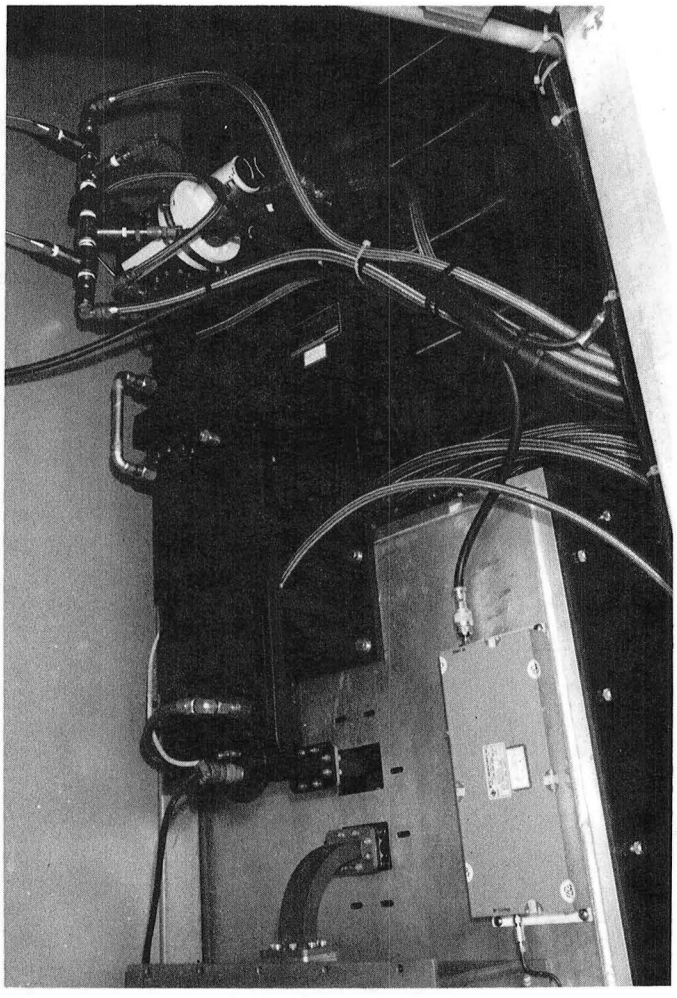


Fig. 4. Output waveguide

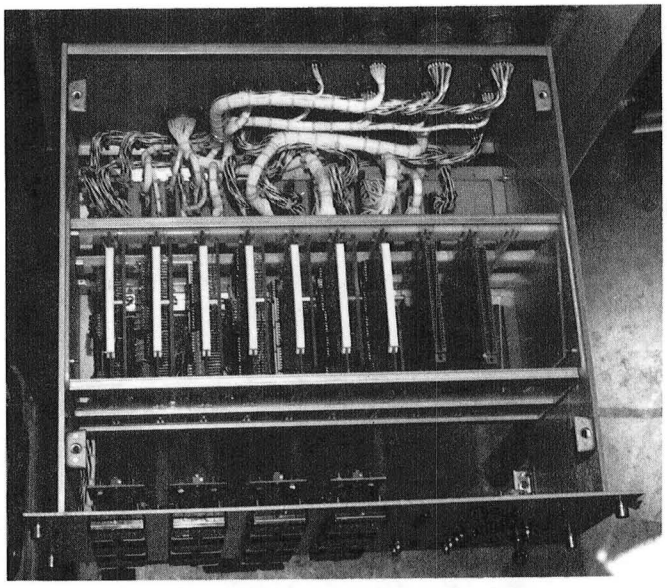


Fig. 5. Control/monitor interface chassis

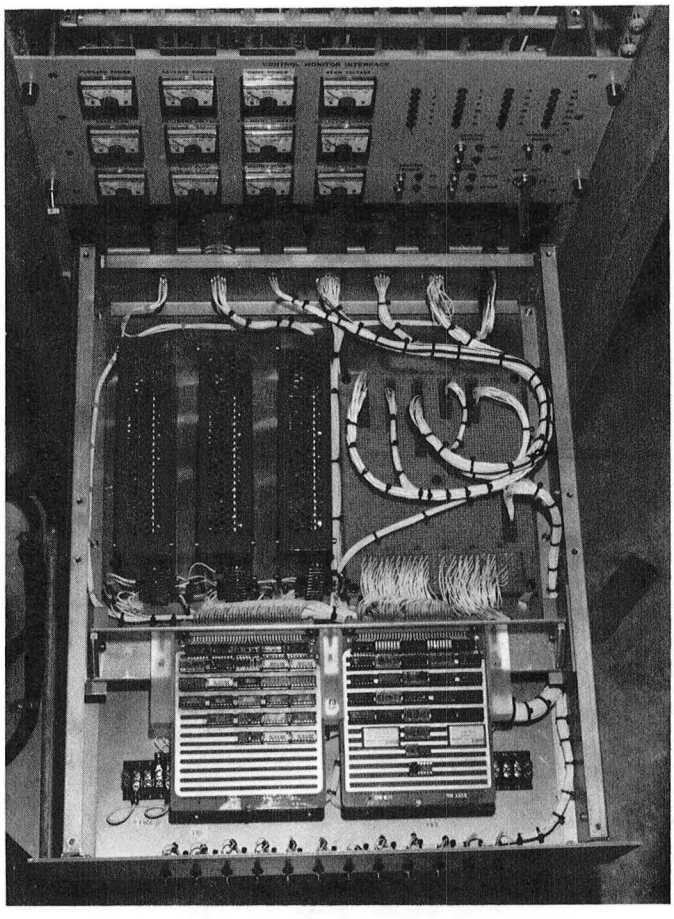


Fig. 6. Buffer control chassis

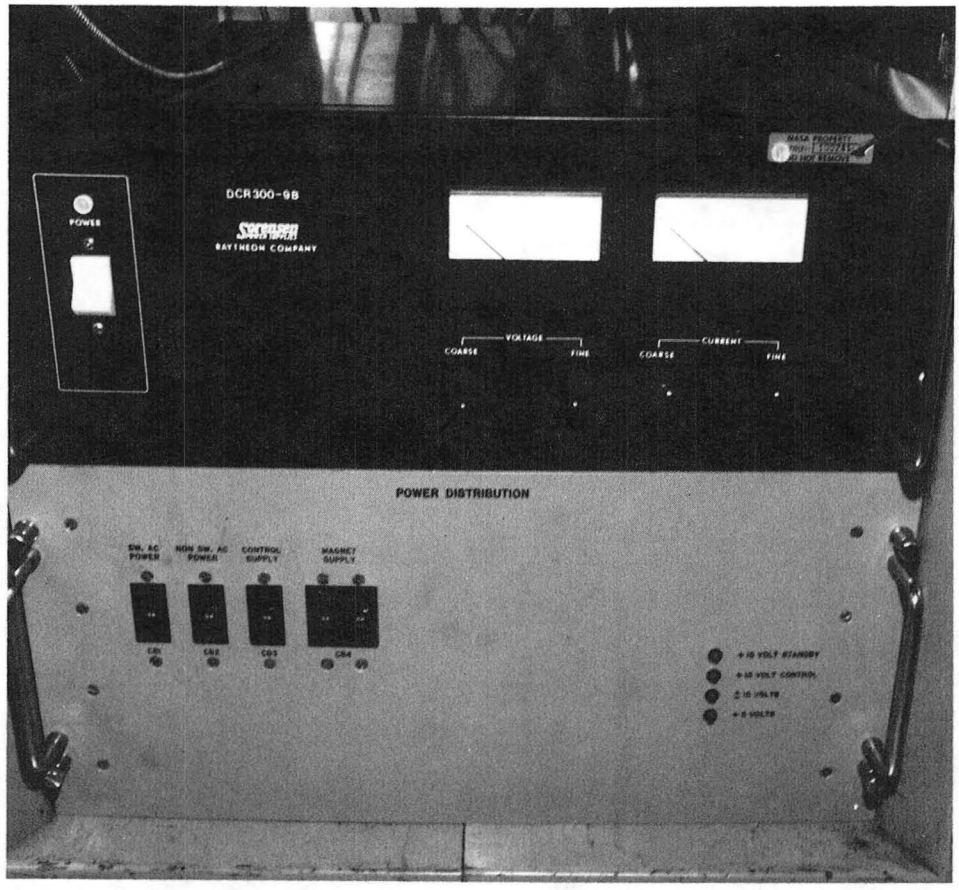


Fig. 7. Magnet and system DC power supplies

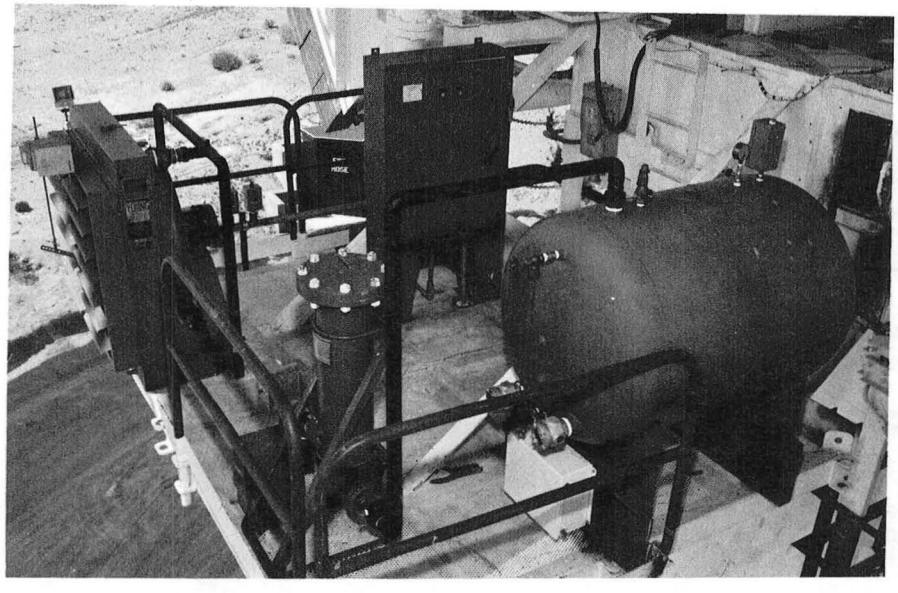


Fig. 8. Heat exchanger

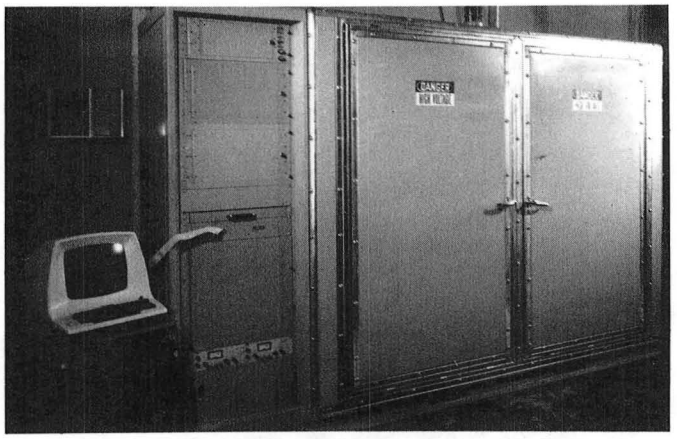


Fig. 9. High voltage assembly

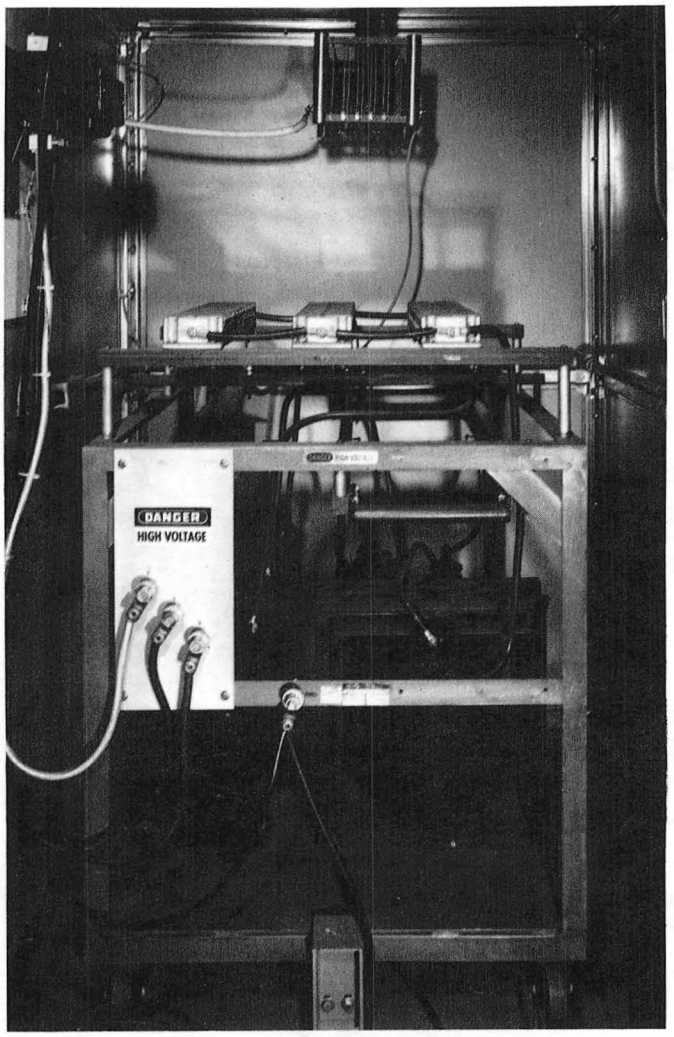


Fig. 10. High voltage section

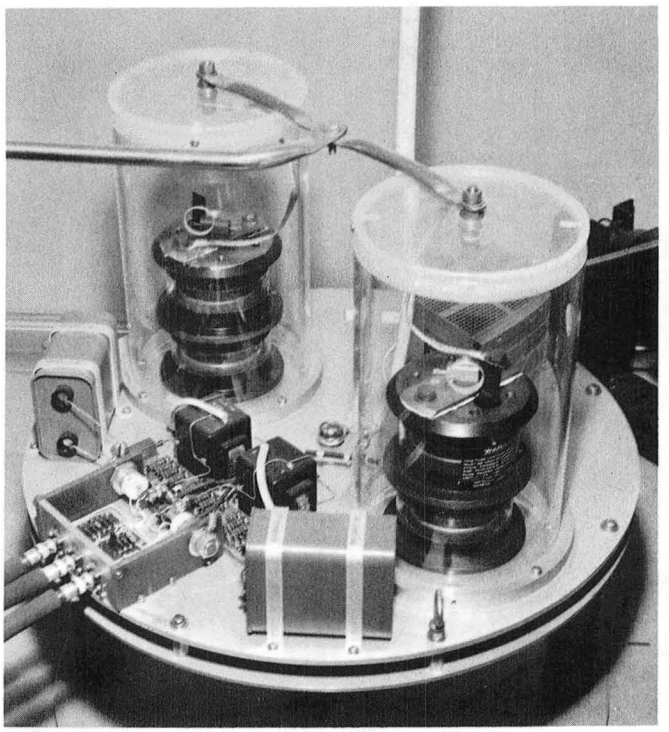


Fig. 11. Crowbar assembly

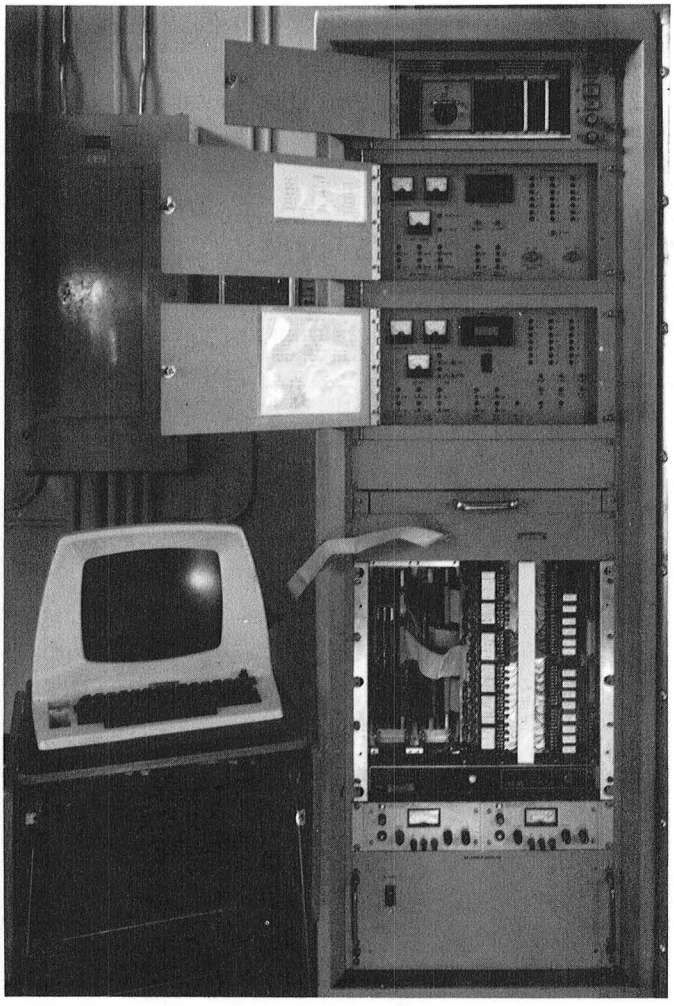


Fig. 12. Instrumentation panels

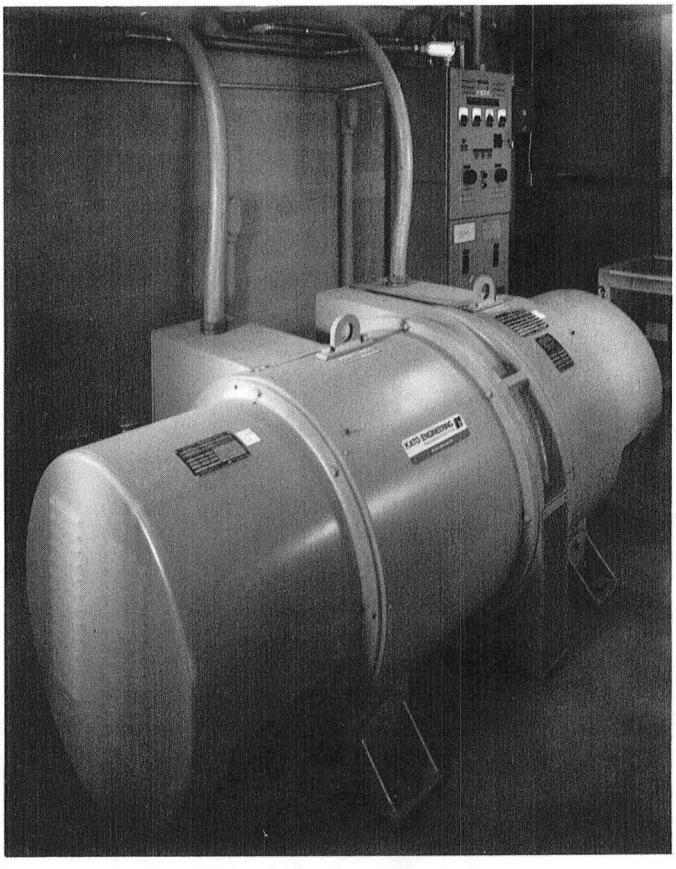


Fig. 13. Motor generator

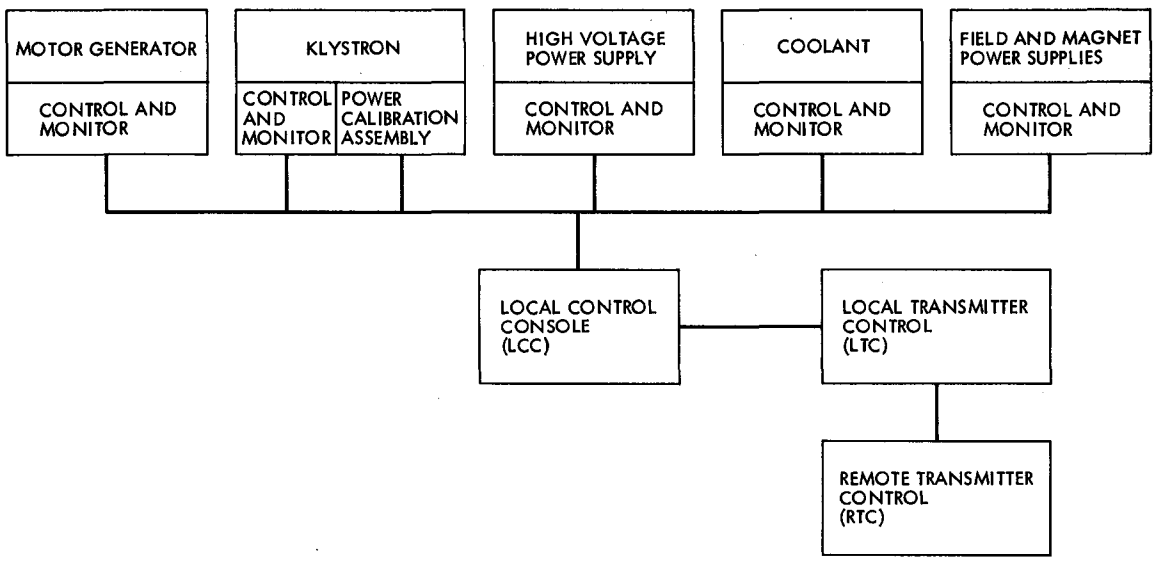


Fig. 14. Transmitter subsystem configuration X-band transmitter automation subsystem

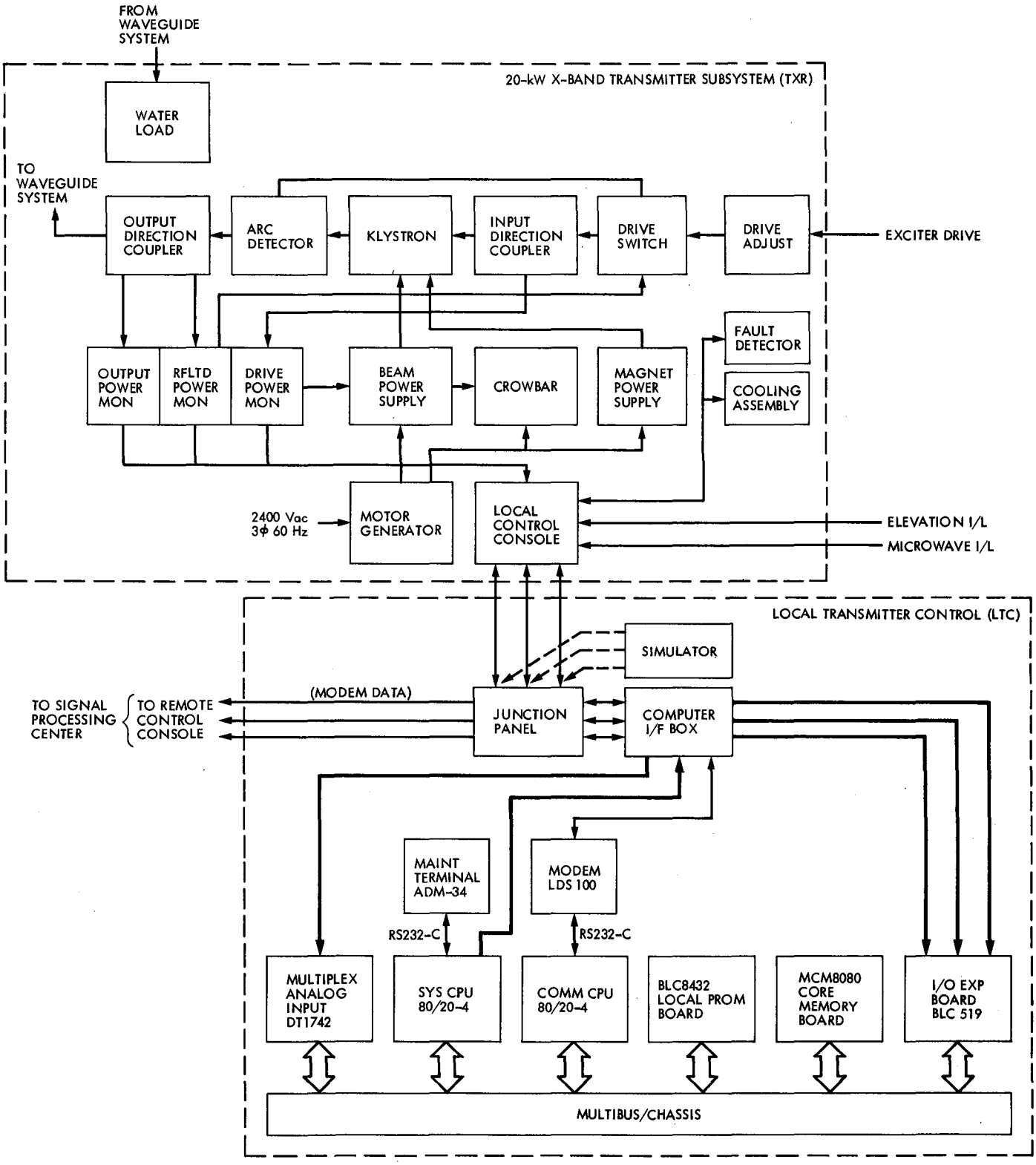


Fig. 15. 20-kW X-band transmitter

Superconducting NbTi and Pb(Cu) Bandpass Filters

J.J. Bautista and S.M. Petty
Radio Frequency and Microwave Subsystem Section

Superconducting $Nb_{0.45}Ti_{0.55}$ and Pb electroplated on Cu bandpass microwave filters have been constructed to investigate their low-loss properties at 4.7 K. An interdigital stripline filter configuration was selected as the optimum structure for future applications. The filter was designed to operate at 8.45 GHz with an equal ripple bandwidth of 0.15 GHz. The insertion loss and the noise temperature contribution were measured at 4.7 K. In addition, the insertion loss was studied as a function of out-of-band power (up to 1 watt) and temperature (4.7 K up to T_c). Results of an oxygen-free high conductivity Cu filter are included for comparison.

I. Introduction

Communication with deep-space probes requires an extremely sensitive system that has an extraordinarily low system noise temperature, 13 to 25 K. The increased bandwidth and sensitivity of the Deep Space Network (DSN) maser-based receiver systems along with the increase in worldwide microwave spectrum usage have dictated the need for employing measures to protect these systems from radio frequency interference (RFI). The effect of RFI on the maser performance is primarily determined by the level of the RFI, its frequency, and how it relates to the frequency of the maser pump source(s). Both in-band and out-of-band RFI can result in gain loss and/or spurious output signals in the maser signal bandpass (Ref. 1).

The objective of this work is to develop filters that will protect the cryogenically-cooled front-end receiver components (e.g., masers, mixers, and upconverters) from the effects of out-of-band RFI. This has to be accomplished without significantly degrading the system noise temperature as well as other microwave performance specifications (e.g., VSWR and group delay variation). Clearly, the very low-loss RF properties of superconducting materials make them attractive choices for use as microwave circuit elements in very-low-noise receivers operating at liquid-helium temperatures.

II. Filter Structure

An interdigital balanced-stripline structure with an air or vacuum dielectric was selected as the optimal design because of its favorable performance characteristics and ease of fabrication from commercially available materials. This structure is compact and has the highest unloaded resonator Q of the commonly used structures (with the exception of the waveguide filter structure) (Ref. 2).

The prototype NbTi filter shown in Fig. 1 consists of five TEM-mode stripline resonators between parallel ground planes.

Each foreshortened resonator is approximately a quarter-wavelength long at the midband frequency and is short-circuited at one end and open-circuited at the other. Final bandpass tuning is accomplished by varying the end capacitance of the resonators with the adjustable screws opposite the open-circuited end. The ground plane contacts are lapped planar and are attached to the filter body with as many screws as possible to reduce the losses at the joints (Ref. 3).

This particular structure was designed with the aid of the low-pass prototype synthesis methods outlined by Matthei, Young, and Jones (Ref. 4). It is an 0.05-dB equal-ripple filter with an equal-ripple bandwidth of 0.15 GHz centered at 8.45 GHz.

Four identical filter structures were machined for the purposes of this work: two from OFHC copper and two from niobium titanium (Ref. 5). All the dimensional tolerances were kept to ±0.051 mm (±0.002 in.) and the cavity surfaces received a 32 machine finish. The center conductor pins were then welded to the input and output launchers while in an inert environment.

Before testing, each filter was given a different final surface preparation. The copper filters were successively rinsed in acetone, dilute nitric acid, and Bright Dip (BD-295). One of the copper filters was next electroplated with a thin layer of lead approximately 10 μm in thickness, chemically polished, and then stored in a dry nitrogen environment (Refs. 6 and 7). Of the two NbTi filters, one (NbTi-II) was mechanically polished with 5-μm alumina powder suspended in water, cleaned, and then given a final rinsing in acetone. The other filter (NbTi-IIA) was left as machined but provided with a machined "knife edge" 0.13 mm (5 mils) wide to further improve the contact at the ground plane joints and then rinsed in acetone.

III. Insertion Loss Measurements

A conventional microwave network analyzer test set was utilized for the return and insertion loss measurements. To measure the losses at 4.7 K, the filters and associated microwave circuit were cooled with a three-stage closed-cycle refrigerator (CCR).

Two copper-plated stainless-steel waveguides (WR-90) with end-on SMA transitions served to carry the input and output signals from 300 K to 4.5 K. These assemblies were tuned to better than 20 dB of return loss across the frequency band of interest.

The reference signal (repeatable to within ±0.05 dB) was obtained by switching the filter out of the microwave circuit with two (commercially available) toggle latching switches modified to operate at 4.5 K. Figure 2 contains a schematic representation of the cryogenic microwave circuit and the measurement setup.

IV. Noise Temperature Measurements

An X-band traveling-wave maser amplifier with an input flange temperature of approximately 4.5 K was used to determine the noise temperature contribution of the filter. The noise temperature of the maser with filter installed and the maser alone was measured using a standard DSN radiometric substitution method (Refs. 8 and 9). In this Y-factor technique, a calibrated horn antenna is used at the receiver input (in this case a Block-IIA prototype DSN maser) to utilize the sky (antenna, atmosphere, and cosmic background) as the cold noise source and an ambient load (Rantec Microwave Absorber positioned over the horn antenna) as the hot noise source.

The respective noise sources are amplified by the maser and the followup field-effect transistor (FET) amplifier, filtered by an adjustable bandpass filter, and the output monitored by a power meter. (Figure 3 contains a schematic representation of the Y-factor method used.) The measured noise powers are then ratioed and substituted in the following expression to obtain the receiver noise temperature (T_r):

$$T_r = T_m + T_f = (T_{am} - YT_{sky})/(Y - 1) \qquad (1)$$

where T_m = noise temperature of the maser, T_f = noise temperature contribution at the maser input due to the followup FET, T_{sky} = noise temperature contribution of horn + atmosphere + cosmic background, T_{am} = noise temperature of ambient load, and Y = ratio of system noise power with hot load to noise power with cold load (Refs. 8 and 9).

The difference between the receiver plus filter noise temperature and the noise temperature of the receiver is taken to be the filter noise temperature contribution.

V. Results and Discussion

A. Insertion Loss

At midband for a Tchebyscheff equal-ripple filter, the insertion loss is given by the expression:

$$L_0\,(\text{dB}) = 8.686\,(C_n/WQ_u) \qquad (2)$$

where W = fractional bandwidth, Q_u = unloaded resonator Q, and C_n = a coefficient determined by the filter order and its inband ripple value (Ref. 10).

At X-band frequencies, the classical-to-anomalous skin-effect transition occurs at approximately 50 K (Ref. 11). Since the insertion loss is proportional to the surface resistance, the loss is thus expected to remain constant below 50 K. For the prototype filter investigated, the minimum insertion loss for Cu is limited by the anomalous skin effect and is one-quarter to one-fifth of its room temperature value (Refs. 12 and 13). Thus, the evaluation of Eq. (2) yields a lower limit of 0.1 dB.

It is well established that at X-band frequencies and at 4.5 K, suitably prepared Pb (T_c = 7.2 K) and NbTi (T_c = 9.8 K) can have surface resistance values three orders of magnitude less than Cu (Refs. 6, 7, 14 and 15). The expected insertion loss for a similar filter made of these superconductors is thus approximately 10^{-4} dB at 4.7 K.

As expected, at room temperature the NbTi and Pb(Cu) filters exhibited higher losses (4.5 and 2.0 dB, respectively) than did the Cu filter (1.0 dB). It was further noticed that the loss of the NbTi was independent of the temperature from 300 to 9.8 K, while the Pb(Cu) displayed a continuous improvement in insertion loss as it cooled from 300 to 4.7 K. Although the predicted superconductor losses at 4.7 K are well below the measurable limits of the test equipment used, measurable losses were present. The lowest value measured for Cu was 0.5 ± 0.06 dB, while NbTi and Pb(Cu) both yielded 0.1 ± 0.06 dB at 4.7 K. (Figure 4 shows the measured insertion loss of the Pb(Cu) filter.)

At the start of this investigation, it was discovered that inadequate ground-plane contacts for the NbTi could result in insertion loss values no better than 0.5 dB at 4.7 K. This was improved by reducing the number of filter assembly parts, maximizing the number of assembly screws, and carefully lapping the contact surfaces before final assembly. The fabrication of a 0.13-mm (5-mil) "knife edge" (NbTi-IIA) at the ground-plane contacts and mechanical polishing (NbTi-II) of the inner filter surfaces did not produce any further measurable improvements.

In addition, it was also found that from 6.0 to 4.7 K the measured insertion loss for the Pb(Cu) and NbTi remained constant. (Figure 5 shows several insertion loss plots at different temperatures for the NbTi-II filter.)

It is suspected that the measured losses are probably associated with the SMA connections at 4.7 K, which were determined to contribute approximately 0.06 ± 0.06 dB.

B. Power Level Effects

Inband and out-of-band CW power tests showed that the inband loss of the Pb(Cu) filter was independent of frequency and power up to 1 watt. The NbTi filter loss, on the other hand, was seriously degraded depending on the CW power level and its frequency relative to the filter's center frequency.

For out-of-band CW signal levels up to 1 watt and frequencies greater than 1.5 GHz on either side of the center frequency, the NbTi inband insertion loss remained unaffected. However, as the center frequency was approached, the affect became more pronounced. For example, at 1.0 GHz on either side of 8.45 GHz for 1 watt of power, the loss increased to 2.0; at 0.5 GHz it increased to 3.0 dB. Inband CW power slightly above 100 mwatts was sufficient to degrade the loss by 2.0 dB. At approximately 50 mwatts, the increase was 0.1 dB. Below 50 mwatts, no inband loss affects were detected.

In addition, inband signal power of approximately 150 mwatts gradually increased the loss 4.0 dB in a manner similar to "thermal runaway." This is caused by the poor thermal conductivity of NbTi (which is 1/1000 times the value for Cu at 4.5 K) and the nearly exponential temperature dependence of the surface resistance below 9.8 K. This difficulty is not present in Pb(Cu) (at least for the power levels of interest) and can be partially alleviated in NbTi by improving the thermal contact.

C. Noise Temperature Contribution

For a low loss and reasonably matched filter, the effective input noise temperature (T_e) of the filter plus maser amplifier can be approximated by the following expression:

$$T_e = (((L - 1) + r^2)\, T_L + LT_m)/(1 - r^2) \qquad (3)$$

where T_L = physical temperature of the filter, T_m = noise temperature of the maser amplifier, L = 1/(transmission coefficient)^2, and r = reflection coefficient (Ref. 16). In the case of a superconducting filter, L is approximately 1, so that Eq. (3) reduces to:

$$T_e = (r^2 T_L + T_m)/(1 - r^2) \qquad (4)$$

Thus, for a filter having r = 0.1 (return loss = 20 dB) and T_L = 4.7 K, the noise temperature contribution (T_e - T_m) will be less than 0.05 K. This would represent an increase of approximately 1% in the input noise temperature of a typical X-band maser (T_m = 4.0 - 4.5 K).

In Fig. 6 are shown the maser and the maser plus filter (NbTi-IIA) input noise temperatures along with their associated return losses. One can see from the figure that the noise temperature of the maser is increased only in the regions where the input return loss is poor. For these measurements

the maser input return loss was laboriously and carefully adjusted to a value of 18 dB or better across the band of interest (0.12 GHz).

For these noise temperature measurements it was found that the largest source of error was associated with the repeatability of the SMA connections and subsequent thermal cycling. This was determined by repeating the input noise temperature measurements of the maser alone and maser plus filter after reconnecting the appropriate low temperature SMA connectors. After several tests of both maser configurations, the repeatability was found to be ±0.5 K between cool-down cycles, while during the particular run the repeatability was ±0.3 K. The worst-case error for these measurements is then ±0.8 K.

To better illustrate the affects of a poor match and to show that the noise temperature measurements are in agreement with the insertion loss measurements, the results in Fig. 6 are replotted in Fig. 7 as the noise temperature contribution. Also shown are the results predicted by Eq. (4) using measured return loss data (continuous curve) assuming a lossless filter. Included for comparison are the results for the Cu filter and the values predicted by Eq. (3) assuming $L(\mathrm{dB}) = 0.5$ dB (dashed curve). Figure 8 contains the same information for the NbTi-II and Pb(Cu) filters, again assuming a lossless filter. It is thus evident from Figs. 7 and 8 that the loss and noise temperature measurements are in agreement. Finally, it is noted that the superconducting filter with the best match (NbTi-II) also yielded the lowest noise temperature contribution.

VI. Conclusion

It has been demonstrated that bandpass, interdigital filters constructed from low-loss superconducting NbTi and Pb(Cu) can be utilized to protect maser amplifiers from out-of-band RFI. In circumstances where power level dependence of the NbTi filter is undesirable, it was shown that Pb(Cu) is the more suitable material.

It is also evident that the noise temperature contribution of superconducting filters to ultra-low-noise maser amplifier systems is more a function of input return loss than dissipation loss. Hence, for the noise temperature contribution of a Block-IIA maser to remain below 0.2 K, the maser-plus-filter-input return loss must be 17 dB or better.

VII. Acknowledgements

The authors wish to thank and express their indebtness to G. J. Dick and E. Boud of Caltech who performed the lead plating and polishing. The authors gratefully acknowledge the aid of D. Zanteson, R. Stuber, and D. Norris for their help in obtaining some of the data. Useful discussions with T. Y. Otoshi and R. C. Clauss were also greatly appreciated.

References

1. Bautista, J. J., and S. M. Petty, "Cryogenic Filters for RFI Protection," *TDA Progress Report 42-65*. Jet Propulsion Laboratory, Pasadena, Calif., July-Aug. 1981, p. 94.
2. Matthaei, G. L., L. Young, and E. M. T. Jones, *Microwave Filters, Impedance-Matching Networks, and Coupling Structures*. Artech House Books, Dedham, Mass., 1980, pp. 421-427.
3. Oelfke, W. C., and W. O. Hamilton, "Design and Preparation of High-Q Niobium Reentrant Cavities for Physical Measurements," *Rev. Sci. Instrum.*, Vol. 54, No. 4, pp. 410-414, April 1983.
4. Matthaei, G. L., et. al., pp. 614-626, op. cit.
5. Niobium titanium stock obtained from Teledyne Wah Chang, Albany, Oregon.
6. Dick, G. J., J. R. Delayen, and H. C. Yen, "A Polishing Procedure for High Surface Electric Fields in Superconducting Lead Resonators," *IEEE Trans. Nucl. Sci.*, Vol. NS-24, pp. 1130-1132, June 1977.
7. Dick, G. J., and J. R. Delayen, "A New Chemical Polishing Procedure for Lead-Plated Copper Superconducting Accelerating Resonators," *IEEE Trans. on Magnetics*, Vol. MAG-19, No. 3, pp. 1315-1317, May 1983.
8. Stelzried, C. T., "Operating Noise Temperature Calibrations of Low-Noise Receiving Systems," *Microwave Journal*, pp. 1-7, June 1971.
9. Stelzried, C. T., *The Deep Space Network – Noise Temperature Concepts, Measurements, and Performance*, Publication 82-33. Jet Propulsion Laboratory, Pasadena, Calif., September 15, 1982, pp. 11.1-12.8.
10. Matthaei, G. L., et. al., pp. 150-151, op. cit.
11. Rathke, J. E., *The Transient Analysis of Coaxial Cables at Low Temperatures Considering Anomalous and Classical Skin Effects with the Inclusion of Electron Relaxation Phenomena*, Ph.D. Thesis, University of Kansas (1969).
12. Pippard, A. B., "Metallic Conduction at High Frequencies and Low Temperatures," *Advances in Physics*, Vol. VI, pp. 1-44, Academic Press, Inc., New York, N.Y., 1954.
13. Bernard, J., N. H. Minyawi and N. T. Viet, "Reduction of RF Losses at 35 GHz in High Purity *Copper* Resonant Cavities by Cooling to Cryogenic Temperatures," *Revue De Physique Appliquie*, Vol. 13, pp. 483-487, 1978.
14. Giordano, S., H. Hahn, H. J. Halama, T. S. Luhman and W. Bauer, "Investigation of Microwave Properties of Superconducting $Nb_{0.4}Ti_{0.6}$," *IEEE Trans. on Magnetics*, Vol. MAG-11, No. 2, pp. 437-440, March 1975.
15. Momose, T., T. Yamashita and Y. Onodera, "Surface Resistance due to Trapped Magnetic Flux of Superconducting Lead at X-band," *J. Appl. Phys.*, Vol. 54, No. 7, pp. 4044-4049, July 1983.
16. Otoshi, T. Y., "The Effect of Mismatched Components on Microwave Noise-Temperature Calibrations," *IEEE Trans. on Microwave Theory and Techniques*, Vol. MTT-16, No. 9, pp. 675-686, Sept. 1968.

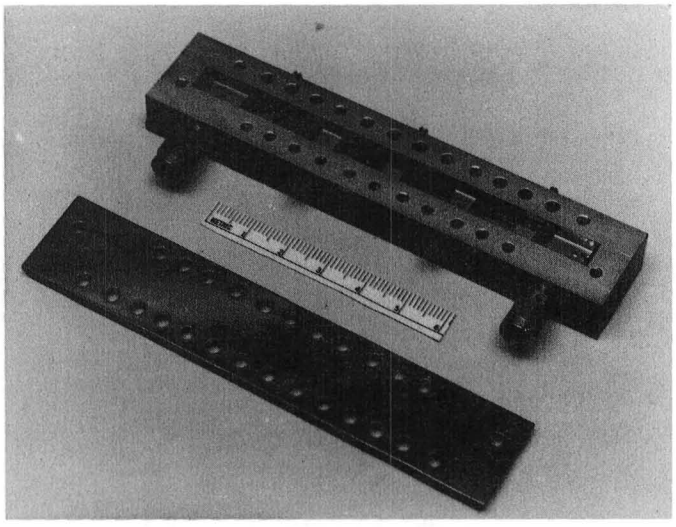


Fig. 1. Photograph of NbTi-II filter

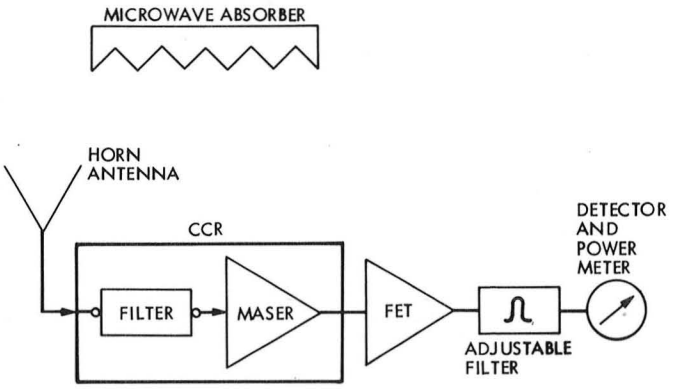


Fig. 3. Noise temperature measurement setup

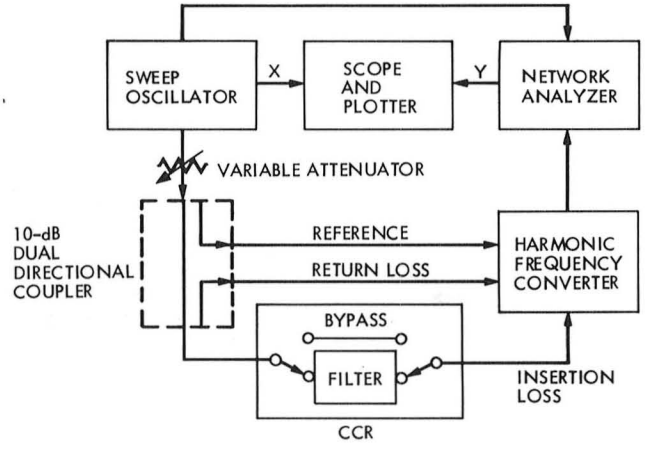


Fig. 2. Insertion loss measurement setup

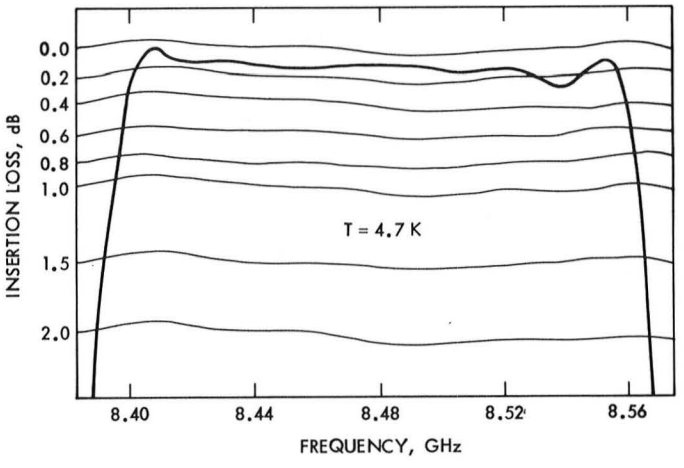


Fig. 4. Insertion loss of the Pb(Cu) filter at 4.7 K

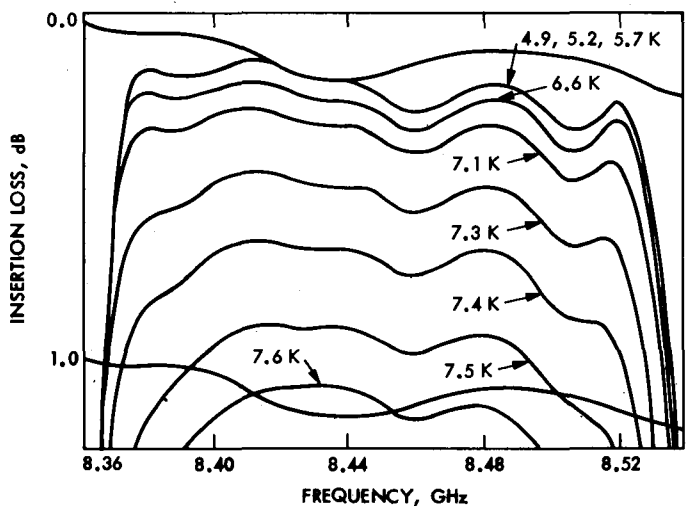


Fig. 5. Insertion loss of the NbTi-II filter for various temperatures

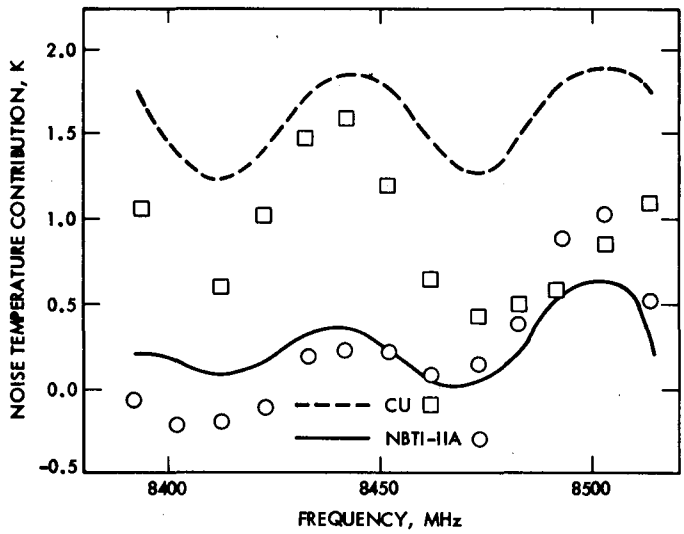


Fig. 7. Noise temperature contribution of NbTi-IIA and Cu filters

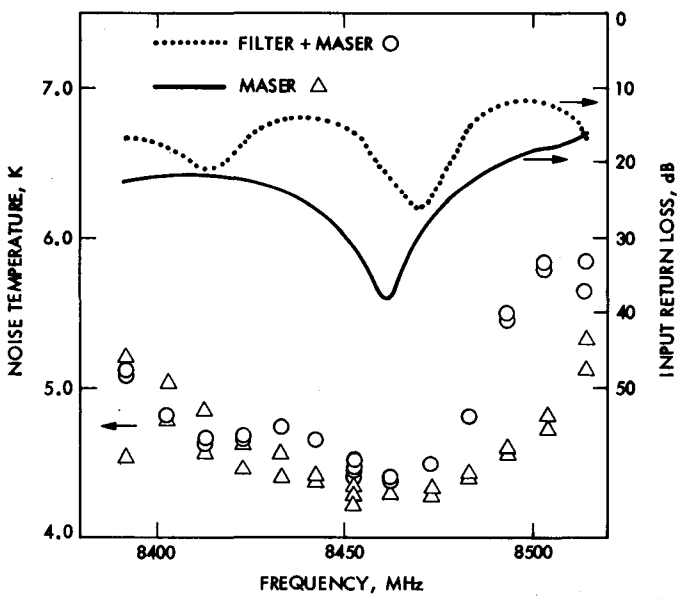


Fig. 6. Measured noise temperature and return loss of filter plus maser and maser alone

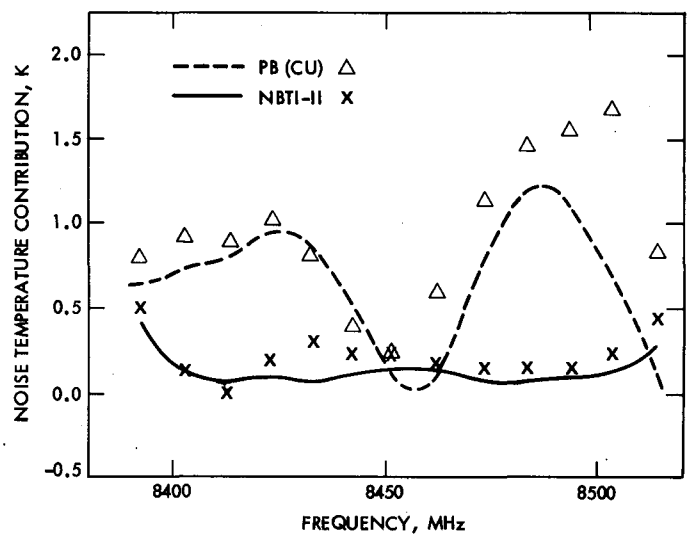


Fig. 8. Noise temperature contributions of Pb(Cu) and NbTi-II filter

Avalanche Photodiode Statistics in Triggered-Avalanche Detection Mode

H. H. Tan
Communications Systems Research Section

The output of a triggered-avalanche mode avalanche photodiode is modeled as Poisson-distributed primary avalanche events plus conditionally Poisson-distributed trapped-carrier-induced secondary events. The moment generating function as well as the mean and variance of the diode output statistics are derived. The dispersion of the output statistics is shown to always exceed that of the Poisson distribution. Several examples are considered in detail.

I. Introduction

When a sufficiently large enough voltage is applied across an avalanche photodiode (APD), free carriers accelerated by the electric field attain enough energy to generate secondary carrier pairs by impact ionization. This leads to carrier avalanche and at a high enough voltage to avalanche breakdown. The normal operating mode of APDs biased below the avalanche breakdown voltage does not, however, have sufficiently favorable noise characteristics for photon counting applications. Recent work (Refs. 1 through 3) has indicated that the APD operated in a triggered-avalanche detection (TAD) mode might be effective for photon-counting applications.

Suppose initially an APD is biased below avalanche breakdown and is sufficiently cooled and shielded to eliminate all free carriers in the APD junction region. The bias voltage is then increased above the avalanche breakdown voltage threshold, but below the zener breakdown voltage. In the absence of free carriers in the junction, no avalanche breakdown can occur. The presence of a single free carrier, for example as a result of photon absorption, can initiate a self-sustaining avalanche breakdown in which the current flow in the diode grows exponentially until the bias voltage is reduced for a sufficient period of time to quench the avalanche discharge and sweep all the free carriers from the diode junction. The large bias can then be reapplied so that the diode can again detect photons. This is called a triggered-avalanche detection (TAD) mode (Refs. 1 and 2) or a Geiger-counter mode (Ref. 3) of operation.

The TAD-mode output pulses are sufficiently large so that no additional amplification is necessary to count them. Hence, single-photon detection is feasible. Moreover, TAD-mode APDs can have significant quantum efficiency improvements over the photomultiplier tubes commonly used for photon counting. TAD-mode APDs, however, have a problem with carrier trapping. During each avalanche, carriers can become trapped at dislocations and impurities within the diode. Those trapped carriers not swept out during the quenching period can subsequently initiate avalanches when they are released

from the traps. Experimental results give evidence of trap lifetimes of up to minutes in duration (Refs. 1 to 3) over which carriers are released. It appears that an evaluation of the photon counting capability of the TAD-mode APD has to include the effect of carrier-induced avalanches.

This article considers a simple and most possibly naive statistical model to account for carrier trapping in a TAD-mode APD. This model is then analyzed to obtain the statistics of the diode output. The moment generating function of the diode output statistics is derived in Section II along with the mean and variance. Specific examples are considered in Section III and a discussion of the results is given in Section IV.

II. Moment Generating Function of Count Statistics

Avalanche discharges in a TAD-mode APD are initiated either by thermally or photon-absorption-induced free carriers or by trap-released free carriers. The first two types of free-carrier-induced discharges can be regarded as primary discharge events. Since trapped carriers exist as a result of previous avalanche discharges, the discharges generated by these carriers can be regarded as secondary events. We shall make the assumption here that trapped-carrier-induced secondary avalanche discharges do not produce more trapped carriers. That is, all the trapped carriers in the APD were generated by previous primary avalanche discharge events. This simplifying assumption is made to make the subsequent analysis tractable.

To describe the diode output statistics, let us define for each $t > 0$, the counting processes:

$N_1(t)$ = number of primary avalanche events due to either thermally or photon-induced free carriers occurring in a time interval $[0, t)$.

$N_2(t)$ = number of secondary trapped-carrier-induced avalanche events in a time interval $[0, t)$.

It can be assumed that $N_1(t)$ is a Poisson process with intensity rate $\lambda(t)$ equal to the average rate at which primary avalanche discharge events occur. This intensity rate $\lambda(t)$ typically includes a constant rate at which thermally generated carriers initiate avalanche discharges and a rate at which photoelectron-initiated avalanches occur on the average.

We assume next that each primary avalanche event occurring at time τ will generate secondary avalanche events in the time interval $[\tau, \infty)$ at an average rate $h(t - \tau)$. It is assumed that this rate function $h(t)$ of generating trapped-carrier avalanche events by a single primary event is fixed and deterministic. So, if $\{t_i\}$ are the occurrence times of the primary avalanche events, then conditioned on the counting process $N_1(t)$, $N_2(t)$ is a conditional Poisson process with intensity rate $\lambda_2(t)$ given by

$$\lambda_2(t) = \sum_{i=1}^{N(t)} h(t - t_i), \qquad t \geq 0 \tag{1}$$

Since $\lambda_2(t)$ is a random process, $N_2(t)$ is effectively a Poisson process with random intensity rate. Such processes are called doubly stochastic Poisson processes (Ref. 4). Moreover, $\lambda_2(t)$ is a filtered Poisson process or shot-noise process (Ref. 4) with filter impulse response $h(t)$.

The process $N_1(t) + N_2(t)$ then gives the total number of avalanche events in the TAD-mode APD output during the time period $[0, t)$. It is interesting to determine the statistics of the TAD-mode APD output in some time interval after the device has been operating for a long period of time. To determine such statistics, let $T > 0$ and $\Delta T > 0$ and define

$\Delta N_1 = N_1(T + \Delta T) - N_1(T)$
= number of primary avalanche events in the time interval $[T, T + \Delta T)$

$\Delta N_2 = N_2(T + \Delta T) - N_2(T)$
= number of secondary avalanche events in the time interval $[T, T + \Delta T)$

The random variables of interest are then

$\Delta N_1 + \Delta N_2$ = total number of avalanche events in the TAD-mode APD in the time interval $[T, T + \Delta T)$

Assume that the device started operation at $t = 0$. Then $\Delta N_1 + \Delta N_2$ represents the number of avalanche events in the TAD-mode APD output during a time interval of length ΔT after the device has already been operating for a time period of length T. The case of particular interest is then when T is significantly larger than ΔT.

Teich and his collaborators (Ref. 5) have studied doubly stochastic Poisson processes such as $N_2(t)$. They have determined the moment generating function and several of the moments of $\Delta N_1 + \Delta N_2$ only in the case when $T = 0$ and hence are not useful here. Our goal is to determine the moment generating function (mgf)

$$\phi(s) = E\left[e^{-s(\Delta N_1 + \Delta N_2)}\right] \tag{2}$$

of $\Delta N_1 + \Delta N_2$ for general $\lambda(t)$ and $h(t)$. The moments of $\Delta N_1 + \Delta N_2$ can also be determined using this mgf. A detailed derivation given in Appendix A shows that

$$\phi(s) = \phi_1(s) \cdot \phi_2(s) \tag{3}$$

where

$$\phi_1(s) = \exp\left\{\int_0^T \lambda(\tau) \times \left[\exp\left\{(e^{-s} - 1)\int_T^{T+\Delta T} h(t-\tau)\,dt\right\} - 1\right] d\tau\right\} \quad (4)$$

and

$$\phi_2(s) = \exp\left\{\int_T^{T+\Delta T} \lambda(\tau) \times \left[e^{-s}\exp\left\{(e^{-s} - 1)\int_T^{T+\Delta T} h(t-\tau)\,dt\right\} - 1\right] d\tau\right\} \quad (5)$$

To understand the significance of $\phi_1(s)$ and $\phi_2(s)$, consider the following special cases. First, suppose that $\lambda(t) = 0$ for $0 \leqslant t \leqslant T$. Then $\phi(s) = \phi_2(s)$. In this case there are no primary events in the time interval $[0, T)$. Hence, we can conclude that $\phi_2(s)$ is the mgf of the sum of the primary events in the time interval $[T, T + \Delta T)$ plus the secondary events in this time period caused by these primaries. Next, consider when $\lambda(t) = 0$ for $T \leqslant t \leqslant T + \Delta T$. Then $\phi(s) = \phi_1(s)$. In this case, there are no primary events in the time interval $[T, T + \Delta T)$. Hence $\phi_1(s)$ is the mgf of the secondary events in the time interval $[T, T + \Delta T)$ caused by the primary events in the interval $[0, T)$. Since $\phi_1(s)$ depends only on $\lambda(t)$ for $0 \leqslant t \leqslant T$ and $\phi_2(s)$ depends only on $\lambda(t)$ for $T \leqslant t \leqslant T + \Delta T$, these two conclusions hold in general. Also, note the statistical independence of these two count sums, which is a result of the independent increments property of $N_1(t)$.

Finally, derivatives of $\phi(s)$ at $s = 0$ can be determined to obtain the moments of $\Delta N_1 + \Delta N_2$. It can be shown that

$$-\phi_1'(0) = \int_0^T \lambda(\tau)\left[\int_T^{T+\Delta T} h(t-\tau)\,dt\right] d\tau \quad (6)$$

$$\phi_1''(0) = \left[\phi_1'(0)\right]^2 + \int_0^T \lambda(\tau) \times \left[\int_T^{T+\Delta T} h(-\tau)\,dt + \left(\int_T^{T+\Delta T} h(t-\tau)\,dt\right)^2\right] d\tau \quad (7)$$

$$-\phi_2'(0) = \left[\int_T^{T+\Delta T} \lambda(\tau)\int_T^{T+\Delta T} h(t-\tau)\,dt\right] d\tau + \int_T^{T+\Delta T} \lambda(\tau)\,d\tau \quad (8)$$

$$-\phi_2''(0) = [\phi_2'(0)]^2 + \int_T^{T+\Delta T} \lambda(\tau) \times \left[\int_T^{T+\Delta T} h(t-\tau)\,dt + \left(\int_T^{T+\Delta T} h(t-\tau)\,dt\right)^2\right] d\tau + 2\int_T^{T+\Delta T} \lambda(\tau)\left[\int_T^{T+\Delta T} h(t-\tau)\,dt\right] d\tau + \int_T^{T+\Delta T} \lambda(\tau)\,d\tau \quad (9)$$

Hence, it follows from Eqs. (3), (6), and (8) that the mean of $\Delta N_1 + \Delta N_2$ is given by

$$E[\Delta N_1 + \Delta N_2] = \int_T^{T+\Delta T} \lambda(\tau)\left[\int_T^{T+\Delta T} h(t-\tau)\,dt\right] d\tau + \int_T^{T+\Delta T} \lambda(\tau)\,d\tau \quad (10)$$

Since

$$\int_T^{T+\Delta T} \lambda(\tau)\,d\tau$$

is the average number of primary events in the time interval $[T, T + \Delta T)$, it follows that the first term in Eq. (10) is the average number of secondary events in that interval. Moreover, Eq. (6) gives the average number of secondaries in $[T, T + \Delta T)$ that are a result of primaries in $[0, T)$.

The variance of $\Delta N_1 + \Delta N_2$ is obtained from Eqs. (3), (7), and (9) and is given by

$$\text{var}\,(\Delta N_1 + \Delta N_2) = \int_0^{T+\Delta T} \lambda(\tau)\left[\int_T^{T+\Delta T} h(t-\tau)\,dt + \left(\int_T^{T+\Delta T} h(t-\tau)\,dt\right)^2\right]d\tau + 2\int_T^{T+\Delta T} \lambda(\tau) \times \left[\int_T^{T+\Delta T} h(t-\tau)\,dt\right]d\tau + \int_T^{T+\Delta T} \lambda(\tau)\,d\tau \qquad (11)$$

Note that

$$\int_T^{T+\Delta T} \lambda(\tau)\,d\tau$$

is the variance of the primary count in $[T, T+T)$. Note from Eq. (7) that the variance of the secondary count in $[T, T+T)$ due to primaries in $[0, T)$ is given by

$$\int_0^T \lambda(\tau)\left[\int_T^{T+\Delta T} h(t-\tau)\,dt + \left(\int_T^{T+\Delta T} h(t-\tau)\,dt\right)^2\right]d\tau \qquad (12)$$

Hence, it follows that

$$\int_T^{T+\Delta T} \lambda(\tau)\left[\int_T^{T+\Delta T} h(t-\tau)\,dt + \left(\int_T^{T+\Delta T} h(t-\tau)\,dt\right)^2\right]d\tau \qquad (13)$$

must be the variance of the secondary count in $[T, T+\Delta T)$ due to primaries in that same interval. So the first term in Eq. (11) represents the variance of the total secondary count in $[T, T+\Delta T)$. Moreover, an examination of Eqs. (9) and (13) yields the obvious conclusion that the second term in Eq. (11) is the covariance between the primary and secondary counts in $[T, T+\Delta T)$.

Another interesting statistic for counting distributions is the dispersion or variance-to-mean ratio of the distribution. The dispersion of a Poisson counting distribution is equal to one. It is then of some interest to compare dispersions with that of the Poisson distribution. Here

$$\frac{\text{var}\,(\Delta N_1 + \Delta N_2)}{E[\Delta N_1 + \Delta N_2]} = 1 + \frac{\displaystyle\int_0^{T+\Delta T} \lambda(\tau)\left[\int_T^{T+\Delta T} h(t-\tau)\,dt\right]^2 d\tau + 2\int_T^{T+\Delta T} \lambda(\tau)\left[\int_T^{T+\Delta T} h(t-\tau)\,dt\right]di}{\displaystyle\int_0^{T+\Delta T} \lambda(\tau)\left[\int_T^{T+\Delta T} h(t-\tau)\,dt\right]d\tau + \int_T^{T+\Delta T} \lambda(\tau)\,d\tau} \qquad (14)$$

Hence, the quotient term in Eq. (14) gives the excess dispersion over the dispersion of the Poisson distribution. This implies that the distribution of $\Delta N_1 + \Delta N_2$ has a wider spread about its mean than the Poisson distribution.

III. Examples

Example 1: Constant Light Source and Constant Rate of Trapped Carrier Induced Events

Consider when

$$\lambda(t) = \bar{\lambda} = \text{constant}$$

and

$$h(t) = \begin{cases} A, & 0 \leq t \leq T_h \\ 0, & \text{otherwise} \end{cases}$$

where

$$\Delta T \leq T_h \leq T \text{ (see Fig. 1)} \qquad (15)$$

An average constant rate $\bar{\lambda}$ of primary avalanche events per unit time is generated here as a result of either photoelectrons or thermally induced free carriers in the diode. Moreover, it

is assumed that each primary avalanche event will, on the average, generate trapped induced secondary events at a constant rate A per unit time over a time period of length T_h. We assume that $T \geqslant T_h$ so that the diode has reached a steady-state condition. We also assume for convenience that the avalanche event counting interval $\Delta T \leqslant T_h$. Experimental work (Refs. 1 and 3) has reported T_h ranging from a few milliseconds to minutes. Since $N_2(t)$ is a doubly stochastic Poisson process, AT_h represents the average number of secondary avalanche events generated per primary event. There is experimental work (Ref. 3) that indicates that $AT_h = 0.1$ may be achieved with commercial APDs.

For this example, it is easy to show that

$$\int_T^{T+\Delta T} h(t-\tau)\,dt = \begin{cases} 0, & 0 \leqslant \tau \leqslant T - T_h \\ A(\tau - T + T_h), & T - T_h \leqslant \tau \leqslant T - T_h + \Delta T \\ A(\Delta T), & T - T_h + \Delta T \leqslant \tau \leqslant T \\ A(T + \Delta T - \tau), & T \leqslant \tau \leqslant T + \Delta T \end{cases} \tag{16}$$

Substitution of Eq. (16) into Eqs. (3) through (5) yields

$$\phi(s) = \exp\left\{\frac{\bar{\lambda}\,[\exp\{(e^{-s}-1)\,A(\Delta T)\}-1]}{A(e^{-s}-1)} - \bar{\lambda}(\Delta T)\right\}$$
$$\cdot \exp\{\bar{\lambda}\,(T_h - \Delta T)\,[\exp\{(e^{-s}-1)\,A(\Delta T)\}-1]\}$$
$$\cdot \exp\left\{\frac{\bar{\lambda}\,e^{-s}\,[\exp\{(e^{-s}-1)\,A(\Delta T)\}-1]}{A(e^{-s}-1)} - \bar{\lambda}(\Delta T)\right\} \tag{17}$$

The product of the first two exponential factors in Eq. (17) is the mgf $\phi_1(s)$ and the last exponential factor is the mgf $\phi_2(s)$. The second exponential factor turns out to be the mgf of a two-parameter Neyman Type-A probability distribution (Ref. 6). Hence all three exponential factors in Eq. (17) are mgfs.

Substitution of Eq. (16) into Eqs. (10) and (11) yields

$$E[\Delta N_1 + \Delta N_2] = \bar{\lambda}\Delta T + \bar{\lambda}T_h(A\Delta T) \tag{18}$$

$$\text{var}\,(\Delta N_1 + \Delta N_2) = \bar{\lambda}T_h(A\Delta T) + \bar{\lambda}\Delta T + \bar{\lambda}T_h(A\Delta T)^2 + \bar{\lambda}\Delta T(A\Delta T) - \frac{\bar{\lambda}\Delta T(A\Delta T)^3}{3} \tag{19}$$

The first term in Eq. (18) is the average number of primaries and the second term is the average number of secondaries. Note that the average number of secondaries is equal to (the average number of primaries over a time interval of length T_h) X (average number of secondaries per primary) $= (\bar{\lambda}T_h) \times (A\Delta T)$. This is about $(\bar{\lambda}\Delta T)\,(A\Delta T)$ less than one would expect at a first glance. The loss is due to edge effects as evidenced in Eq. (16). The dispersion is given by

$$\frac{\text{var}\,(\Delta N_1 + \Delta N_2)}{E\,[\Delta N_1 + \Delta N_2]} = 1 + \frac{\bar{\lambda}T_h(A\Delta T)^2 + \bar{\lambda}\Delta T(A\Delta T) - \dfrac{(\bar{\lambda}\Delta T)\,(A\Delta T)^2}{3}}{\bar{\lambda}\Delta T + \bar{\lambda}T_h(A\Delta T)} \tag{20}$$

The excess dispersion over the dispersion of the Poisson distribution is of the order $A\Delta T + 0(\Delta T)^2$ as $\Delta T \to 0$. So, as $\Delta T \to 0$, the dispersion approaches that of the Poisson distribution. In fact, the distribution of $\Delta N_1 + \Delta N_2$ tends to a Poisson distribution as $\Delta T \to 0$. To show this, assume that ΔT is small so that

$$\exp\{(e^{-s}-1)\,A(\Delta T)\} \approx 1 + (e^{-s}-1)\,A(\Delta T) + \frac{(e^{-s}-1)^2\,A^2(\Delta T)^2}{2} \tag{21}$$

Substitution of the approximation (21) into Eq. (17) yields

$$\phi(s) \approx \exp\left\{\bar{\lambda}(e^{-s}-1)\,(\Delta T) + \frac{\bar{\lambda}e^{-s}(e^{-s}-1)\,A(\Delta T)^2}{2}\right\}$$
$$\cdot \exp\left\{\frac{\bar{\lambda}(e^{-s}-1)\,A(\Delta T)^2}{2}\right\}$$
$$\cdot \exp\left\{\bar{\lambda}(e^{-s}-1)\,A(\Delta T)\,(T_h - \Delta T) + \frac{\bar{\lambda}(e^{-s}-1)\,(A\Delta T)^2\,(T_h - \Delta T)}{2}\right\}$$
$$= \exp\{[\bar{\lambda}\Delta T + \bar{\lambda}T_h(A\Delta T)]\,(e^{-s}-1)\}$$
$$\cdot \exp\left\{[\bar{\lambda}\Delta T\,(A\Delta T) + \bar{\lambda}(T_h - \Delta T)\,(A\Delta T)^2]\,\frac{(e^{-s}-1)^2}{2}\right\}$$
$$\approx \exp\{[\bar{\lambda}\Delta T + \bar{\lambda}T_h(A\Delta T)]\,(e^{-s}-1)\} \tag{22}$$

which is a Poisson distribution with mean $\bar{\lambda}\Delta T + \bar{\lambda}T_h(A\,\Delta T)$. (The last approximation in Eq. (22) drops terms of order $0((\Delta T)^2)$). The final approximation has the same mean (Eq. (18)) as $\phi(s)$.

Example 2: Constant Light Source and Exponential Rate of Trapped Carrier Induced Events

Consider when $\lambda(t)$ is again constant as given by Eq. (6) and

$$h(t) = (aA)\, e^{-t/(T_h/a)} \tag{23}$$

(see Fig. 2) where A and T_h can be taken to be the same parameters as those in Eq. (15) for comparisons with the rectangular $h(t)$ of Example 1. Since T_h/a is the time constant in Eq. (23), the case where $a \gg 1$ is of most interest for comparison to the rectangular case. Also note that the areas under Eqs. (15) and (23) are both equal to AT_h.

It is easy to show that

$$\int_T^{T+\Delta T} h(t-\tau)\,dt = \begin{cases} AT_h\left(1 - e^{-\frac{\Delta T}{(T_h/a)}}\right) e^{-(T-\tau)/(T_h/a)}, & 0 \leqslant \tau \leqslant T \\ AT_h\left(1 - e^{-(T+\Delta T-\tau)/(T_h/a)}\right), & T \leqslant \tau \leqslant T+\Delta T \end{cases} \tag{24}$$

Substitution of Eq. (24) into Eqs. (3) through (5) yields

$$\phi(s) = \exp\left\{\bar{\lambda}\int_0^T [\exp\{-K_1\, e^{-\beta(T-\tau)}\} - 1]\,d\tau\right\} \cdot \exp\left\{\bar{\lambda}\int_T^{T+\Delta T} [e^{-s}\exp\{-K_2 \cdot (1 - e^{-\beta(T+\Delta T-\tau)})\} - 1]\,d\tau\right\} \tag{25}$$

where

$$K_1 = (1 - e^{-s})(AT_h)\left(1 - e^{-\frac{\Delta T}{T_h/a}}\right) \tag{26}$$

$$K_2 = (1 - e^{-s})(AT_h) \tag{27}$$

$$\beta = (T_h/a)^{-1} \tag{28}$$

The integrals in Eq. (25) can be expressed in terms of variants of the Exponential Integral function (Ref. 7, Chapter 5). It can be shown that for $s \geqslant 0$,

$$\phi(s) = \exp\left\{\frac{\bar{\lambda}}{\beta}[-E_1(K_1) + E_1(K_1\, e^{-\beta T}) - \beta T]\right\} \cdot \exp\left\{\frac{\bar{\lambda}}{\beta}[e^{-(s+K_2)}\{Ei(K_2) - Ei(K_2\, e^{-\beta\Delta T})\} - \beta\Delta T]\right\} \tag{29}$$

and for $s < 0$,

$$\phi(s) = \exp\left\{\frac{\bar{\lambda}}{\beta}[Ei(-K_1) - Ei(-K_1\, e^{-\beta T}) - \beta T]\right\} \cdot \exp\left\{\frac{\bar{\lambda}}{\beta}[e^{-(s+K_2)}\{-E_1(-K_2) + E_1(-K_2\, e^{-\beta\Delta T})\} - \beta\Delta T]\right\} \tag{30}$$

where the Exponential Integral functions $E_1(\cdot)$ and $Ei(\,)$ are given in Ref. 7, Chapter 5. Finally substitution of Eq. (24) into Eqs. (10) and (11) yields

$$E[\Delta N_1 + \Delta N_2] = \bar{\lambda}\Delta T + \bar{\lambda}T_h(A\Delta T) - e^{-\frac{T}{(T_h/a)}}\left(1 - e^{-\frac{\Delta T}{(T_h/a)}}\right)\left(\frac{\bar{\lambda}AT_h^2}{a}\right) \tag{31}$$

$$\begin{aligned}\text{var}(\Delta N_1 + \Delta N_2) = {} & E[\Delta N_1 + \Delta N_2] \\ & + \frac{(\bar{\lambda}T_h)(AT_h)^2}{2a}\left(1 - e^{-\frac{\Delta T}{T_h/a}}\right)^2\left(1 - e^{-\frac{2T}{(T_h/a)}}\right) \\ & + \bar{\lambda}\Delta T(AT_h)^2 + 2\bar{\lambda}\Delta T(AT_h) \\ & - \frac{2\bar{\lambda}T_h(AT_h)(1+AT_h)}{a}\left(1 - e^{-\frac{\Delta T}{T_h/a}}\right) \\ & + \frac{(\bar{\lambda}T_h)(AT_h)^2}{2a}\left(1 - e^{-\frac{2\Delta T}{T_h/a}}\right)\end{aligned} \tag{32}$$

Note that as $T \to \infty$, the average count given by Eq. (31) approaches the average count (Eq. (18)) in the rectangular $h(t)$ case. As $\Delta T \to 0$, the variance approaches the mean and so the dispersion again tends to that of a Poisson distribution. In fact, the distribution of $\Delta N_1 + \Delta N_2$ again tends to a

Poisson distribution as $\Delta T \to 0$. This can be established by using the approximations

$$E_1(x) = -\gamma - \ln x + x$$

$$Ei(x) = \gamma + \ln x + x$$

for small $x > 0$ (where γ is Euler's constant) in Eqs. (29) and (30), assuming that $aA\Delta T \ll 1$ and $\Delta T/(T_h/a) \ll 1$ and $T \to \infty$. Under these assumptions it can be shown that

$$\phi(s) \approx \exp\{[\bar{\lambda}\Delta T + \bar{\lambda}T_h(A\Delta T)](e^{-s} - 1)\} \quad (33)$$

which is the same asymptotic Poisson distribution Eq. (22) as that in the rectangular $h(t)$ case when $\Delta T \to 0$. This is so because the primary events are Poisson distributed. So, given a fixed number of events in an interval, the occurrence times are uniformly distributed. This would then tend to average out the exponential $h(t)$ to produce an effective rectangular $h(t)$.

Example 3: Pulsed Light Source and Constant Rate of Trapped Carrier Induced Events

Consider a rectangular $h(t)$ given by Eq. (15) and a pulsed intensity rate $\lambda(t)$ given in Fig. 3. That is, $\lambda(t)$ is periodic of period T_p and pulse duration t_p and pulse height $\bar{\lambda}$. We assume that $T \geq \tau_p$ and $T + \Delta T \leq T_h$ and $T_p = \tau_p + T_h$. In this case, $\Delta N_1 + \Delta N_2$ will be the total count in any ΔT-second interval between pulses in steady state. So

$$\int_T^{T+\Delta T} h(t-\tau)\,dt = A\Delta T, \quad 0 \leq \tau \leq \tau_p$$

and hence Eqs. (3) through (5) yield

$$\phi(s) = \exp\{(\bar{\lambda}t_p)[\exp\{(e^{-s} - 1)A\Delta T\} - 1]\} \quad (34)$$

This is the mgf of a two-parameter Neyman Type-A distribution. Here

$$E[\Delta N_1 + \Delta N_2] = (\bar{\lambda}\tau_p)(A\Delta T) \quad (35)$$

$$\text{var}(\Delta N_1 + \Delta N_2) = (1 + A\Delta T)E[\Delta N_1 + \Delta N_2] \quad (36)$$

Again, as $\Delta T \to 0$, $\phi(s)$ approaches a Poisson mgf with the same mean (Eq. (35)).

Example 4: Pulsed Light Source and Exponential Rate of Trapped Carrier Induces Events

Consider $\lambda(t)$ given as in Fig. 3 and an exponential $h(t)$ given by Eq. (23). We assume here $T \geq \tau_p$ and that $T_p \gg (T_h/a)$, the time constant of the exponential $h(t)$. We also assume that $T + \Delta T \leq T_p$. Then, as shown for Eqs. (24) and (25), it can be shown that

$$\phi(s) = \exp\left\{\bar{\lambda}\int_{T-\tau_p}^{T}[\exp\{-K_1 e^{-\beta x}\} - 1]\,dx\right\} \quad (37)$$

where K_1 and β are given by Eqs. (26) and (28), respectively. The integral in Eq. (37) can be expressed in terms of Exponential Integral functions as in Eqs. (29) and (30).

In fact,

$$\phi(s) = \begin{cases} \exp\left\{\frac{\bar{\lambda}}{\beta}[-E_1(K_1 e^{-\beta(T-\tau_p)}) + E_1(K_1 e^{-\beta T}) - \beta\tau_p]\right\}, \\ \text{if } s \geq 0 \\ \exp\left\{\frac{\bar{\lambda}}{\beta}[Ei(-K_1 e^{-\beta(T-\tau_p)}) - Ei(-K_1 e^{-\beta T}) - \beta\tau_p]\right\}, \\ \text{if } s < 0 \end{cases} \quad (38)$$

It can also be shown that

$$E[\Delta N_1 + \Delta N_2] = \frac{(\bar{\lambda}T_h)(AT_h)}{a}\left(1 - e^{-\frac{\Delta T}{(T_h/a)}}\right)\left(e^{\frac{\tau_p}{T_h/a}} - 1\right)e^{-\frac{T}{(T_h/a)}} \quad (39)$$

and

$$\text{var}(\Delta N_1 + \Delta N_2) = E[\Delta N_1 + \Delta N_2] + \frac{\bar{\lambda}T_h(AT_h)^2}{2a}\left(1 - e^{-\frac{\Delta T}{T_h/a}}\right)^2\left(e^{\frac{2\tau_p}{(T_h/a)}} - 1\right)e^{-\frac{2T}{(T_h/a)}} \quad (40)$$

Note that as $\Delta T \to 0$, the variance approaches the mean. In fact, it can be shown that as $\Delta T \to 0$, $\phi(s)$ approaches a Poisson mgf with mean given by

$$(\bar{\lambda}T_h)(A\Delta T)\left(e^{\frac{\tau_p}{T_h/a}} - 1\right)e^{-\frac{T}{T_h/a}} \quad (41)$$

IV. Conclusions

In this article, moment generating functions of the output count statistics of a TAD-mode APD were derived under the assumptions that the primary avalanche events are Poisson distributed and the secondary trapped-carrier-induced avalanche events generated by each primary event are conditionally Poisson. The mean and variance of the distributions were also derived. Examples of constant light source and pulsed light source excitation and constant and exponential decaying rate of secondaries generated per primary were considered. It is shown that the dispersion of the distribution is always larger than that of the Poisson. In each of the examples, the distribution approached a Poisson distribution when the count interval tended to zero.

We should note that trapped carriers generated by a particular primary event were assumed to be unaffected by future avalanche events. It was also assumed that trapped-carrier-induced avalanche events do not themselves produce more trapped carriers. These simplifying assumptions were made in the interest of analytical tractability. Hence the usefulness of these results need to be established experimentally.

References

1. Ekstrom, P. A., "Triggered-Avalanche Detection of Optical Photons," *J. Applied Physics*, Vol. 52, No. 11, November 1981, pp. 6974-6979.
2. Ekstrom, P. A., "Single Photon Counters for the Infrared," *SPIE Instrumentation in Astronomy IV*, Vol. 331, 1982, pp. 2-8.
3. Ingerson, T. E., et. al., "Photon Counting with Photodiodes," *Applied Optics*, Vol. 22, No. 13, July 1983, pp. 2013-2018.
4. Snyder, D. L., *Random Point Processes*, Wiley, New York, 1975.
5. Saleh, B., and Teich, M. C., "Multiplied-Poisson Noise in Pulse, Particle and Photon Detection," *IEEE Proceedings*, Vol. 70, No. 3, March 1982, pp. 229-245.
6. Neyman, J., "On a New Class of Contagious Distributions, Applicable in Entomology and Bacteriology," *Annals Math. Stat.*, Vol. 10, 1939, pp. 35-57.
7. Alramowitz, M., and Stegun, A., *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, D.C. 1972.

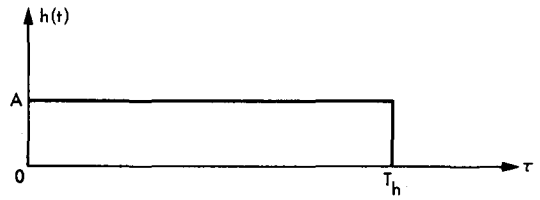


Fig. 1. Rectangular *h(t)*

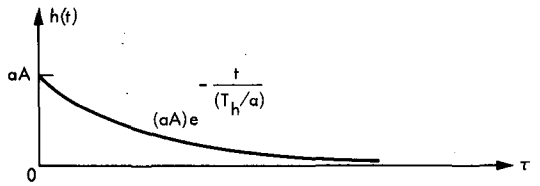


Fig. 2. Exponential *h(t)*

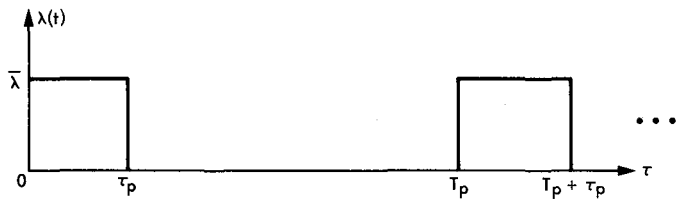


Fig. 3. Pulsed light λ(*t*)

Appendix A
Derivation of $\phi(s)$

First we note that

$$\begin{aligned}
\phi(s) &= \sum_{k=0}^{\infty} e^{-sk} P[\Delta N_1 + \Delta N_2 = k] \\
&= \sum_{k=0}^{\infty} e^{-sk} \sum_{i=0}^{\infty} \sum_{\ell=0}^{k} P[\Delta N_1 + \Delta N_2 = k, \Delta N_1 = \ell, N_1(T) = i] \\
&\overset{(1)}{=} \sum_{k=0}^{\infty} \sum_{\ell=0}^{k} e^{-sk} \left\{ \sum_{i=0}^{\infty} P[\Delta N_2 = k - \ell \mid \Delta N_1 = \ell, N_1(T) = i] \cdot P[\Delta N_1 = \ell] \cdot P[N_1(T) = i] \right\} \\
&\overset{(2)}{=} \sum_{j=0}^{\infty} \sum_{\ell=0}^{\infty} e^{-s(j+\ell)} \left\{ \sum_{i=0}^{\infty} P[\Delta N_2 = j \mid \Delta N_1 = \ell, N_1(T) = i] \cdot P[\Delta N_1 = \ell] \cdot P[N_1(T) = i] \right\} \\
&= \sum_{i=0}^{\infty} P[N_1(T) = i] \left[\sum_{\ell=0}^{\infty} e^{-s\ell} P[\Delta N_1 = \ell] \cdot \left\{ \sum_{j=0}^{\infty} e^{-sj} P[\Delta N_2 = j \mid \Delta N_1 = \ell, N_1(T) = i] \right\} \right]
\end{aligned} \tag{A-1}$$

where Subequation (1) is because $N_1(t)$ has independent increments and Subequation (2) is obtained by setting $j = k - 1$ and noting that

$$\sum_{k=0}^{\infty} \sum_{\ell=0}^{k} = \sum_{j=0}^{\infty} \sum_{\ell=0}^{\infty}$$

Next consider $i \geqslant 1$ and $\ell \geqslant 1$ and suppose that $N_1(T) = i$ and $\Delta N_1 = \ell$. Then let $t_1, \cdots, t_i$ be the primary event occurrence times in $[0, T)$ and $t_{i+1}, \cdots, t_{i+\ell}$ the occurrence times in $[T, T + \Delta T)$. Then

$$\begin{aligned}
\sum_{j=0}^{\infty} e^{-sj} P[\Delta N_2 = j \mid \Delta N_1 = \ell, N_1(T) = i] &= E_{t_1, \cdots, t_{i+\ell}} \left[\sum_{j=0}^{\infty} e^{-sj} \frac{\left(\sum_{m=1}^{i+\ell} \int_T^{T+\Delta T} h(t - t_m)\, dt \right)^j}{j!} \right. \\
&\quad \left. \cdot \exp\left(- \sum_{m=1}^{i+\ell} \int_T^{T+\Delta T} h(t - t_m)\, dt \right) \cdot \mid \Delta N_1 = \ell, N_1(T) = i \right] \\
&= E_{t_1, \cdots, t_{i+\ell}} \left[\prod_{m=1}^{i+\ell} \exp\left\{ (e^{-s} - 1) \int_T^{T+\Delta T} h(t - t_m)\, dt \right\} \mid \Delta N_1 = \ell, N_1(T) = i \right]
\end{aligned} \tag{A-2}$$

$$
\begin{aligned}
&\overset{(1)}{=} E_{t_1,\cdots,t_i}\left[\prod_{m=1}^{i}\exp\left\{(e^{-s}-1)\int_T^{T+\Delta T} h(t-t_m)\,dt\right\} \mid N_1(T)=i\right] \\
&\quad\cdot E_{t_{i+1},\cdots,t_{i+\ell}}\left[\prod_{m=i+1}^{i+\ell}\exp\left\{(e^{-s}-1)\int_T^{T+\Delta T} h(t-t_m)\,dt\right\} \mid \Delta N=\ell\right] \\
&\overset{(2)}{=} K^i\cdot H^\ell
\end{aligned}
\qquad \text{(A-2)}
$$

where

$$
K=\frac{\displaystyle\int_0^T \lambda(\tau)\exp\left[(e^{-s}-1)\int_T^{T+\Delta T} h(t-\tau)\,dt\right]d\tau}{\displaystyle\int_0^T \lambda(\tau)\,d\tau} \qquad \text{(A-3)}
$$

and

$$
H=\frac{\displaystyle\int_T^{T+\Delta T} \lambda(\tau)\exp\left[(e^{-s}-1)\int_T^{T+\Delta T} h(t-\tau)\,dt\right]d\tau}{\displaystyle\int_T^{T+\Delta T} \lambda(\tau)\,d\tau} \qquad \text{(A-4)}
$$

The independent increments property of $N_1(t)$ is used to obtain Subequation (1) in Eq. (A-2) and follows from known properties of the conditional distribution of occurrence times, given the number of events in a fixed time interval for a Poisson process (Ref. 4).

Next, we note from the derivation that the final result in Eq. (A-2) is valid when either $i = 0$ or $\ell = 0$ or both $i = \ell = 0$. Finally, substitution of Eqs. (A-2), (A-3), and (A-4) into Eq. (A-1) yields:

$$
\begin{aligned}
\phi(s) &= \left\{\sum_{i=0}^{\infty} K^2\,P[N_1(T)=i]\right\}\left\{\sum_{\ell=0}^{\infty}(He^{-s})^\ell\,P[\Delta N_1=\ell]\right\} \\
&= \left\{\sum_{i=0}^{\infty} K^i\frac{\left(\displaystyle\int_0^T\lambda(\tau)\,d\tau\right)^i}{i!}\exp\left(-\int_0^T\lambda(\tau)\,d\tau\right)\right\} \\
&\quad\cdot\left\{\sum_{\ell=0}^{\infty}(He^{-s})^\ell\frac{\left(\displaystyle\int_T^{T+\Delta T}\lambda(\tau)\,d\tau\right)^\ell}{\ell!}\exp\left(-\int_T^{T+\Delta T}\lambda(\tau)\,d\tau\right)\right\}
\end{aligned}
\qquad \text{(A-5)}
$$

$$
\left.
\begin{aligned}
&= \exp\left\{\int_0^T \lambda(\tau)\left[\exp\left\{(e^{-s}-1)\int_T^{T+\Delta T} h(t-\tau)\,dt\right\}-1\right]dt\right\} \\
&\cdot \exp\left\{\int_T^{T+\Delta T} \lambda(\tau)\left[e^{-s}\exp\left\{(e^{-s}-1)\int_T^{T+\Delta T} h(t-\tau)\,dt\right\}-1\right]d\tau\right\}
\end{aligned}
\right\} \quad \text{(A-5)}
$$

Design and Performance of Sampled Data Loops for Subcarrier and Carrier Tracking

S. Aguirre and W. J. Hurd
Communications Systems Research Section

Design parameters and resulting performance are presented for the sampled data analogies of continuous-time phase-locked loops of second and third order containing perfect integrators. Expressions for noise-equivalent bandwidth and steady-state errors are given. Then, stability and gain margin are investigated using z-plane root loci. Finally, an application is presented for Voyager subcarrier and carrier tracking under the dynamics of the encounters with Uranus and Neptune. For carrier tracking, loop bandwidths narrow enough for satisfactory loop SNR_s can be achieved using third-order loops without rate aiding, whereas second-order loops would require aiding. For subcarrier tracking, third-order loops can be used when the sampling rate is limited to approximately once per second, as in the Baseband Assembly, whereas second-order loops sufficiently wide to track the dynamics have stability problems at that sampling rate.

I. Introduction

Phase-locked loops for communications systems synchronization must satisfy requirements for noise performance, dynamic tracking, process or oscillator noise tracking, stability and gain margin, acquisition, and signal-to-noise ratio threshold. Digital technology enables the use of sampled data loops with programmable or even adaptively controlled parameters. These loops have significant advantages over continuous-time analog loops, but also have some characteristics requiring different design considerations.

Two major advantages of sampled data loops that enhance the dynamic tracking ability for a fixed-loop bandwidth are that perfect integrators can be realized, resulting in type-two second-order loops and type-three third-order loops when desired, and that third-order loops become more feasible because the parameters can be accurately controlled and varied with time. A major potential disadvantage of sampled data loops is that they are never unconditionally stable: high loop gains always result in instability due to the inherent transport lag. This is as opposed to first- and second-order continuous-time loops, which are normally unconditionally stable.

Stability problems with sampled data loops occur when the loop bandwidth is not sufficiently small compared to the sampling rate, i.e., the loop filter update rate. This case is of significant interest because the sampling rates for typical implementations are limited by mechanization considerations. This article investigates the stability and gain margin of second- and third-order loops when the loop bandwidth compared with sampling rate is wide enough that continuous loop

analysis is inadequate. Achievable loop bandwidths and steady-state phase-lag errors due to dynamics are discussed.

The Voyager II encounters with Uranus in January, 1986, and with Neptune in August, 1989, are of particular interest. The performance is presented for various loops for both subcarrier and carrier tracking under the dynamics at the encounters. It is shown that third-order loops result in excellent performance at both encounters, whereas second-order loops are marginal at Uranus and unsatisfactory at Neptune, for both carrier and subcarrier tracking.

II. Approach

The design of sampled data loops is more complex than that of continuous loops because the physical parameters of the loops cannot be calculated directly, as in the case of continuous loops, from the desired loop bandwidth, damping ratio, and, for third-order loops, the third-order gain parameter. Furthermore, stability must be investigated, because sampled data loops can be only conditionally stable. On the other hand, there is some simplification because perfect integrators can be used. This also enhances dynamic performance.

There is a folklore rule of thumb (Ref. 1) that sampled data loops perform much like the continuous-time equivalents whenever the loop bandwidth is less than one-tenth of the sampling rate. It is shown here that this is a good folklore regarding loop bandwidth, but not loop stability. For a nominal bandwidth of one-tenth of the sampling rate and for typical damping and transport lag, sampled data loops become unstable when the loop gain increases by only approximately 7 dB. This is true for both second- and third-order loops. The third-order loops also are unstable at low gain. Thus, stability should always be investigated for sampled data loops.

The approach taken here is to choose a nominal loop bandwidth as a fraction of sampling rate, and to pick other parameters based upon continuous time theory. Stability and gain margin are then investigated using z-plane root loci. Actual loop bandwidth and steady-state phase error due to dynamics can then be calculated as a function of loop gain for gains resulting in stability.

III. Review of Continuous Time Phase-Locked Loops

In this section we present a review of results widely used by designers of continuous-time phase-locked loops. Specifically, we review the expressions obtained from linear analysis of noise equivalent bandwidth and steady-state phase errors for perfect integrator second- and third-order loops.

Then, in Section IV, we discuss how to obtain a sampled data loop that contains perfect integrators, and whose gain parameters can be conveniently expressed in terms of the continuous-loop parameters r, $B_L T$, and k. We then calculate the noise equivalent bandwidth for the digital loops and the steady-state phase errors for specified input phase dynamics.

A. Second-Order Continuous Loop

The open-loop transfer function of the loop is written as $AKF(s)$, where AK is the loop gain and $F(s)$ is the loop filter transfer function. The parameter A is due to the phase detector and the signal amplitude, and K is selected by the designer and includes the VCO gain.

The perfect integrator (type 2) second-order continuous loop has a loop filter transfer function of the form (Refs. 2 and 3)

$$F(s) = \frac{1 + \tau_2 s}{\tau_1 s} \tag{1}$$

and typically $\tau_2 \ll \tau_1$. For a given loop gain AK, the loop damping parameter r is defined as

$$r = \frac{AK\,\tau_2^2}{\tau_1} \tag{2}$$

The closed-loop transfer function $H(s)$ is given by

$$H(s) = \frac{AKF(s)}{s + AKF(s)} \tag{3}$$

and its associated one-sided loop noise bandwidth B_L (in Hz) is

$$B_L = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(j\omega)|^2 \, d\omega \cong \frac{r+1}{4\,\tau_2} \tag{4}$$

Often in practice, the values $r = 2$ or 4 are selected since they correspond to damping ratios (from standard control theory notation) of $\zeta = 0.707$ or 1 respectively (note that $r = 4\zeta^2$).

Now consider tracking the carrier or subcarrier signal from a spacecraft with dynamics. The steady-state phase error for this loop due to an assumed instantaneous doppler of the form

$$d(t) = \frac{\omega_i}{c} (\Omega_0 + \Lambda_0 t) \tag{5}$$

where

ω_i = subcarrier (carrier) frequency (rad/s)

c = speed of light (m/s)

Ω_0 = spacecraft speed (m/s) at $t = 0$ (m/s)

Λ_0 = spacecraft acceleration (m/s^2)

can be calculated via the final-value theorem, and is

$$\phi_{ss} = \lim_{s\to 0} \frac{\omega_i}{c} \cdot s\,[1 - H(s)] \left[\frac{\Omega_0}{s^2} + \frac{\Lambda_0}{s^3}\right]$$

$$= \frac{\omega_i}{c} \cdot \frac{\Lambda_0}{r} \cdot \left(\frac{r+1}{4B_L}\right)^2 \tag{6}$$

B. Third-Order Continuous Loops

The loop filter and the corresponding closed-loop transfer function of a perfect third-order loop can be represented respectively as (Refs. 2 and 4)

$$F(s) = \frac{1 + \tau_2 s}{\tau_1 s} + \frac{1}{\tau_1 \tau_3 s^2} \tag{7}$$

$$H(s) = \frac{rk + r\tau_2 s + r(\tau_2 s)^2}{rk + r\tau_2 s + r(\tau_2 s)^2 + (\tau_2 s)^3} \tag{8}$$

where, as before, $r = AK\tau_2^2/\tau_1$ and with $k = \tau_2/\tau_3$. The corresponding one-sided loop bandwidth B_L (in Hz) is

$$B_L = \frac{r}{4\tau_2}\left(\frac{r - k + 1}{r - k}\right) \tag{9}$$

notice that if $k = 0$, then our loop reduces to a perfect integrator second-order loop.

When the spacecraft experiences jerk, the Laplace transform of the phase error is

$$\Phi(s) = \frac{\omega_i}{c}\,[1 - H(s)] \left[\frac{\Omega_0}{s^2} + \frac{\Lambda_0}{s^3} + \frac{J_0}{s^4}\right]$$

and the resulting steady-state phase error is

$$\phi_{ss} = \frac{\omega_i}{c}\,\frac{J_0}{rk}\left[\frac{r}{4B_L}\left(\frac{r - k + 1}{r - k}\right)\right]^3 \tag{10}$$

IV. Sampled Data Loop Noise and Dynamic Responses

A general block diagram for a sampled data loop is shown in Fig. 1. We resort to the z-transform technique because of the discrete nature of the system. In this linear model, θ_n symbolizes the value of the input phase at sampling instant t_n, and $\hat{\theta}_n$ its corresponding estimate. AK represents the loop gain, including the numerically controlled oscillator (NCO) gain, $F(z)$ is an arbitrary loop filter, and $N(z)$ is characteristic of the particular implementation, which includes the modeling of the NCO, phase detector, and computational delays. The open-loop transfer function is $G(z) = AKF(z)N(z)$. A factor $D(z)$ may be added as a cascaded compensator to enhance stability.

In this article we concentrate exclusively on the $N(z)$ that arises in the proposed implementation for the DSN Advanced Receiver (Ref. 5). In this case, the sampling interval (filter update interval) is of length T. Phase estimates are averages of $\theta - \hat{\theta}$ over one interval, and there is a transport lag of T seconds from the end of the measurement time until the NCO is updated. For this important case, $N(z)$ is given by

$$N(z) = \frac{T(z+1)}{2z^2(z-1)} \tag{11}$$

The derivation of this transfer function and extension to other transport lags is found in Appendix A.

Regarding the loop filter $F(z)$, we deal with the sampled data versions of filters comprising single and double integrators as defined by Eqs. (1) and (7).

A. Second-Order DPLL

The digital equivalent of the loop filter in the perfect second-order analog PLL has a transfer function of the form $F(z) = G_1 + G_2 Tz/(z - 1)$. To establish a correspondence between the digital PLL (DPLL) and the analog PLL, let $G_1 = \tau_2/\tau_1$ and $G_2 = 1/\tau_1$ where τ_1 and τ_2 are the time constants of the filter defined by Eq. (1). Then using Eqs. (2) and (4) we can express the open-loop transfer function as

$$G(z) = \frac{\frac{r}{2}(z+1)\left[\frac{4B_L T}{r+1}(z-1) + \left(\frac{4B_L T}{r+1}\right)^2 z\right]}{(z-1)^2 z^2} \tag{12}$$

Note that B_L is the noise bandwidth of the analog loop, not the bandwidth of the digital loop. The closed-loop transfer function $H(z) = G(z)/(1 + G(z))$ can be represented as

$$H(z) = \frac{b_2 z^2 + b_3 z + b_4}{a_0 Z^4 + a_1 z^3 + a_2 z^2 + a_3 z + a_4} \qquad (13)$$

where

$$b_2 = \frac{r}{2}(d^2 + d) \qquad a_2 = 1 + \frac{r}{2} d + \frac{r}{2} d^2$$

$$b_3 = \frac{r}{2} d^2 \qquad a_3 = \frac{r}{2} d^2$$

$$b_4 = -\frac{r}{2} d \qquad a_4 = -\frac{r}{2} d$$

$$a_0 = 1 \qquad d = \frac{4B_L T}{r+1}$$

$$a_1 = -2$$

Notice that this is actually a fourth-order loop in z, but we discuss it as a second-order loop because it is analogous to a type-2 second-order continuous loop.

1. Noise Equivalent Bandwidth of the Second-Order DPLL. For the DPLL, the one-sided noise equivalent bandwidth B_L^* (Hz) is given by

$$B_L^* \triangleq \frac{1}{2T} \frac{1}{H^2(1)} \frac{1}{2\pi j} \oint_{|z|=1} H(z) H(z^{-1}) \frac{dz}{z} \qquad (14)$$

where T is the update time in seconds, and $H^2(1) = 1$, regardless of r and $B_L T$.

Define

$$I_4 \triangleq \frac{1}{2\pi j} \oint_{|z|=1} H(z) H(z^{-1}) \frac{dz}{z} \qquad (15)$$

then from Table III in Ref. 6 we have

$$I_4 = \frac{a_0 B_0 Q_0 - a_0 B_1 Q_1 + a_0 B_2 Q_2 - a_0 B_3 Q_3 + B_4 Q_4}{a_0 [(a_0^2 - a_4^2) Q_0 - (a_0 a_1 - a_3 a_4) Q_1 + (a_0 a_2 - a_2 a_4) Q_2 - (a_0 a_3 - a_1 a_4) Q_3]} \qquad (16)$$

where

$$B_0 = b_0^2 + b_1^2 + b_2^2 + b_3^2 + b_4^2$$

$$B_1 = 2(b_0 b_1 + b_1 b_2 + b_2 b_3 + b_3 b_4)$$

$$B_2 = 2(b_0 b_2 + b_1 b_3 + b_2 b_4)$$

$$B_3 = 2(b_0 b_3 + b_1 b_4)$$

$$B_4 = 2 b_0 b_4$$

$$Q_0 = a_0 e_1 e_4 - a_0 a_3 e_2 + a_4 (a_1 e_2 - e_3 e_4)$$

$$Q_1 = a_0 a_1 e_4 - a_0 a_2 a_3 + a_4 (a_1 a_2 - a_3 e_4)$$

$$Q_2 = a_0 a_1 e_2 - a_0 a_2 e_1 + a_4 (a_2 e_3 - a_3 e_2)$$

$$Q_3 = a_1 (a_1 e_2 - e_3 e_4) - a_2 (a_1 e_1 - a_3 e_3) + a_3 (e_1 e_4 - a_3 e_2)$$

$$Q_4 = -a_4 Q_0 + a_3 Q_1 - a_2 Q_2 + a_1 Q_3$$

$$e_1 = a_0 + a_2$$

$$e_2 = a_1 + a_3$$

$$e_3 = a_2 + a_4$$

$$e_4 = a_0 + a_4$$

$$e_5 = a_0 + a_2 + a_4$$

The integral I_4 can also be evaluated using the residue theorem, i.e.,

$$I_4 = \sum [\text{residues of } H(z) H(z^{-1}) z^{-1} \text{ at the poles inside the unit circle}] \qquad (17)$$

For a single pole of $H(z)H(z^{-1})z^{-1}$ at $z = a$, the residue is simply

$$\text{residue} = \lim_{z \to a} (z - a) H(z) H(z^{-1}) z^{-1}$$

while for a pole at $z = a$ of multiplicity k, the residue is

$$\text{residue} = \lim_{z \to a} \frac{1}{(k-1)!} \frac{d^{k-1}}{dz^{k-1}} \cdot \{(z - a)^k H(z) H(z^{-1}) z^{-1}\}$$

this procedure requires a numerical solution because of the order of the polynomials involved.

A comparison of B_L^* with B_L is shown in Fig. 2 where the value of the parameter k is zero for the second-order loop. The two bandwidths are very close for $B_L T < 0.1$.

2. Dynamic Response of Second-Order DPLL. Under the assumption of linearity, the phase error (no noise) in the z-domain is given by the following expression

$$\Phi(z) = [1 - H(z)]\,\Theta(z) \tag{18}$$

where $H(z)$ is the closed-loop transfer function given by Eq. (13) and $\theta(z)$ is the z-transform of the phase input. If we assume an instantaneous doppler characterized by Eq. (5), and apply the final value theorem to our phase-error equation, we get

$$\phi_{ss} = \frac{\omega_i}{c} \lim_{z \to 1} \left(\frac{z-1}{z}\right) [1 - H(z)]$$

$$\left[\frac{\Omega_0 T z}{(z-1)^2} + \frac{\Lambda_0 T^2 z\,(z+1)}{2\,(z-1)^3}\right]$$

$$= \frac{\omega_i}{c}\,\frac{\Lambda_0 T^2}{r}\left(\frac{r+1}{4B_L T}\right)^2 \tag{19}$$

This is exactly the same result obtained for continuous loops, Eq. (6). Thus, sampling and $N(z)$ do not affect the steady-state response. Again, note that B_L is the bandwidth of the continuous loop.

B. Third-Order DPLL

We now consider an extension of the loop filter employed in the second-order DPLL. We add an additional perfect integrator, so that our new filter has a transfer function of the form $F(z) = G_1 + G_2 Tz/(z-1) + G_3 T^2 z^2 (z-1)^2$. As before, we let $G_1 = \tau_2/\tau_1$, $G_2 = 1/\tau_1$, and set the new constant $G_3 = 1/\tau_1\tau_3$. The closed-loop transfer function $H(z)$ in terms of the analog parameters r, k, and $B_L T$ involves the manipulation of Eq. (9) along with $r = AK\tau_2^2/\tau_1$ and $k = \tau_2/\tau_3$. We obtain

$$H(z) = \frac{b_2 z^3 + b_3 z^2 + b_4 z + b_5}{a_0 z^5 + a_1 z^4 + a_2 z^3 + a_3 z^2 + a_4 z + a_5} \tag{20}$$

where

$$b_2 = \frac{r}{2}(d + d^2 + kd^3)$$

$$b_3 = \frac{r}{2}(kd^3 - d)$$

$$b_4 = -\frac{r}{2}(d^2 + d)$$

$$b_5 = \frac{r}{2}d$$

$$a_0 = 1$$

$$a_1 = -3$$

$$a_2 = 3 + \frac{r}{2}(d + d^2 + kd^3)$$

$$a_3 = -1 + \frac{r}{2}(kd^3 - d)$$

$$a_4 = -\frac{r}{2}(d^2 + d)$$

$$a_5 = \frac{r}{2}d$$

$$d = \frac{4B_L T}{r} \cdot \frac{r-k}{r-k+1}$$

Notice too, that this is actually a fifth-order loop, but it is analogous to the third-order, type-3 continuous loop.

1. Noise Equivalent Bandwidth of the Third-Order DPLL. The one-sided noise equivalent bandwidth B_L^* (Hz) can be written as

$$B_L^* = \frac{1}{2TH^2(1)} \cdot I_5 \tag{21}$$

with I_5 defined in Eq. (15). Again, $H^2(1) = 1$ regardless of r, $B_L T$, and k. Unfortunately, no closed form solution is readily available for the evaluation of I_5. Nevertheless, from Ref. 7, I_5 can be obtained numerically, by solving the following set of simultaneous equations:

$$\sum_{i=0}^{5} (a_{i-r} + a_{i+r}) M_i = B_r$$

$$r = 0, 1, 2 \cdots 5$$

$$M_0 = a_0 I_5 \tag{22}$$

$$B_r = \begin{cases} \sum_{i=0}^{5} b_i^2, & r = 0 \\ 2 \sum_{i=0}^{5-r} b_i b_{i+r}, & r = 1, 2 \cdots 5 \end{cases}$$

where the coefficients a_i and b_i are as shown in Eq. (20). I_5 can also be evaluated using the residue theorem, as outlined at the end of Section A-1.

In Fig. 2 we summarize the numerical results obtained with Eq. (22). For convenience, we plot the normalized product $B_L^* T$ (digital) versus $B_L T$ (analog) for several cases of interest, including $k = 0$, which corresponds to the second-order DPLL. Notice that for values of $B_L T < 0.07$, the resulting $B_L^* T$ is insensitive to either r or k and the digital system resembles the analog system, with $B_L^* T \cong B_L T$. On the other hand, when $B_L T > 0.1$, the ratio of $B_L^* T$ to $B_L T$ increases rapidly with $B_L T$, suggesting that the system is approaching the instability point. This point occurs when $B_L T \cong 0.25$ for all values of r and k. These results are also confirmed later, using the root locus technique.

2. **Dynamic Response of the Third-Order DPLL.** The linear loop equation for the phase error in the absence of noise is given by Eq. (18), where $H(z)$ is now given in Eq. (20), and

$$\Theta(z) = \frac{\omega_i}{c}\left[\frac{\Omega_0 T z}{(z-1)^2} + \frac{\Lambda_0 T^2 z(z+1)}{2(z-1)^3} + \frac{J_0 T^3 z(z^2+4z+1)}{6(z-1)^4}\right]$$

is the z-transform of $\omega_i/c(\Omega_0 t + 1/2\Lambda_0 t^2 + 1/6 J_0 t^3)$, which assumes that a jerk is present.

Using the final value theorem, the static phase error (in radians) is

$$\phi_{ss} = \frac{\omega_i}{c} \cdot \frac{J_0 T^3}{rk}\left[\frac{r(r-k+1)}{4B_L T(r-k)}\right]^3 \text{(rad)} \tag{23}$$

This result is exactly the one obtained for a perfect third-order continuous loop (see Eq. (10)). Note that B_L is the bandwidth of the continuous loop.

V. Stability of Sampled Data Loops

A basic requirement of a control system is that it must be stable under all operating conditions. It is important for the loop to be stable not only at the design point, but at a region of parameter values around the design point as well. This assures stability for variations in the parameters. We utilize the concept of gain margin to determine the region of stability. Gain margin is the ratio of the maximum (or minimum) loop gain for stability to the design-point loop gain. We use the root locus plot to determine stability and gain margin.

The root locus is a pictorial representation of the poles of the closed-loop transfer function as a function of the loop gain. The plot starts at zero gain on the open-loop poles and terminates at infinite gain on the open-loop zeros. For a sampled data loop, the plot is drawn in the z-plane, and the stable region is the interior of the unit circle.

The approach taken here is to choose nominal values for r, $B_L T$, and k, and then to plot the root locus as a function of loop gain. We then evaluate the system gain margin. For second- and third-order loops, note that loop gain is proportional to the parameter r.

A. Stability First-Order DPLL

For a first-sampled data loop with $F(z) = 1$, and with the $N(z)$ of Eq. (11), the closed-loop transfer function is

$$H(z) = \frac{AKT(z+1)}{2z^3 - 2z^2 + AKTZ + AKT} \tag{24}$$

The root locus plot is shown in Fig. 3.

In contrast with a first-order analog loop, which is unconditionally stable, the digital loop becomes unstable for high-loop gains. By using the Routh-Hurwitz criterion, we can determine exactly the minimum value of $r(r_{osc})$ that produces instability. The result is

$$(AK)_{osc} = \frac{\sqrt{8}-2}{T} \simeq \frac{0.828}{T} \tag{25}$$

For continuous loops, we know that (Ref. 2)

$$B_L = \frac{AK}{4} \tag{26}$$

Defining gain margin (gm) as the ratio r_{osc}/r, we obtain after combining Eqs. (25) and (26)

$$gm \simeq \frac{0.21}{B_L T} \tag{27}$$

B. Stability of Second-Order DPLL

In Fig. 4 we show a typical example of the root locus plot for the second-order sampled data loop. In this example, the design point bandwidth is $(B_L T)_0 = 0.10$, and the design point r, denoted r_0, is 4. The main point of the plot is that the loop is stable at low gains and unstable at high gains. This is also true for other nomimal parameters $(B_L T)_0$ and r_0. In this example, there are two underdamped (complex) roots and two overdamped (real) roots for both small and large r, but for $2.55 \lesssim r \lesssim 2.75$, all the roots are real. For some values of $(B_L T)_0$ and r_0, there are two complex roots for all r.

For the first-order loop we found that the gain margin is inversely proportional to $B_L T$. Finding an exact expression similar to Eq. (26) using the Routh-Hurwitz criterion or Jury's test is difficult because of the order of the polynomials involved. Nevertheless, by generating a large number of root loci for different nominal values, we found that

$$gm \simeq \frac{0.25}{B_L T} \tag{28}$$

which is valid when $r = r_0$ (actual design point equal to the nominal point) and $B_L T < 0.10$. Notice that this gain margin is slightly larger than that of the first-order loop.

C. Stability of Third-Order DPLL

In Fig. 5 we present three typical examples for the third-order loop. They show that, in general, the third-order DPLL is unstable for both low and high gains. At low gains, instability occurs if $r < k$, while for high gains, instability occurs if $r \gtrsim k + 0.25 r_0/(B_L T)_0$.

Thus, the "high-gain" margin of the third-order DPLL is approximately the same for the first- and second-order DPLLs. For third-order loops, one also needs to consider the "low-gain" margin, or the amount by which the gain can be reduced before instability occurs. This "low-gain" margin is approximately k/r_0. Reduction of k improves "low-gain" margin at the expense of steady-state phase errors due to jerk (Eq. (23)). This follows intuition, since $k = 0$ corresponds to a second-order loop.

D. Stability Improvement by Compensation

For critical applications, typically when $B_L T$ is not sufficiently small to have adequate gain margins, some improvement in stability may be achieved by lead, lag, or tuned compensation networks $D(z)$. Study of loops with compensation is outside the scope of this article.

VI. Application to Voyager

In this section, we determine the loop-bandwidth requirements for second- and third-order PLLs for dynamic tracking at the Voyager II encounters with Uranus and Neptune. Our criteria for good dynamic tracking is that the steady-state phase error be 1 degree or less. For DPLLs, the sampling rate should be greater than 10 B_L to have an adequate gain margin of approximately 7 dB. In this case, $B_L^* \cong B_L$, so the bandwidths are approximately the same for continuous and sampled data loops.

A. Bandwidth Requirements for Second-Order PPLs

From Eq. (19), for second-order loops the B_L required for a given steady-state phase error due to acceleration is

$$B_L = \frac{r+1}{4} \frac{1}{\sqrt{r}} \Lambda_0^{1/2} \sqrt{\frac{\omega_i}{c} \cdot \frac{1}{\phi_{ss}}} \tag{29}$$

For example, for the Voyager subcarrier frequency of 360 kHz and a steady-state error of 1°, Eq. (29) simplifies to

$$B_L = 0.164 \frac{r+1}{\sqrt{r}} \Lambda_0^{1/2}$$

In Table 1 we present the required minimum second-order loop bandwidth for Voyager subcarrier and carrier tracking (at 8.4 GHz) in order to achieve 1° steady-state phase error at the encounters. We assume acceleration values of $\Lambda_0 = 0.32$ m/s² and 4 m/s² for Uranus and Neptune, respectively.

B. Bandwidth Requirements for Third-Order PLLs

From Eq. (23), for third-order loops the B_L required for a given steady-state phase error due to jerk is

$$B_L = \frac{r}{4} \left(\frac{r-k+1}{r-k} \right) \left(\frac{\omega_i}{c} \cdot \frac{J_0}{rk\phi_{ss}} \right)^{1/3} \tag{31}$$

For 1° steady-state error when the spacecraft experiences jerk, and for the 360-kHz subcarrier frequency,

$$B_L = 0.189 \frac{r}{(rk)^{1/3}} \cdot \left(\frac{r-k+1}{r-k} \right) J_0^{1/3}$$

In Fig. 6 we plot the normalized ratio $B_L/J_0{}^{1/3}$ for different values of r and k. Notice that for fixed k the required bandwidth increases as r increases. Table 2 summarizes the bandwidth requirements for third-order loops at encounters. We

assume jerk values of $J_0 = 0.83 \times 10^{-4}$ m/s^3 and 0.29×10^{-2} m/s^3 for Uranus and Neptune, respectively.

The reader should not conclude from Fig. 6 that small values of r are optimum, considering all effects. Specifically, the problem of acquisition must be considered. This will be done in a future report.

C. Conclusions for Subcarrier Tracking

For subcarrier tracking, there appear to be no fundamental problems with either second- or third-order DPLLs. The widest bandwidth required to track at Neptune with a second-order loop is 0.82 Hz, which results in an adequate loop SNR of 39 dB under weak signal conditions of symbol SNR of 0 dB and symbol rate of 20 ksymbols/s. For the existing Baseband Assembly (BBA) Demodulator Synchronizer (DSA), the only potential problem is the sampling rate. The DSA sampling rate is currently limited to slightly over 1/s. With this sampling rate, use of third-order loops is indicated for both encounters.

D. Conclusions for Carrier Tracking

For carrier tracking, loop SNR is a serious problem, and it is important to use the narrowest loop bandwidth that is consistent with both dynamic tracking and with tracking of the oscillator instabilities.

With the current DSN receivers, the Voyager spacecraft is tracked in the cruise mode with threshold bandwidths of $B_L = 6$ Hz. The actual B_L depends on SNR, but is on the order of 10 Hz. Thus, it is known that 10-Hz bandwidths are satisfactory for tracking oscillator instabilities.

Dynamic tracking at the encounters with second-order loops require bandwidths of 30 to 130 Hz. With third-order loops, the dynamic tracking limitation is only 1 to 4 Hz. Thus, use of third-order loops is indicated. Considering both dynamics and oscillator instabilities, a bandwidth of approximately $B_L = 10$ Hz may be appropriate. The loop must be implemented with an adequate sampling rate of at least 100/s.

References

1. Lindsey, W.C., 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. Holmes, J.K., *Coherent Spread Spectrum Systems*, John Wiley & Sons Inc., New York, 1982.
3. Tausworthe, R.C., *Theory and Practical Design of Phase-Locked Receivers*, JPL Technical Report No. 32-819. Jet Propulsion Laboratory, Pasadena, Calif., Feb. 1966.
4. Tausworthe, R.C., and R.B. Crow, *Practical Design of Third-Order Phase-Locked Loops*, JPL Technical Report 900-450. Jet Propulsion Laboratory, Pasadena, Calif., April 27, 1971.
5. Sfeir, R., S. Aguirre, and W.J. Hurd, "Coherent Digital Demodulation of a Residual Carrier Signal Using IF Sampling", *TDA Progress Report 42-78*, pp. 135-142. Jet Propulsion Laboratory, Pasadena, Calif., April 1984.
6. Jury, E.I., *Theory and Application of the Z-Transform Method*, Wiley, New York, 1964.
7. Winkelstein, R.A., "Closed Form Evaluation of Symmetric Two-Sided Complex Integrals", *TDA Progress Report 42-65*, pp. 133-144. Jet Propulsion Laboratory, Pasadena, Calif., July 1981.

Table 1. Second-order PLL bandwidths for 1° steady-state phase error at encounters

r_0	Bandwidth B_L, Hz			
	Subcarrier		Carrier	
	Uranus	Neptune	Uranus	Neptune
2	0.20	0.70	30	106
4	0.23	0.82	35	126

Table 2. Third-order PLL bandwidths for 1° steady-state phase error at encounters

r_0	k	Bandwidth B_L, Hz			
		Subcarrier		Carrier	
		Uranus	Neptune	Uranus	Neptune
2	1/4	0.033	0.11	0.94	3.0
	1/3	0.030	0.10	0.86	2.8
4	1/4	0.042	0.14	1.2	3.9
	1/3	0.038	0.13	1.1	3.6

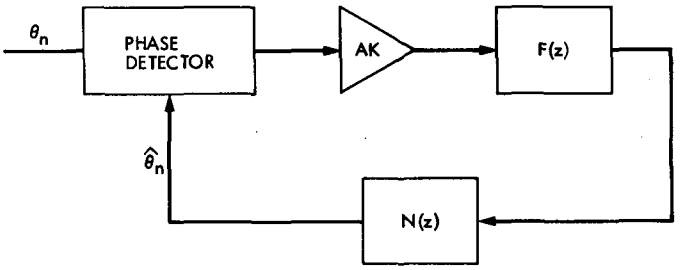


Fig. 1. Linearized model of the sampled data loop

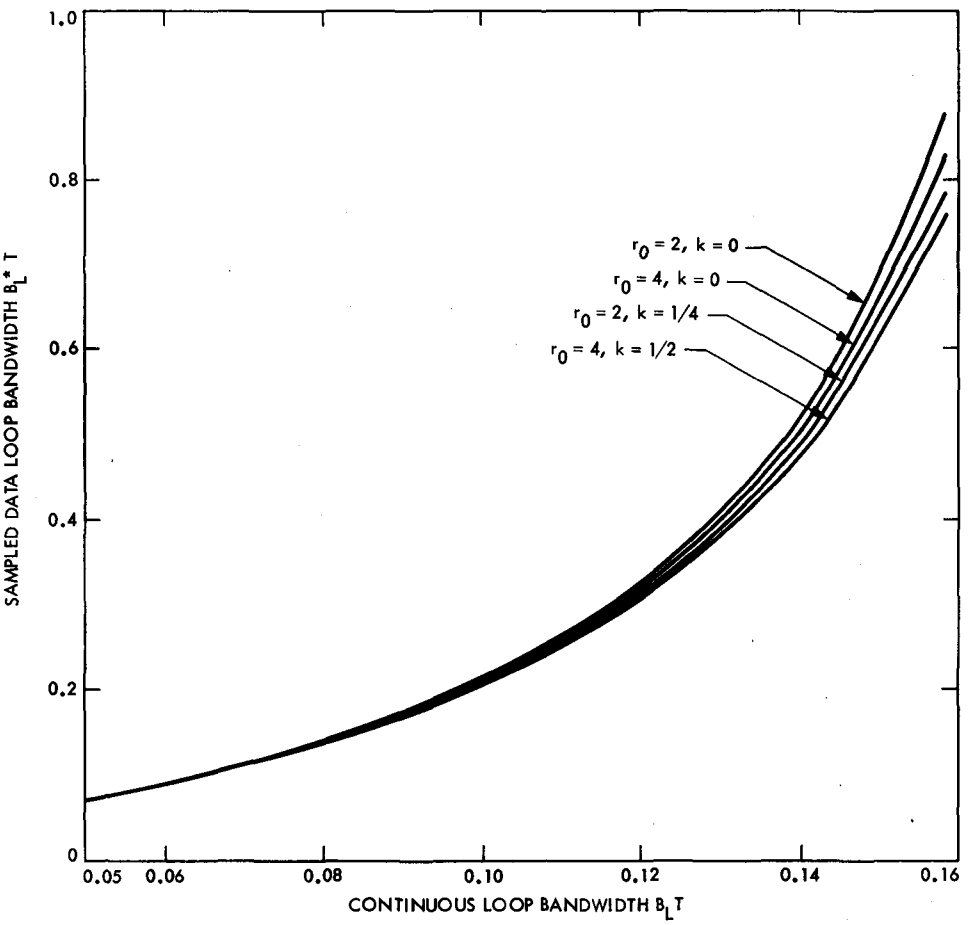


Fig. 2. Normalized bandwidth of the sampled data loop

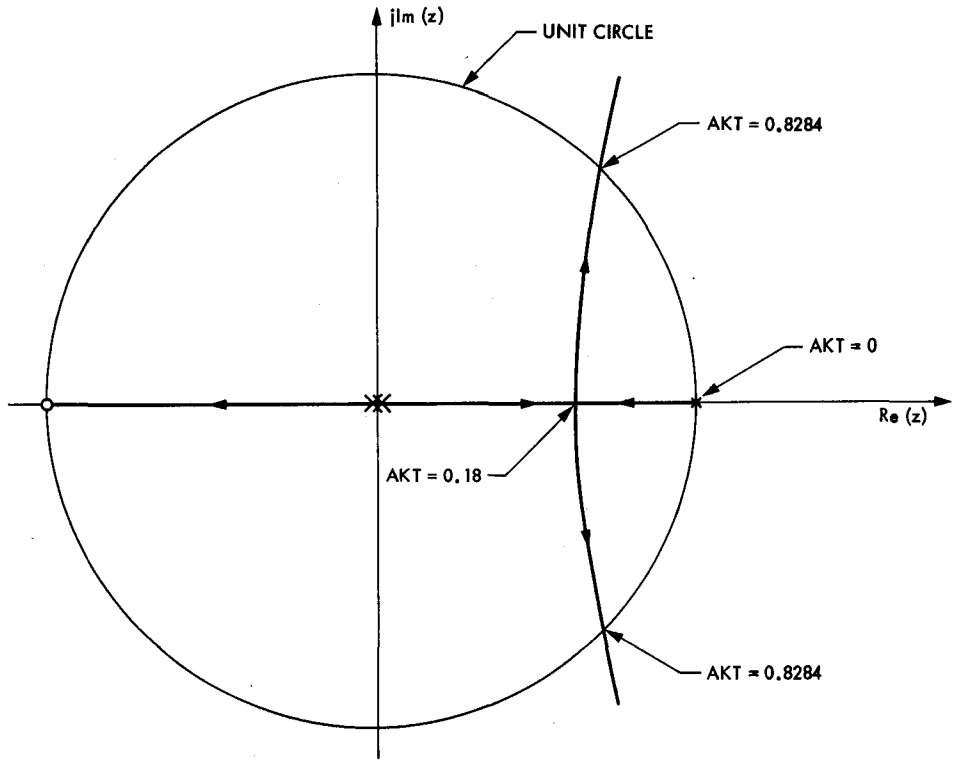


Fig. 3. Root locus of the first-order DPLL

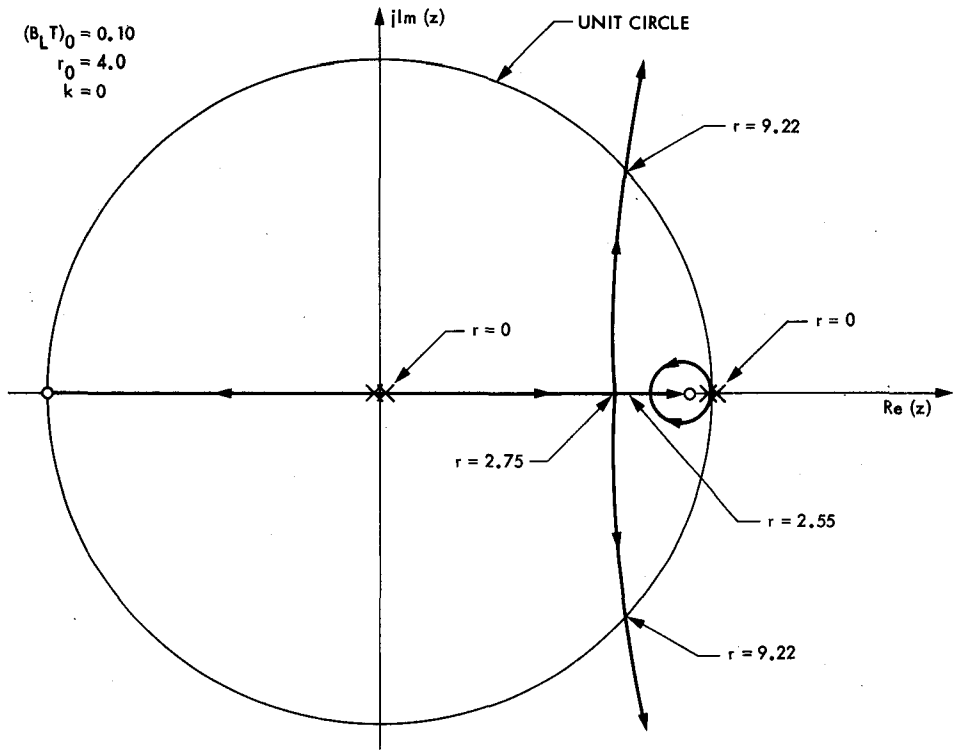


Fig. 4. Root locus of the second-order DPLL

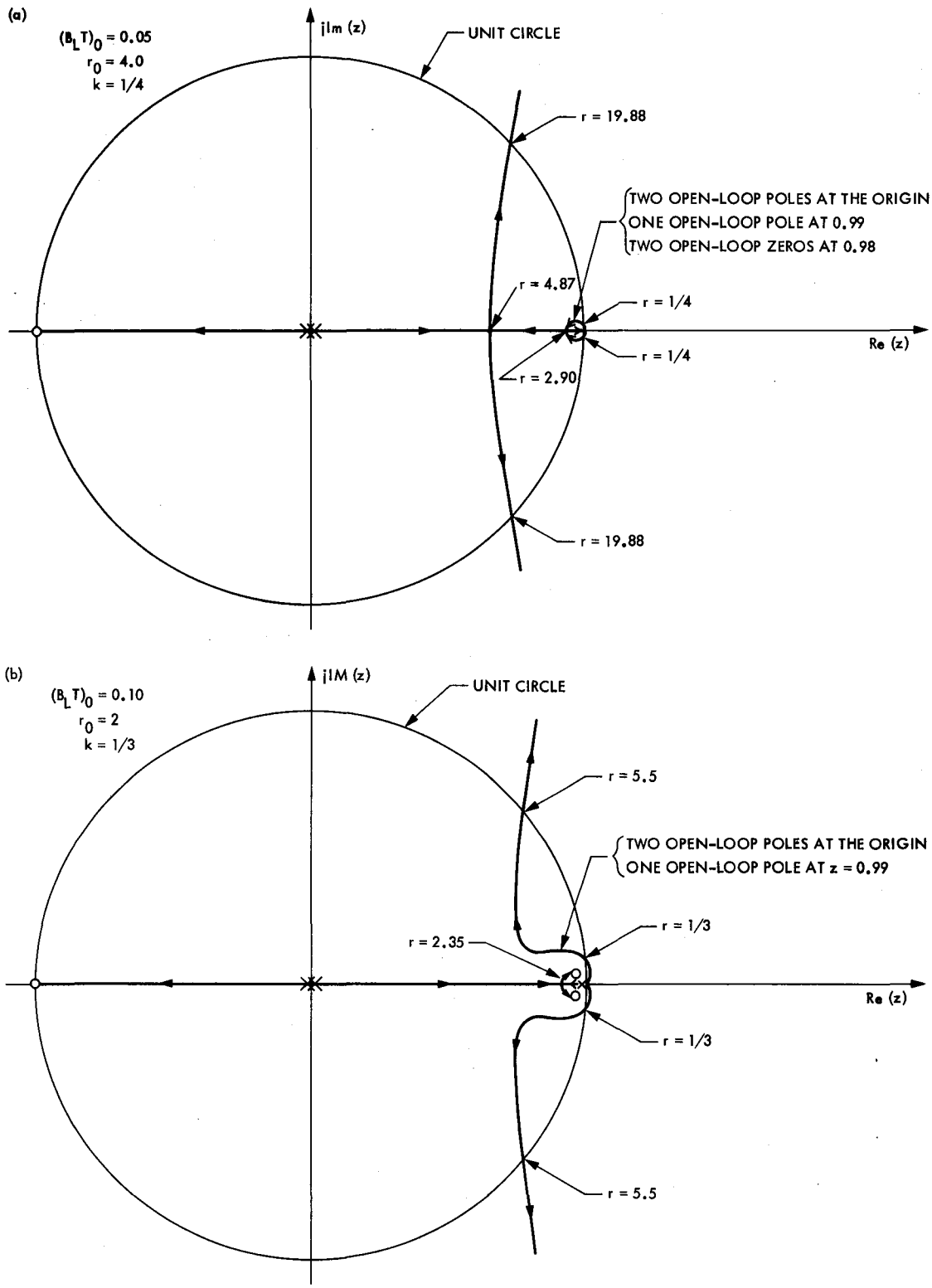


Fig. 5. Root loci of the third-order DPLL: (a) narrow bandwidth; (b) wide bandwidth, underclamped; (c) wide bandwidth with region of real roots

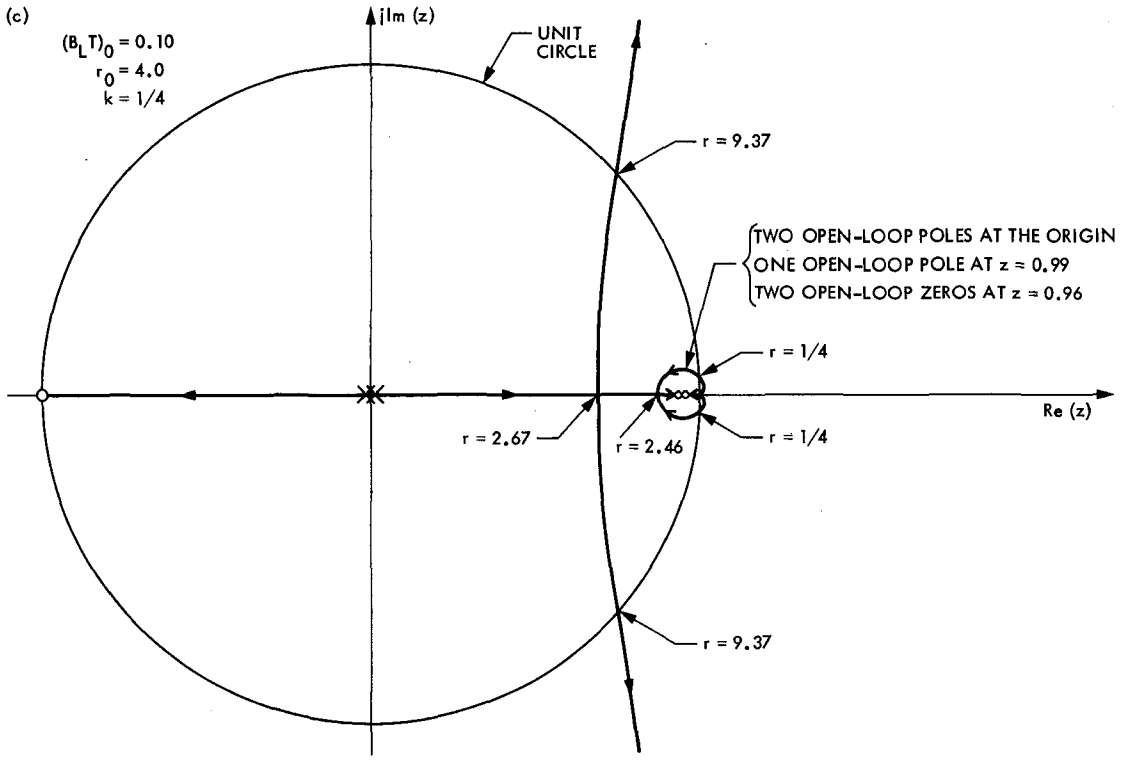


Fig. 5 (contd)

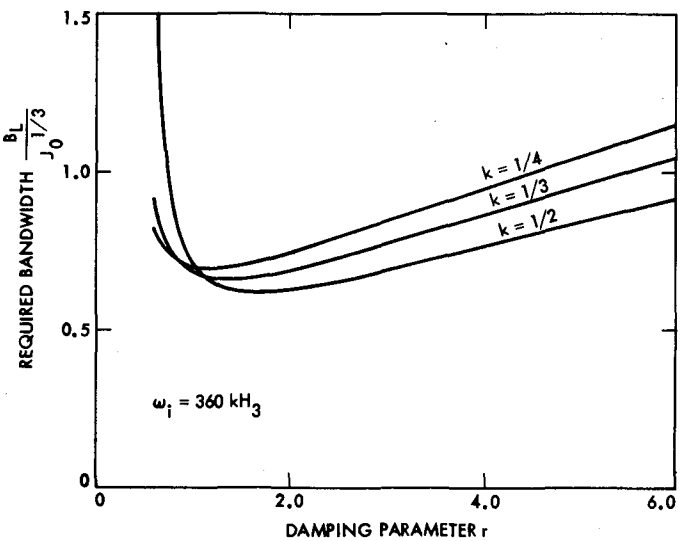


Fig. 6. Bandwidth requirements of third-order loops for 1° steady-state error

Appendix A
Derivation of N(z)

Figure A-1 is a block diagram of a DPLL with computation delay gT, i.e., for the case when the NCO input is updated gT seconds after the phase measurement is made. At the end of an integration interval of T seconds, the phase detector renders the difference of the average input phase and the average NCO phase.

Denote by $t_n = nT$ the sampling time, and by y_n the output of the NCO. We represent y_n with a linear piecewise function (see Fig. A-2) for g from 0 to 1,

$$y(t) = \begin{cases} y_{n-1+g} + x_{n-1} \cdot (t - t_{n-1+g}) & t_{n-1+g} \leqslant t \leqslant t_{n+g} \\ y_{n+g} + x_n \cdot (t - t_{n+g}) & t_{n+g} \leqslant t \leqslant t_{n+1+g} \end{cases} \tag{A-1}$$

the feedback function $\hat{\theta}_{n+1}$ is the averaged phase output over the interval from t_n to t_{n+1}, i.e.,

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \frac{1}{T}\int_{t_n}^{t_{n+1}} y(\tau)\, d\tau \tag{A-2}$$

Performing the integration for intervals n and $n+1$, we obtain

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \frac{1}{2}T[(1-g)^2 x_n + (1 + 2g - 2g^2) x_{n-1} + g^2 x_{n-2}] \tag{A-3}$$

Transforming to the z-domain, $N(z)$ is

$$N(z) \triangleq \frac{\hat{\theta}(z)}{x(z)} = \frac{T[(1-g)^2 z^2 + (1 + 2g - 2g^2)z + g^2]}{2z^2 (z-1)} \tag{A-4}$$

For the advanced receiver implementation $g = 1$,

$$N(z) = \frac{T(z+1)}{2z^2 (z-1)} \tag{A-5}$$

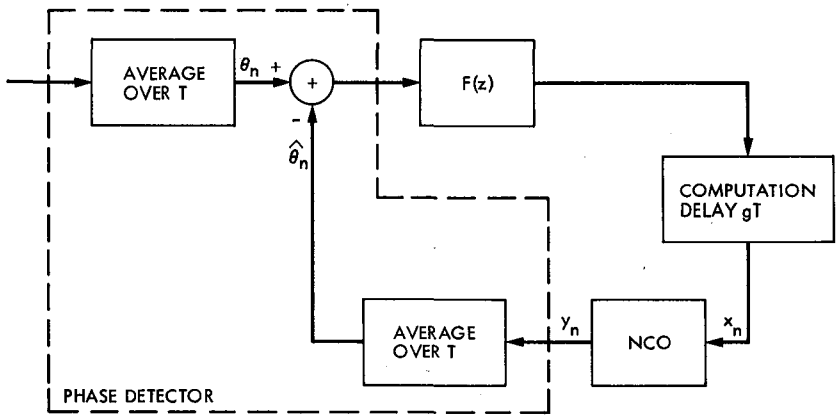


Fig. A-1. DPLL with computation delay *gT*

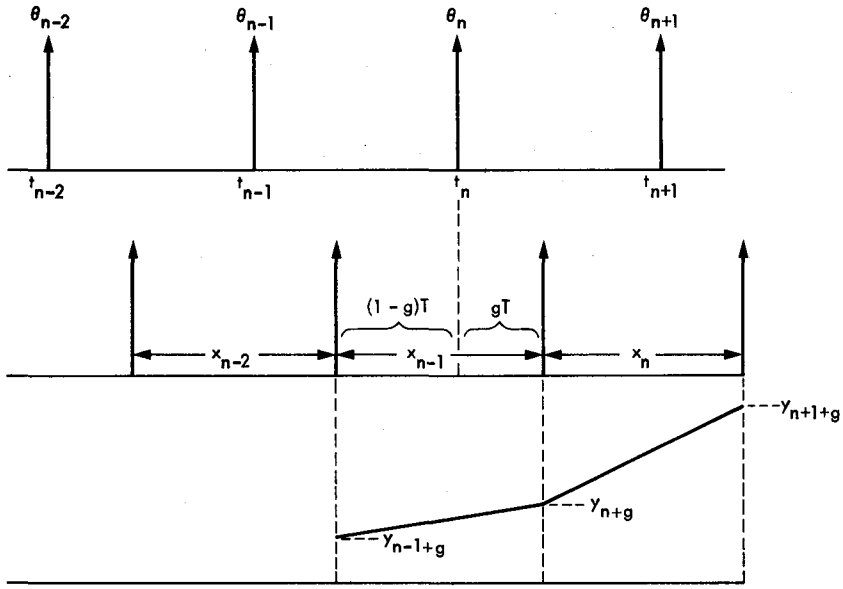


Fig. A-2. Timing for DPLL with computation delay *gT*

Improved Carrier Tracking by Smoothing Estimators

C. A. Pomalaza Raez
Clarkson University

W. J. Hurd
Communications Systems Research Section

Smoothing as a way to improve the carrier phase estimation is proposed and analyzed. The performance of first- and second-order Kalman optimum smoothers are investigated. This performance is evaluated in terms of steady-state covariance error computation, dynamic tracking, and noise response. It is shown that with practical amounts of memory, a second-order smoother can have a position error due to an acceleration or jerk step input less than any prescribed maximum. As an example of importance to the NASA Deep Space Network, a second-order smoother can be used to track the Voyager spacecraft at Uranus and Neptune encounters with significantly better performance than a second-order phase-locked loop.

I. Introduction

Carrier tracking (carrier phase estimation) is traditionally accomplished using phase-locked loops (PLLs), which may be residual carrier tracking loops, Costas loops, or sideband-aided loops (combination of residual carrier tracking and Costas loops). All these loops are causal, i.e., only past and present data are used to estimate phase at the present time. It is reasonable to expect an improved estimation by using future data, i.e., by using noncausal filters. Implementing this kind of estimator requires storage of the data for the length of time necessary to acquire the future data. This is now feasible for DSN carrier tracking, using memory densities available with current technology.

Linear estimators are classified as performing prediction, filtering, or smoothing according to whether the parameter estimates for the present time are based on past data only, past and present data, or past, present, and future data. PLLs are causal; analog loops normally perform filtering, and digital or sampled data loops perform prediction, because of the transport lag inherent in the sampled data implementation.

The sources of tracking errors for a given estimator can be classified as those due to observation noise, modeled and unmodeled state noise, and unmodeled dynamics. In a DSN receiver, the observation noise is primarily due to receiving system noise; the state noise is due to the oscillator, transmitter, receiver, and propagation media instabilities; and the unmodeled dynamics are due to spacecraft acceleration, Earth rotation, or to the error in modeling these effects. By using a smoothing estimator, one might intuitively expect to reduce the phase-error variance due to observation noise

by a factor of two, perhaps by using future data in a similar manner to past data. As shown later, this is true for first-order systems, but significantly more improvement is possible for higher-order systems. Finally, some cases of potential interest to the DSN advanced receiver are considered in this paper.

II. Practical Realization of the Smoothing Estimator

The proposed approach uses a Kalman estimation procedure as a method to extend the filter solution to a smoothing technique. Since the assumptions made to derive the Kalman solution may be different from the real case, it is necessary to evaluate the performance of the estimators for other cases. In this paper, in addition to the steady-state Kalman solution, a steady-state dynamic response and noise analysis are also presented.

Figure 1 shows a block diagram of a possible implementation of a smoothing estimator. The implementation shown is for a suppressed carrier signal with Costas-type phase detection. Residual carrier systems are similar except for the phase detection. The signal $r(t)$ is

$$r(t) = D(t)\cos(\omega_c t + \phi_c(t)) + v'(t) \quad (1)$$

where

$D(t)$ = binary data modulation

$\phi_c(t)$ = carrier phase

$v'(t)$ = narrowband white noise process

ω_c = received frequency

Also in Fig. 1,

$\theta(t)$ = phase error, $\theta(t) = \phi_c(t) - \phi_0(t)$

$\phi_0(t)$ = carrier phase estimate provided by the Costas Loop

$v''(t)$ = assumed white noise process

$v(n)$ = sample of a white noise process

$\hat{\phi}_c(n)$ = smoother estimate of the carrier phase

The suppressed carrier waveform $r(t)$ is initially tracked by the Costas loop, which provides the estimate $\phi_0(t)$. The input to the smoother is the sampled carrier phase $\phi_c(n)$ plus a noise term, $v(n)$, that is related to $v'(t)$. $\phi_c(n)$ is modeled as a state noise process driven by random noise, but actually also has variations due to unmodeled dynamics. The output of the smoother is an improved, albeit delayed, estimation of $\phi_c(n)$. Note that whereas the actual sampling of the signal is performed at the data symbol rate, T_s, the estimates $\theta(n)$ and $\phi_0(n)$ are based on the averages of those samples; i.e., the sampling interval for $\phi_c(n)$ is $T = M T_s$.

In the proposed implementation, the initial phase estimates are made at the channel symbol rate, but are then averaged over M symbols before application to the real-time PLL and to the smoother. The smoother update rate of $1/M$ times the symbol rate can be chosen appropriately for the system parameters. A rate of 10 to 20 times the bandwidth of the filter is typical, i.e., of the filter that forms the basis of the smoother. As shown later, the smoother delay should be several correlation times. Suppose the delay is 0.5 s, which might be typical for DSN carrier tracking. At a symbol rate of 60 k symbols/s, a high rate for Voyager, only 30,000 symbols need be stored, which is practical with current memories.

III. Smoother Mathematical Model and General Solution

This section presents the mathematical model and the general solutions to the filtering and smoothing problems. Then the steady-state solutions of the smoother are investigated for the first- and second-order cases in Sections IV and V.

In navigation problems, the phase is related to position and the phase rate to velocity. Also, to use the terminology in several of the references, the following change of variables is used in the remainder of this article. For the second-order system, define the state vector $\mathbf{x}(n)$ at sampling time nT as

$$\mathbf{x}(n) = \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} \quad (2)$$

where

$$x_1(n) = \phi_c(n) = \text{position}$$

$$x_2(n) = \dot{\phi}_c(n) = \text{velocity}$$

We also consider a first-order system that has only the position variable.

A. Nomenclature

The following notation and assumptions are to be used:

Φ = state transition matrix (assumed constant)

$\mathbf{H}$ = observation matrix (assumed constant)

$\mathbf{K}(n)$ = Kalman gain matrix

$\mathbf{P}_F(n)$ = filter error covariance

$\mathbf{P}_P(n)$ = predictor error covariance

$\mathbf{P}_S(n)$ = smoother error covariance

$\mathbf{v}(n)$ = observation noise sample vector (white)

$\mathbf{w}(n)$ = state noise sample vector (white)

$\mathbf{x}(n)$ = state vector

$\mathbf{y}(n)$ = measurement vector

$\mathbf{Q}$ = covariance of the process $\mathbf{w}(n)$ (assumed constant)

$\mathbf{R}$ = covariance of the process $\mathbf{v}(n)$ (assumed constant)

The smoother proposed is an optimum linear Kalman smoother. Since the various equations that describe the estimator are given in terms of the Kalman filter equations, some of these equations are repeated here; however, the detailed description and solutions of them are left to the references. Noting the assumptions stated above, the following system model equations are considered:

$$\mathbf{x}(n+1) = \Phi\,\mathbf{x}(n) + \mathbf{w}(n)$$

$$\mathbf{Q} = E\,[\mathbf{w}(n)\mathbf{w}^T(n)] \tag{3}$$

$$\mathbf{y}(n) = \mathbf{H}\,\mathbf{x}(n) + \mathbf{v}(n)$$

$$\mathbf{R} = E\,[\mathbf{v}(n)\mathbf{v}^T(n)] \tag{4}$$

B. Filter Equations

For the filter estimation, i.e., calculation of the estimate $\mathbf{x}_F(n)$ of $\mathbf{x}(n)$ using observations up to the present sampling time, n, the general Kalman filter equations are (Ref. 1)

filter state:

$$\mathbf{x}_F(n) = \Phi\,\mathbf{x}_F(n-1) + \mathbf{K}(n)\,[\mathbf{y}(n) - \mathbf{H}\Phi\,\mathbf{x}_F(n-1)] \tag{5}$$

prediction error covariance matrix:

$$\mathbf{P}_P(n) = \Phi\,\mathbf{P}_F(n-1)\,\Phi^T + \mathbf{Q} \tag{6}$$

filter error covariance matrix:

$$\mathbf{P}_F(n) = [\mathbf{I} - \mathbf{K}(n)\mathbf{H}]\,\mathbf{P}_P(n) \tag{7}$$

Kalman gain matrix:

$$\mathbf{K}(n) = \mathbf{P}_P(n)\,\mathbf{H}^T\,[\mathbf{H}\mathbf{P}_P(n)\,\mathbf{H}^T + \mathbf{R}]^{-1} \tag{8}$$

The initial conditions are given as

$$E\,[\mathbf{x}(0)] = \mathbf{x}_F(0)$$

$$E\,[(\mathbf{x}(0) - \mathbf{x}_F(0))\,(\mathbf{x}(0) - \mathbf{x}_F(0))^T] = \mathbf{P}_F(0)$$

C. Smoother Equations

The following equations are for the fixed-lag smoother; i.e., a smoothing solution that estimates $\mathbf{x}(n)$ on the basis of measurements up to time $n + N$. From Refs. 1 and 2, the equations for such an estimate, denoted $\mathbf{x}_S(n)$, are

$$\mathbf{x}_S(n+1) = \Phi\,\mathbf{x}_S(n) + \mathbf{Q}\Phi^T\,\mathbf{P}_F^{-1}\,[\mathbf{x}_S(n) - \mathbf{x}_F(n)] + \mathbf{B}(n+1+N)\,\mathbf{K}(n+1+N)\,[\mathbf{y}(n+1+N) - \mathbf{H}\Phi\,\mathbf{x}_F(n+N)] \tag{9}$$

where

$$\mathbf{B}(n+1+N) = \prod_{i=n+1}^{n+N} \mathbf{A}(i), \quad \mathbf{A}(i) = \mathbf{P}_F(i)\Phi^T\,\mathbf{P}_P^{-1}\,(i+1)$$

$$n = 0, 1, \cdots$$

The initial state is $\mathbf{x}_S(0)$. The error covariance matrix of the smoothed estimates is

$$\mathbf{P}_S(n+1) = \mathbf{P}_P(n+1) - \mathbf{B}(n+1+N)\,\mathbf{K}(n+1+N)\,\mathbf{H}\,\mathbf{P}_P(n+1+N)\,\mathbf{B}^T(n+1+N) - \mathbf{A}^{-1}(n)\,[\mathbf{P}_F(n) - \mathbf{P}_S(n)]\,(\mathbf{A}^T(n))^{-1} \tag{10}$$

for $n = 0, 1, 2, \cdots$, where the initial condition is $\mathbf{P}_S(0)$. Both initial conditions, $\mathbf{x}_S(0)$ and $\mathbf{P}_S(0)$, are computed by using the fixed-point smoother equations described next (Refs.1 and 2). Let $\mathbf{x}_S(0\,|i)$ denote the optimum smoothing solution for the initial time 0 when i measurements have been taken. Then

$$\mathbf{x}_S(0\,|i) = \mathbf{x}_S(0\,|i-1) + \mathbf{B}(i-1)\,[\mathbf{x}_F(i) - \Phi\,\mathbf{x}_F(i-1)] \tag{11}$$

where

$$\mathbf{B}(i-1) = \prod_{j=0}^{i-1} \mathbf{A}(i), \qquad \mathbf{A}(i) = \mathbf{P}_F(j)\Phi^T\,\mathbf{P}_P^{-1}\,(j+1)$$

$$\mathbf{x}_S(0\,|0) = \mathbf{x}_F(0)$$

$$\mathbf{P}_S(0|i) = \mathbf{P}_S(0|i-1) + \mathbf{B}(i-1)\,[\mathbf{P}_F(i) - \mathbf{P}_P(i)]\,\mathbf{B}^T(i-1)$$

$$\mathbf{P}_S(0|0) = \mathbf{P}_F(0) \tag{12}$$

The initial conditions for the smoother with delay N are then $\mathbf{x}_S(0) = \mathbf{x}_S(0|N)$ and $\mathbf{P}_S(0) = \mathbf{P}_S(0|N)$.

The steady-state solutions of the first- and second-order smoothers are investigated in Sections IV and V.

IV. First-Order Smoother

In this case we consider only the estimation and measurement of position. Thus all the variables are scalars:

$$\begin{aligned} x &= \text{position} \\ \Phi &= 1 \\ \mathbf{H} &= 1 \\ \mathbf{Q} &= qT = \sigma_v^2\,T^2 \\ \mathbf{R} &= \sigma_0^2 \end{aligned} \tag{13}$$

where T is the sampling interval, q is the spectral density of the continuous white velocity process, and σ_0^2 is the variance of the position measurement.

Define the parameter $s = \sigma_0/(\sigma_v T)$. The value of s completely characterizes the solution. For the steady-state solution, $\mathbf{P}_F(n) = \mathbf{P}_F(n-1) = \mathbf{P}_F$, $\mathbf{P}_P(n) = \mathbf{P}_P(n-1) = \mathbf{P}_P$ and $\mathbf{K}(n) = \mathbf{K}(n-1) = \mathbf{K}$. Solving the corresponding Eqs. (6 through 8), the following solutions are obtained for the filter estimator, where P_P, P_F, P_S, and K are the scalar values of the corresponding matrices:

$$\frac{P_P}{\sigma_0^2} = \frac{\sqrt{1/4 + s^2} + 1/2}{s^2} \tag{14}$$

$$\frac{P_F}{\sigma_0^2} = \frac{\sqrt{1/4 + s^2} - 1/2}{s^2} \tag{15}$$

$$K = \frac{1}{\sqrt{1/4 + s^2} + 1/2} \tag{16}$$

$$\frac{P_S}{P_F} = \frac{1}{1+A} + \frac{KA^{2N+1}}{1-A^2} \tag{17}$$

where

$$A = P_F P_P^{-1} = 1 - K.$$

It is interesting to observe that as N increases, P_S/P_F approaches $(1 + A)^{-1}$. If also s is large, P_S/P_F approaches 1/2. Thus the variance of position error is reduced by a factor of 2 for the assumed conditions. This is in accordance with intuition as stated in the introduction. These results say nothing about the performance for other conditions.

A. Dynamic Tracking Performance

The dynamic tracking error of an estimation system is an important performance measure that is not obtained from the Kalman analysis. To obtain this performance, a z-transform approach is used.

Assuming steady-state values for the various parameters and taking the z-transform of the corresponding equations, the following transfer functions are obtained. For filtering,

$$F_F(z) = \frac{X_F(z)}{y(z)} = \frac{Kz}{z - (1-K)} \tag{18}$$

$$E_F(z) = \frac{y(z) - X_F(z)}{y(z)} = \frac{(z-1)(1-K)}{z-(1-K)} \tag{19}$$

where $E_F(z)$ is the error transfer function of the filter.

The smoother and smoother error transfer functions are

$$F_S(z) = \frac{X_S(z)}{y(z)} = \frac{z\,[A^N K z^{N+1} - A^N K z^N - 1/s^2]}{(z - 1/A)(z - A)} \tag{20}$$

$$E_S(z) = \frac{(z-1)(z - 1 - A^N K z^{N+1})}{(z - 1/A)(z - A)} \tag{21}$$

1. Response to step velocity. The steady-state position error due to a 1-m/s velocity step input can be computed by using the final value theorem. For this kind of input, $Y(z) = Tz\,(z-1)^{-2}$ and the position errors for filtering and smoothing are

$$e_F(\infty) = \frac{(1-K)\,T}{K} \tag{22}$$

$$e_S(\infty) = \frac{(1-K)^{N+1}\,T}{K} \tag{23}$$

The importance of this result is that the error of a first-order smoother to a step velocity input can be made arbitrarily small by choosing N sufficiently large since $e_S(\infty) \to 0$ as $N \to \infty$.

2. Response to step acceleration. For infinite delay smoothing, the error transfer function is ($N = \infty$ in Eq. (21)):

$$E_S(z) = \frac{(z-1)^2}{\left(z - \frac{1}{A}\right)(z-A)} \tag{24}$$

The z-transform of a 1-m/s^2 acceleration step input is $Y(z) = T^2 z(z+1)(z-1)^{-3}/2$. The position error for infinite delay smoothing is

$$e_S(\infty) = \lim_{z \to 1} \frac{(z-1)}{z} E_S(z) Y(z) = -\frac{(1-K)T^2}{K^2} \tag{25}$$

For $s > 5$ this error reduces approximately to

$$e_S(\infty) \approx -s^2 T^2 \tag{26}$$

For finite N, the error due to an acceleration step increases linearly with time. Figure 2 illustrates the position error as a function of time when N is finite and the input is a 1-m/s^2 acceleration step. The value of s is 5. It can be deduced from the results in Fig. 2 that for N finite, the error due to an acceleration step increases with time (t), approximately as $e_S(\infty)t$ where $e_S(\infty)$ is the position final error due to a 1-m/s velocity step and is given by Eq. (23). Thus for a known maximum observation time t, the error due to step acceleration can be made arbitrarily small by choosing N sufficiently large.

B. Noise Response

The Kalman solution gives the estimate error covariance due to the state noise and the observation noise. To obtain the error covariance due to the observation noise only, a z-transform approach is again used. The noise response of the kind of linear estimator discussed here is given by

$$\left(\frac{\sigma^2_{out}}{\sigma^2_{in}}\right) = \frac{1}{2\pi j} \oint_{\substack{UNIT \\ CIRCLE}} dz\, F(z) F(z^{-1}) z^{-1} \tag{27}$$

Substituting for $F_F(z)$ from Eq. (18) into Eq. (27), the noise reduction of the filter is

$$\left(\frac{\sigma^2_{out}}{\sigma^2_{in}}\right)_F = \frac{K}{2-K} \tag{28}$$

Computing Eq. (27) for the smoother usually requires some kind of numerical method unless $N \to \infty$, in which case

$$\left(\frac{\sigma^2_{out}}{\sigma^2_{in}}\right)_S = \frac{(1-K)^2\,(1+(1-K)^2)}{s^4\,(1-(1-K)^2)^3} \tag{29}$$

For this case of no state noise, Fig. 3 shows $(\sigma^2_{out}/\sigma^2_{in})$ vs NT/τ_1, where τ_1 is the time constant (correlation time) of the filter. Of course the filter and smoother parameters are based on the nonzero state covariance. The time constant is

$$\tau_1 = -\frac{T}{\ln(1-K)} \tag{30}$$

Note that K depends on s. It can be observed in Fig. 3 that if the smoothing time lag NT is approximately $4\tau_1$, then the noise response approximately reaches its asymptotic value given by Eq. (29).

V. Second-Order Smoother

For the second-order case, the state variable is a vector as in Eq. (2) and

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\Phi = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{H} = [1 \quad 0] \tag{31}$$

$$\mathbf{Q} = qT \begin{bmatrix} T^2/3 & T/2 \\ T/2 & 1 \end{bmatrix}, \qquad q = \sigma_0^2 T$$

$$\mathbf{R} = R = \sigma_0^2$$

where q is the spectral density of the assumed continuous acceleration process, σ_0^2 is the variance of the position measurement, and σ_a^2 is the variance of the acceleration noise. The covariance matrix $\mathbf{Q}$ is chosen to be consistent with previous work (Ref. 3).

The steady-state solution for the filter gains and error covariance matrices are obtained after considerable algebraic manipulations and can be found in Ref. 3. In Ref. 3, all the

solutions are expressed in terms of two parameters denoted there as r and s. For the case presented in this paper, s is infinity since there is no measurement of velocity. Therefore, all the solutions are functions of r defined as

$$r = \frac{4\sigma_0}{\sigma_a T^2} \tag{32}$$

From Eq. (10), the smoother steady-state error covariance $\mathbf{P}_S$ is the solution of

$$\mathbf{P}_S = \mathbf{P}_P - \mathbf{B}\,\mathbf{K}\,\mathbf{H}\,\mathbf{P}_P\,\mathbf{B}^T - \mathbf{A}^{-1}\,[\mathbf{P}_F - \mathbf{P}_S]\,(\mathbf{A}^T)^{-1} \tag{33}$$

where $\mathbf{P}_P$, $\mathbf{P}_F$, $\mathbf{K}$ are the steady-state filter solutions (Refs. 3 and 4). The asymptotic improvement due to smoothing $(\mathbf{P}_F - \mathbf{P}_S)$ may be shown to be (Ref. 5)

$$\mathbf{P}_F - \mathbf{P}_S = \mathbf{P}_F \sum_{i=0}^{N} \{[\tilde{\Phi}^T]^i\,\mathbf{H}^T\,[\mathbf{H}\,\mathbf{P}_F\,\mathbf{H}^T + \mathbf{R}]^{-1}\,\mathbf{H}\tilde{\Phi}^i\}\,\mathbf{P}_F \tag{34}$$

where $\tilde{\Phi} = \Phi - \mathbf{KH}$. The smoothing gain is obviously a monotonic function of N with most of the gain realized within a few time constants of the filter. Letting N approach infinity, we obtain

$$[\mathbf{P}_F - \mathbf{P}_{S,\infty}] - \mathbf{P}_F\,\tilde{\Phi}^T\,\mathbf{P}_F^{-1}\,[\mathbf{P}_F - \mathbf{P}_{S,\infty}]\,\mathbf{P}_F^{-1}\,\tilde{\Phi}\,\mathbf{P}_F = \mathbf{P}_F\,\mathbf{H}^T\,\left[\mathbf{H}\mathbf{P}_F\,\mathbf{H}^T + \mathbf{R}\right]^{-1}\,\mathbf{H}\mathbf{P}_F \tag{35}$$

The above is a linear equation in the elements of $(\mathbf{P}_F - \mathbf{P}_{S,\infty})$ and can be solved explicitly. Also, $\mathbf{A} = \mathbf{P}_F\,\Phi^T\,\mathbf{P}_P^{-1}$ and $\mathbf{B} = \mathbf{A}^N$.

$\mathbf{P}_S$ as a function of r and the delay N is determined by Eq. (33). Figure 4 shows the smoothed position accuracy, $P_S(1, 1)/\sigma_0^2$, as a function of r and N. The case $N = 0$ corresponds to the filter solution. For large observation noise compared to process noise (large r), the mean square position accuracy improvement due to smoothing over filtering is almost a factor of 4 for large N ($N \gtrsim 20$). This is almost twice the improvement realized in the first-order case. Intuitively, the additional improvement for the second-order case comes from the implicit estimation of rate as well as position.

As in the first-order case, dynamic tracking performance and noise response analyses are presented next.

A. Dynamic Tracking Performance

Taking the z-transforms of the corresponding recursive equations, the following transfer functions are obtained for the filter and smoother estimators.

$$\frac{X_{F1}(z)}{Y(z)} = F_F(z) = \frac{K_1 z^2 + (TK_2 - K_1)\,z}{z^2 + (K_1 + K_2 T - 2)\,z + (1 - K_1)} \tag{36}$$

$$\frac{X_{S1}(z)}{Y(z)} = F_S(z) = F_F(z) + \frac{z(z-1)^2\,\mathbf{H}\,[\mathbf{I} - z^N\mathbf{A}^N]\,[\mathbf{I} - \mathbf{A}z]^{-1}\mathbf{AK}}{z^2 + (K_1 + K_2 T - 2)z + (1 - K_1)} \tag{37}$$

where $\mathbf{K} = (K_1, K_2)^T$ is the Kalman gain vector, and $X_{F1}(z)$ and $X_{S1}(z)$ stand for the first elements of the vectors $\mathbf{X}_F(z)$ and $\mathbf{X}_S(z)$, respectively. Equation (37) has been derived from a numerically more stable recursive equation than of Eq. (9). This relation is

$$\mathbf{x}_S(n) = \mathbf{x}_F(n) + \mathbf{A}[\mathbf{x}_F(n+1) - \Phi\mathbf{x}_F(n)] + \cdots + \mathbf{A}^N[\mathbf{x}_F(n+N) - \Phi\mathbf{x}_F(n+N-1)] \tag{38}$$

The error transfer functions are

$$E_F(z) = \frac{(z-1)^2\,(1 - K_1)}{(z-1)^2 + (K_1 + K_2 T)(z-1) + K_2 T} \tag{39}$$

$$E_S(z) = E_F(z) - F_S(z) = \frac{(z-1)^2\,[1 - K_1 - z\mathbf{H}\,[\mathbf{I} - z^N\,\mathbf{A}^N]\,[\mathbf{I} - \mathbf{A}z]^{-1}\mathbf{AK}]}{(z-1)^2 + (K_1 + K_2 T)(z-1) + K_2 T} \tag{40}$$

1. Response to step acceleration. The z-transform of a 1-m/s^2 acceleration step input is $Y(z) = T^2 z(z+1)(z-1)^{-3}/2$. The steady-state position errors for the filter and smoother are

$$e_F(\infty) = \frac{(1 - K_1)T}{K_2} \tag{41}$$

$$e_S(\infty) = \frac{(1 - K_1 - \mathbf{H}[\mathbf{I} - \mathbf{A}^N]\,[\mathbf{I} - \mathbf{A}]^{-1}\mathbf{AK})T}{K_2} \tag{42}$$

The filter has finite but nonzero error due to step acceleration, but the smoother error approaches zero for large N.

Figure 5 shows the steady-state position error computed from Eqs. (41) and (42) vs the normalized smoothing time lag, NT/τ_2. The time constant, τ_2, of the second-order Kalman filter, Eq. (35), is obtained by using the mapping between the s plane and z plane (Ref. 6). The time constant is

$$\tau_2 = -\frac{2T}{\ln(1 - K_1)} \tag{43}$$

In Fig. 5, note the oscillatory nature of $|e_S(\infty)|$. One should not design N to operate on a null of this function because differences between the model and the actual system will affect the locations of the nulls. Arbitrarily small error can be achieved by choosing N large enough so that the error is sufficiently small for this N and all larger N.

2. **Response to step jerk.** For infinite delay smoothing, the position error transfer function is ($N \to \infty$ in Eq. (40))

$$E_S(z) = \frac{(z - 1)^4 (1 - K_1)}{[(z - 1)^2 + (K_1 + K_2 T)(z - 1) + K_2 T]\,[(z - 1)^2 - K_1 z(z - 1) + K_2 z]} \tag{44}$$

The z-transform of a 1-m/s^3 step jerk is

$$Y(z) = \frac{z(z^2 + 4z + 1)}{(z - 1)^4} T^3 \tag{45}$$

From Eqs. (44) and (45) and the final value theorem, it is easy to see that in this case $e_S(\infty) \to 0$. Thus, for infinite smoothing, the steady-state error to step jerk is zero.

For finite N, the error due to a step jerk increases linearly with time. Figure 6 shows the position error as a function of time when N is finite. The value of r is 500. It can be seen in Fig. 6 that for N finite, the error due to a step jerk increases with time approximately as $e_S(\infty)t$ where $e_S(\infty)$ is the position final error due to a 1-m/s^2 acceleration step and is given by Eq. (42). This error can be made arbitrarily small for any finite time interval.

3. **Summary of dynamic responses.** Table 1 shows a summary of the dynamic responses for the first- and second-order Kalman filters and smoothers.

B. Noise Response

As in the first-order case, Eq. (27), the noise response of the second order Kalman filter with no state noise is given by

$$\left(\frac{\sigma_{out}^2}{\sigma_{in}^2}\right)_F = \frac{2K_2 T - 3K_1 K_2 T + 2K_1^2}{K_1(4 - 2K_1 + K_2 T)} \tag{46}$$

The smoother responses are given in Eqs. (34) and (35). From Eq. (37), a specific expression for the smoother transfer function for $N = \infty$ is

$$F_S(z) = \frac{z[(2K_2 - K_2 K_1 - K_1^2)(z - 1)^2 + K_2^2 z]}{[(z - 1)^2 + (K_1 + K_2)(z - 1) + K_2]\,[(z - 1)^2 - (K_1 + K_2)z(z - 1) + K_2 z^2]} \tag{47}$$

Replacing $F_S(z)$ in Eq. (27) and using the results given in Ref. 7, the noise response can be easily computed.

Figure 7 shows the noise response of the Kalman filter and smoother vs NT/τ_2. Note that for any value of the parameter r, a smoothing time delay of roughly 1.5 τ_2 is enough for the smoother to approach its asymptotic noise response value.

VI. Examples for Uranus and Neptune Encounter

Of particular interest in the DSN is the tracking of the Voyager spacecraft in its close encounters with Uranus and Neptune. The approximate dynamics for Voyager encounters are

Planet	Acceleration	Jerk
Uranus	-0.32 m/s^2	0.83×10^{-4} m/s
Neptune	-4.00 m/s^2	0.29×10^{-2} m/s^3

Assuming an X-band carrier frequency of 8.4 GHz, a desired phase error of less than 1° corresponds approximately to $e_S(\infty) = 10^{-4}$m. In this example, it is assumed that the one-sided equivalent noise bandwidth of the Kalman filter used by the smoother is $B_F = 5$ Hz, because the Voyager spacecraft oscillator is known to be stable enough for this loop bandwidth. Also, it is assumed that $B_F T = 0.05$, so $T = 0.01$ s. The assumed Kalman parameters r and s are determined from the noise results of the previous sections and from the relationship $\sigma_{out}^2/\sigma_{in}^2 = 2B_F T$ for the filter ($N = 0$).

A. First-Order Smoother

For an acceleration step input and N finite, the first-order Kalman smoother has a position error increasing linearly with time (see Fig. 2). For $B_F T = 0.05$, the corresponding design parameter obtained from Fig. 3 is $s = 5$. In this case, from Eq. (26) or Fig. 2, even when N is very large the position error to a 1-m/s^2 acceleration step is approximately 0.0025 m, which means that the maximum tolerable position error for both Uranus and Neptune encounter is exceeded.

B. Second-Order Smoother

For $B_F T = 0.05$, the design parameter r obtained from Fig. 7 is $r = 500$. For Uranus encounter, a 10^{-4}-m lag error

with a 0.32-m/s^2 acceleration corresponds to $e_S(\infty)/T = 0.031$ at 1 m/s^2, which requires $NT/\tau_2 \geqslant 3.4$ (Fig. 5). For Neptune encounter, $NT/\tau_2 \geqslant 6.3$.

C. Second-Order Filter

The smoother reduces to the filter for $N = 0$. This corresponds to use of a PLL. The steady-state delay error is 0.01 m for a 1-m/s^2 acceleration, or 0.003 m at Uranus and 0.04 m at Neptune. These errors are not satisfactory.

D. Conclusion for Voyager

For the example parameters, a second-order PLL with a one-sided loop bandwidth of 5 Hz will not yield adequate dynamic performance for the Voyager encounters. Neither will a first-order smoother. A second-order smoother will result in satisfactory performance.

VII. Comments on Stability and Implementation Constraints

The smoother-state estimation equation as in Eq. (37) explicitly shows the dependence of the smoother on the corresponding Kalman filter. Since this dependence is expressed in a form of a finite sum of filter estimates, it follows that if the filter is stable so is the smoother. There are a large number of methods to ensure the filter stability in actual practical implementations (Ref. 1). For the results presented in this paper, a word length of 64 bits has been used to implement the various filters. It was found that a 32-bit word length was not enough to guarantee the stability of the filters, but 64 bits do suffice.

VIII. Conclusions

Use of smoothing filters to improve the phase estimation in carrier tracking systems has been proposed. First-order and second-order Kalman optimum smoothers are analyzed. The steady-state filter gains and error covariances are computed. Dynamic tracking performance and noise response are investigated by means of z-transform techniques. It is shown that a second-order Kalman smoother can keep the position error due to step acceleration or step-jerk inputs less than any prescribed maximum, assuming the step jerk is applied for finite time. It is also shown that the phase-error variance due to observation noise (receiver noise) is reduced by almost a factor of four by use of second-order smoothing rather than second-order filtering or a second-order PLL.

References

1. Gelb A., *Applied Optimal Estimation*, MIT Press, Cambridge, Mass., 1977.
2. Meditch, J. S., *Stochastic Optimal Linear Estimation and Control*, McGraw-Hill, New York, New York, 1969.
3. Ekstrand B., "Analytical Steady-State Solutions for Kalman Tracking Filter," *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 19, November 1983, pp. 815-819.
4. Friedland B., "Optimal Steady-State Positions and Velocity Estimation Using Noisy Sampled Position Data," *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 9, November 1973, pp. 906-911.
5. Anderson, B. D. O., and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, New Jersey, 1979.
6. Phillips, C. L., and H. T. Nagle, *Digital Control Systems Analysis and Design*, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.
7. Jury, E. I., *Theory and Application of the Z-transform Method*, R. E. Krieger, Huntington, New York, 1973.

Table 1. Summary of dynamic responses—steady-state position error, m

	Step velocity V	Step acceleration A	Step jerk J
First-order filter	$\frac{V(1-K)T}{K}$	∞	∞
First-order smoother	$\frac{V(1-K)^{N+1}T}{K}$	∞ (Ramp, Fig. 2)	∞
First-order smoother, $N = \infty$	0	$\frac{-A(1-K)T^2}{K^2}$	∞
Second-order filter	0	$\frac{A(1-K_1)T}{K_2}$	∞
Second-order smoother	0	A · (Eq. (41))	∞ (Ramp. Fig. 6)
Second-order smoother, $N = \infty$	0	0	0

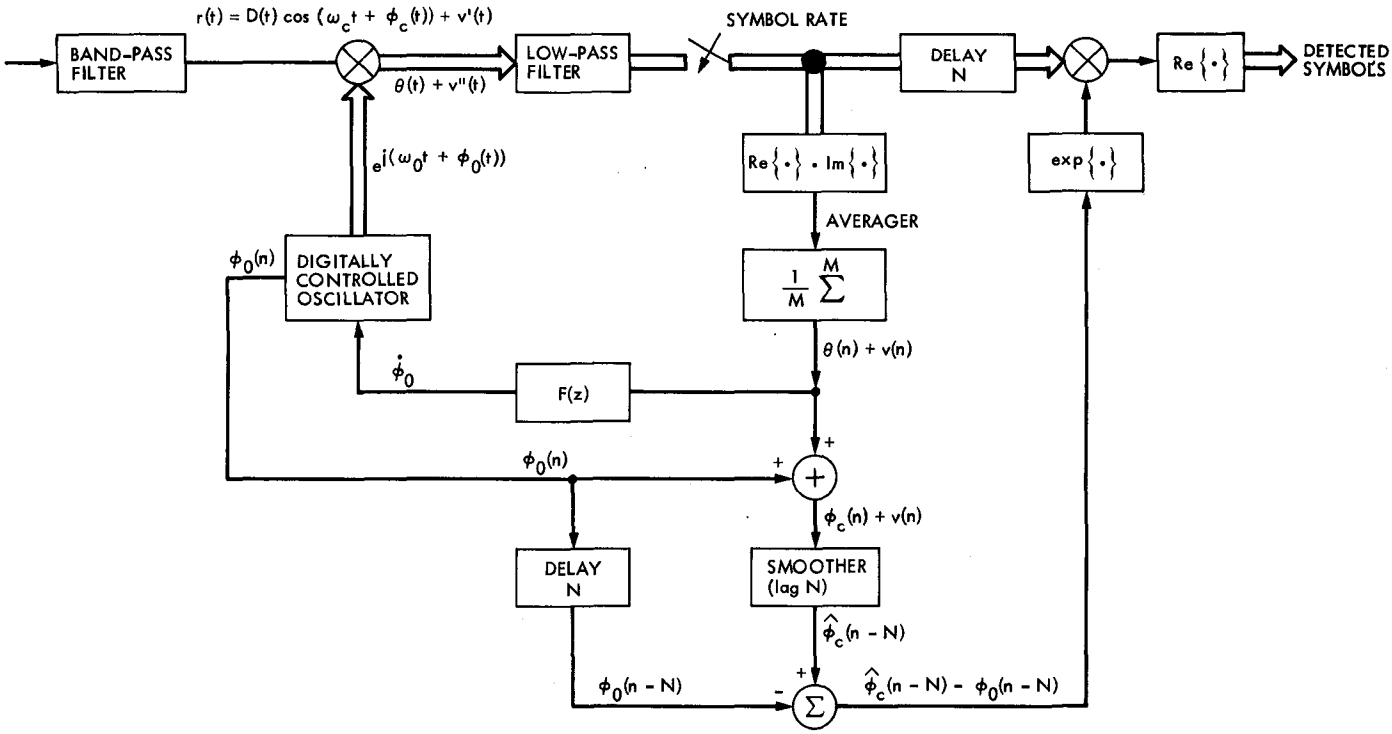


Fig. 1. Possible realization of smoothing estimator and data detection

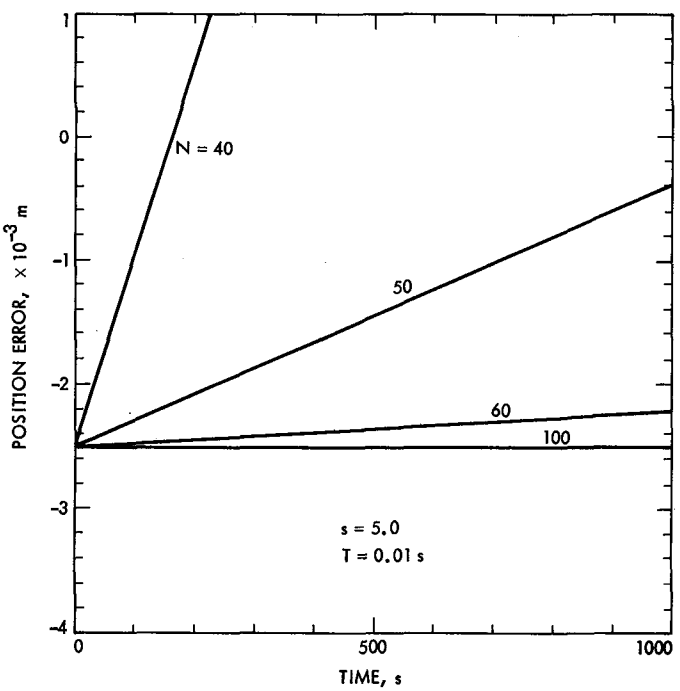


Fig. 2. First-order smoother: steady-state dynamic response to 1−m/s² step acceleration

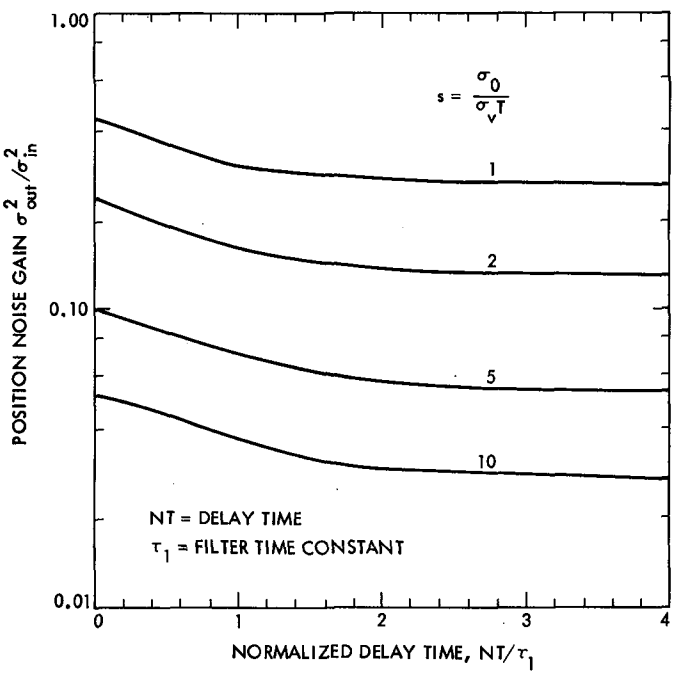


Fig. 3. First-order smoother response to observation noise; no state noise.

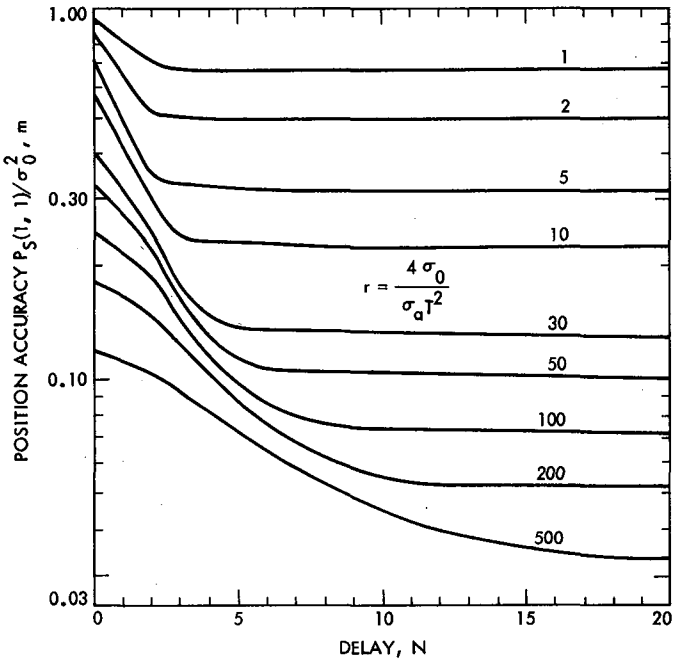


Fig. 4. Second-order smoother position accuracy with both observation and state noise

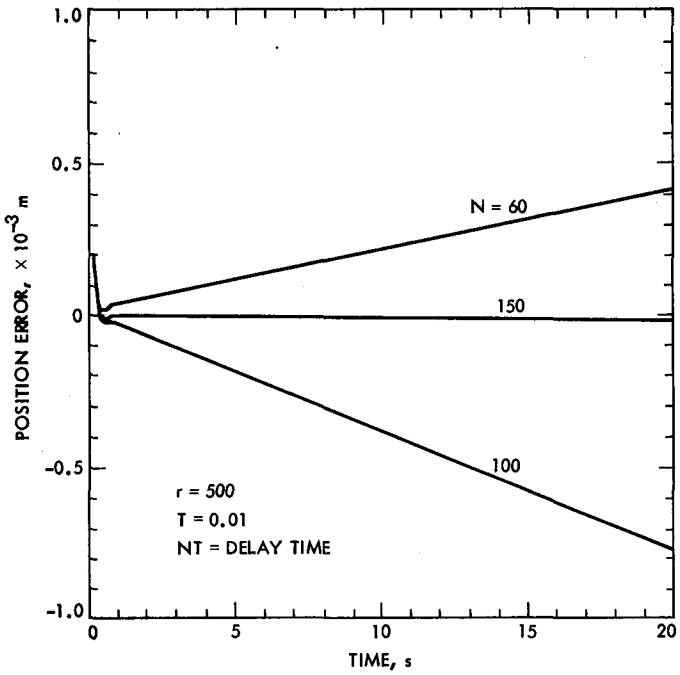


Fig. 6. Second-order smoother transient response to 1−m/s² step jerk

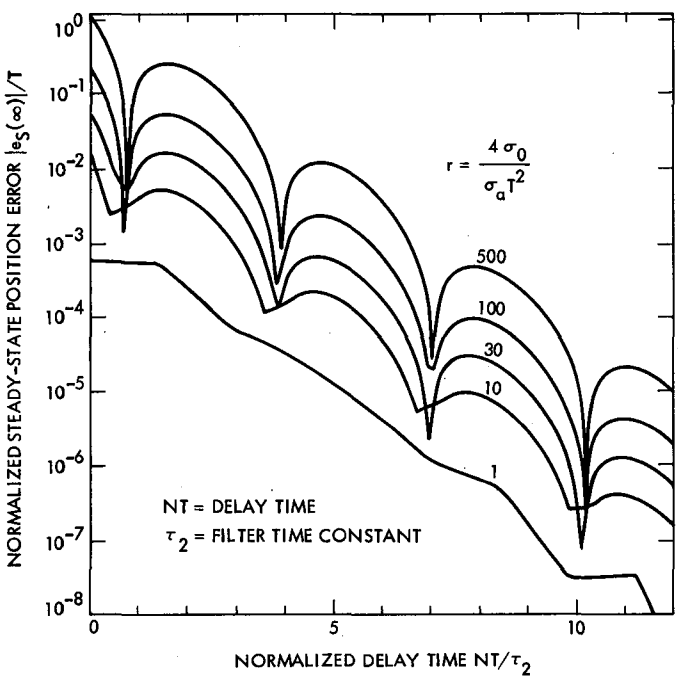


Fig. 5. Second-order smoother steady-state error to 1 m/s² step acceleration

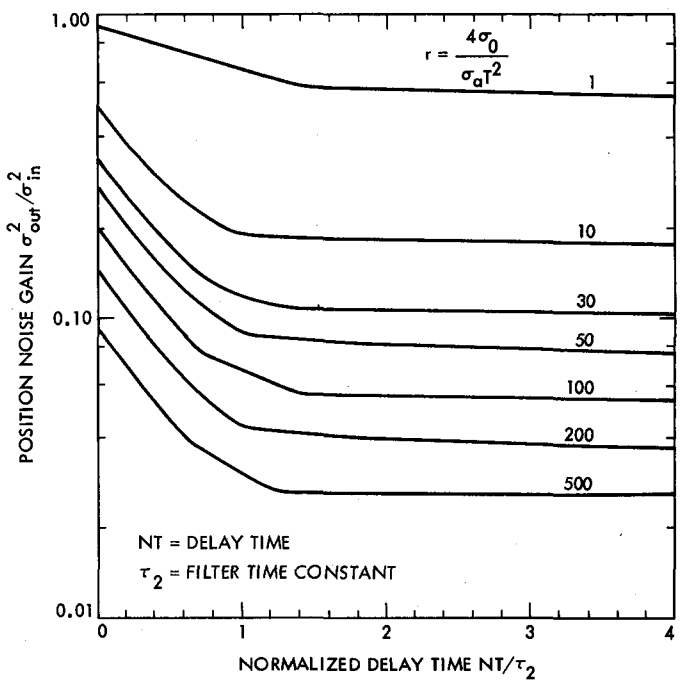


Fig. 7. Second-order smoother response to observation noise; no state noise

Analysis of a Coded, M-ary Orthogonal Input Optical Channel With Random-Gain Photomultiplier Detection

P. J. Lee
Communications Systems Research Section

Performance of two coding systems is analyzed in this study for a noisy optical channel with $M(= 2^L)$-ary orthogonal signaling and random-gain photomultiplier detection. The considered coding systems are the Reed-Solomon (RS) coding with error-only correction decoding and the interleaved binary convolutional system with soft decision Viterbi decoding. The required average number of received signal photons per information bit, $\bar{N}_b$, for a desired bit error rate of 10^{-6} is found for a set of commonly used parameters and with a high background noise level. We find that the interleaved binary convolutional coding system is preferable to the RS coding system in performance-complexity tradeoffs.

I. Introduction

The utilization of optical frequencies for deep-space communications has been considered as an attractive alternative to microwave frequencies, because of the advantages of smaller apparatus sizes and wider bandwidths. $M(= 2^L)$-ary orthogonal signaling is the most popular modulation scheme for energy-efficient optical communications when a large bandwidth-bit time product is available. Such a signaling is implemented by multiple-pulse-position modulation and/or multiple-color (frequency) modulation (Refs. 1 and 2). Additionally, direct detection employing a photomultiplier is often used for reception.

For further energy efficiency, coding over such a channel has been studied. When no background noise and ideal detection are assumed, the above channel can be modeled as an M-ary input, $(M + 1)$-ary output erasure-only channel (MEC). For this simplest channel, Reed-Solomon (RS) coding with erasure-only correction decoding was studied in Ref. 3. In Ref. 4, it was recognized that there is an equivalence between the MEC and the L-parallel correlated binary input, ternary output, erasure-only channels (BECs). Also in Ref. 4, the use of L-parallel, independent binary convolutional coding systems with Viterbi decoding were suggested; these systems exhibited performance favorable to those with similar complexity described in Ref. 3. For the channel with random-gain photomultiplier tube (PMT) reception, an RS coding system with combined error-and-erasure correction decoding was studied in Ref. 5, but with a negligibly low background noise level.

In this article, we consider the M-ary orthogonal input optical channel with random-gain PMT detection and with arbi-

trary background noise levels, as described in the next section. The performance of two coding systems over this channel is then studied. In Section III, the performance analysis of an RS coding system with error-only correction decoding is given. Then by extending the channel decomposition idea of Ref. 4, we propose a method of using interleaved, short-constraint-length binary convolutional codes with soft decision Viterbi decoding for this noisy optical channel, and analyze its performance. A numerical example with a high background noise level is given. We conclude that, as in the case of the noiseless ideal optical channel, the interleaved binary convolutional coding system is preferable to the RS coding system for the noisy optical channel with nonideal photon counting, which results from realistic PMTs.

II. Noisy M-ary Orthogonal Input Optical Channel With a Random-Gain PMT Receiver

Let an optical transmitter emit an optical pulse in one pulse position among M_0 nonoverlapping pulse positions with one color among M_1 nonoverlapping colors at each symbol time T_s. Suppose the receiver can separate the pulse positions and the colors. Then we have an $M(= M_0 \times M_1)$-ary input optical channel. Let X be the channel input symbol random variable, which takes its value from the M-ary channel input symbol alphabet $\mathcal{M} = \{0, \cdots, m, \cdots, M-1\}$. For each T_s, the receiver is to produce an M-dimensional observation vector $\mathbf{y} = (y_0, \cdots, y_m, \cdots, y_{M-1})$ where each y_m is the output at the detector consisting of a PMT and an integrate-and-dump filter for the assigned color-time slot. A model for the detector is shown in Fig. 1, where T_c $(= T_s/M_0)$ is the slot time and R is the PMT anode load resistance. Note that we need such a PMT detector for each M_1 color slot and the color splitter. We assume symmetries and independences for every time and color slot so that the resulting channel becomes an M-ary orthogonal input, symmetric output, memoryless channel. We restrict our attention to $M = 2^L$ cases with integer L.

A PMT is characterized by its average gain A and the number of dynodes ν. The variance of the PMT gain is $2AB$ where $2B = (A-1)/(A^{1/\nu}-1)$. In Ref. 6, a Markov diffusion approximation to the statistics of PMT gain was given. Including the effect of thermal noise, the statistics of the detector output voltage random variable Y, conditioned on the number of input photons a, is given by (Ref. 6)

$$H(y|a) = \Pr\{Y \leqslant y|a\}$$

$$= C \cdot Q(-y/s) + \int_0^\infty Q(-(y-cz)/s) \cdot h(z|a) \cdot dz$$

where

$$h(z|a) = e^{-a} \cdot \sum_{n=1}^{\infty} \frac{1}{n!} (a \cdot e^{-A/B})^n \cdot \sum_{k=1}^{\infty} \frac{(nA/B^2)^k \cdot z^{k-1}}{k!(k-1)!} e^{-z/B}$$

$$Q(y) \equiv \int_y^\infty \exp(-t^2/2) \cdot dt/\sqrt{2\pi},$$

$C = \exp[-a\{1 - \exp(-A/B)\}]$, s^2 is the variance of thermal noise $(= N_0/T_c$, N_0 is the one-sided thermal noise spectral density), and $c = \epsilon R/T_c$ (ϵ is the electron charge).

Let $\bar{N}_n$ be the average number of noise photons per slot and $\bar{N}_s$ be the average number of received signal photons per M-ary symbol. Recall that there is only one active slot among M possible slots. Let $G(y)$ be the cumulative distribution function for the active slot and $F(y)$ be that for a nonactive slot. Then we have

$$G(y) = H(y|a = \bar{N}_s + \bar{N}_n) \text{ and } F(y) = H(y|a = \bar{N}_n)$$

and the channel transition probabilities for each $m \in \mathcal{M}$ can be represented as

$$\Pr\{\mathbf{Y} \leqslant \mathbf{y} \mid X = m\} = G(y_m) \cdot \prod_{m' \neq m} F(y_{m'})$$

III. Reed-Solomon Coding System With Error-Only Correction Decoding

For an M-ary orthogonal input, symmetric output channel, one of the best coding systems is the (N, K) RS coding system whose symbol alphabet size is M, where K and N are the numbers of information and channel symbols in an RS codeword (Ref. 7). The code rate r is K/N [information symbols/channel symbol], and N is usually chosen to be $M-1$ or M. When the decoder is capable of correcting errors only, a hard decision (i.e., M-ary symbol decision) must be performed in front of the decoder. By assuming equiprobable channel input symbols, it is easy to show that the maximum likelihood (ML) symbol decision rule is optimum due to the symmetries of the channel. Since $dG(y)/dF(y)$ is a monotonically increasing function of y for this optical channel, the decision rule simplifies to, "declare m as the transmitted symbol, if the m-th observation value y_m is the largest." The corresponding decision error probability, p, is given by

$$p = (M-1) \cdot \int_{-\infty}^{\infty} G(y) \cdot F^{M-2}(y) \cdot dF(y)$$

The resulting RS decoded bit-error rate (BER) is given by (Ref. 7)

$$BER = \frac{M}{2(M-1)} \cdot \sum_{i=t+1}^{N} \binom{N}{i} \cdot p^i \cdot (1-p)^{N-i} \cdot \frac{i}{N}$$

where t (= integer part of $(N-K)/2$) is the number of correctable error symbols in an RS codeword.

IV. Binary Convolutional Coding System With Soft Decision Viterbi Decoding

For binary-input symmetric-output channels, a short-constraint-length binary convolutional coding system with soft decision Viterbi decoding is one of the most practical and powerful coding systems available. Let $U \in \{0, 1\}$ be the binary channel input symbol and $\mu(\mathbf{y}, U)$ is the metric to be used by the decoder for the channel input symbol U with a given channel output observation vector $\mathbf{y}$. Let D be the Chernhoff bound on the probability that the metric for a transmitted symbol is smaller than that for a nontransmitted symbol. Then we have (Ref. 8)

$$D = \min_{w \geq 0} D(w)$$

$$D(w) = E[\exp\{w \cdot [\mu(\mathbf{Y}, U=1) - \mu(\mathbf{Y}, U=0)]\} \mid U=0]$$

where the operator E is the expectation over the observation random vector $\mathbf{Y}$. Notice that D becomes the Bhattacharyya bound when the metric is an ML metric. For a given binary convolutional encoder having a transfer function $T(Z, I)$, the decoded BER at the output of the associated Viterbi decoder using the metric μ is well bounded by (Ref. 8)

$$BER \leq 1/2 \cdot \frac{\partial}{\partial I} T(Z, I)\Big|_{Z=D,\ I=1}$$

Recall that $M = 2^L$ and each one of the M-ary symbols is assumed to have a one-to-one correspondence with a set of L bits. An example of such a mapping is shown in Table 1 for the $M = 8$ ($L = 3$) case. Suppose we have an MEC with erasure probability of q. That is, the transmitted symbol is either correctly received with probability $1 - q$ or is erased with probability q. Hence, once the receiver has an unerased symbol, all the corresponding L bits are recovered correctly. Alternately, when the receiver has an erased symbol, all the L bits are also erased. Accordingly, the MEC is equivalent to L-parallel binary erasure-only channels (BEC), each with the same erasure probability q. Such component channels are fully correlated, since the erasure events always occur simultaneously for all of L-component channels. Massey, in Ref. 4, noticed this equivalence and suggested use of the binary convolutional coding system with Viterbi decoding for each component binary input channel independently (i.e., by ignoring the correlation). Note that, instead of using L-parallel binary coding systems, one may use one binary coding system with interleaving (Ref. 4). Also, the binary convolutional coding system was shown to be preferable to RS coding system in performance-complexity tradeoffs, for the ideal optical channel with no background noise.

Now we shall extend the Massey's channel decomposition idea for the M-ary optical channel to the case of nonideal detection and with background noise as described in Section II. Let $\mathcal{I}_\ell$ ($\mathcal{J}_\ell$) represent the set of all M-ary channel input symbols having a 0 (1) in the ℓ-th component of their binary representations (respectively). For the example shown in Table 1, $\mathcal{I}_2 = \{0, 1, 4, 5\}$ and $\mathcal{J}_3 = \{1, 3, 5, 7\}$. Let i_ℓ and j_ℓ be the largest value among the channel output observation values in the set $\mathcal{I}_\ell$ and $\mathcal{J}_\ell$. For the example set of observation values in Table 1, $i_2 = \max(3.11, 2.72, 6.67, 4.06) = 6.67$ and $j_3 = 4.06$. We may use this i_ℓ and j_ℓ directly for the metric for the ℓ-th component channel input binary symbol U_ℓ, i.e., $\mu(\mathbf{y}, U_\ell = 0) = i_\ell$ and $\mu(\mathbf{y}, U_\ell = 1) = j_\ell$. This is not an ML metric for such a binary input coding channel, but is a good choice based on the complexity consideration. It is easy to show that the corresponding Chernhoff bounds are given, for each $\ell = 1, 2, \cdots, L$, by

$$D_\ell = D = \min_{w \geq 0} D(w)$$

$$D(w) = \left[(M/2) \cdot \int_{-\infty}^{\infty} \exp(wy) \cdot F^{M/2-1}(y) \cdot dF(y)\right] \cdot \left[\int_{-\infty}^{\infty} \exp(-wy) \left\{\left(\frac{M}{2} - 1\right) \cdot G(y) \cdot F^{M/2-2}(y) \cdot dF(y) + F^{M/2-1}(y) \cdot dG(y)\right\}\right]$$

where we have used the following:

$$\Pr\{I_\ell \leq y \mid U_\ell = 0\} = \Pr\{J_\ell \leq y \mid U_\ell = 1\} = G(y) \cdot F^{M/2-1}(y)$$

$$\Pr\{I_\ell \leq y \mid U_\ell = 1\} = \Pr\{J_\ell \leq y \mid U_\ell = 0\} = F^{M/2}(y)$$

where I_ℓ and J_ℓ are the random variables corresponding to i_ℓ and j_ℓ, respectively.

V. Numerical Example, Discussion, and Conclusions

The coding systems described and analyzed in the previous sections are compared for the specific example of an optical channel with random-gain PMT reception and a very high background noise level of 10^8 noise photons per second. Such a background level is expected when a spacecraft is in front of a bright planet. All the other parameters are chosen to be the same as those in Refs. 5 and 6: the slot time $T_c = 10^{-7}$ s, average PMT gain $A = 10^6$, the number of PMT dynodes $v = 11$, the PMT anode load resistance $R = 50\ \Omega$, and the one-sided thermal noise power spectral density $N_0 = 1.156 \times 10^{-17}$ V/Hz.

With these parameters and the equations in the previous sections, we find the required average number of received signal photons per symbol, $\bar{N}_s$, for a desired BER of 10^{-6} and for some values of channel symbol alphabet size M and code rate r. For the RS coding system, $5 \leqslant L \leqslant 8$ and all possible code rates $r = K/N$ were considered with $N = M$. For the interleaved binary coding system, we consider $4 \leqslant L \leqslant 10$ and $r = 1/3$, 1/2, 2/3, 3/4, and 4/5. For the latter system, there is another choice of code parameter, which is the number of states, S, in the Viterbi decoding. For a regular rate $1/n$ convolutional code with constraint length k, S is given by 2^{k-1}. For this numerical example, we restricted S to be 64 (or equivalently, $k = 7$), and used the (7, 1/2) and (7, 1/3) codes (found in Ref. 9) and the $r = 2/3$, 3/4, and 4/5 punctured codes derived from the (7, 1/2) codes (found in Ref. 10). For such punctured codes, a decoder for the (7, 1/2) code can be used directly with minor modifications on metric forming (Ref. 10).

The quantity of interest for energy-efficient optical communication is the average number of received photons per information bit, $\bar{N}_b$, rather than $\bar{N}_s$. Note that $\bar{N}_b = 1/\rho$ where ρ is the commonly used parameter (Ref. 5) for photon efficiency in [bits/photon]. $\bar{N}_b$ is related to $\bar{N}_s$ by $\bar{N}_b = \bar{N}_s/(rL)$. The curves of required values of $\bar{N}_b$ for BER $\leqslant 10^{-6}$ versus code rate r are plotted in Fig. 2. The solid lines correspond to the RS coding system performances while the broken lines refer to the interleaved binary convolutional coding system performances.

From Fig. 2 we see that RS coding systems perform better for larger values of L (>6), while the interleaved binary 64-state convolutional coding systems perform better for smaller values of L (<6). However, notice that the RS coding system complexity grows exponentially as L increases, while the complexity of the binary coding system remains almost the same for any value of L. Hence, for a fair comparison between the two coding systems, one must consider the system complexity as well as the performance. For a given channel input alphabet size $M = 64$ ($L = 6$), the two coding systems show almost identical performance, but the interleaved 64-state binary Viterbi decoding system is considered to be less complex than the 6-bit RS decoding system.

For better performance with the interleaved binary coding system, one may increase the system complexity by employing longer constraint-length codes. The interleaved binary 256-state ($k = 9$) Viterbi decoded convolutional coding system performs almost the same as the 7-bit RS coding system for the noisy 128-ary optical channel and requires less complexity. From these performance-complexity considerations, we conclude that the interleaved binary convolutional coding system is preferable to the RS coding system for the noisy nonideal channel, similar to the results for the noiseless ideal optical channel considered in Ref. 4.

One may consider some variations of the two coding systems. For the RS coding system, instead of error-only correction decoding, one can consider a combined error-and-erasure correction decoding. This modification can improve performance, but the amount of improvement is usually very small except for the cases of very low background noise level, even with properly designed $(M + 1)$-ary decision rules (see Ref. 7). The metric given in Section IV is the simplest unquantized metric for the component channel. Hence, one may consider better metrics that more closely approximate ML decoding metric. Furthermore, one can consider concatenated coding systems employing the interleaved convolutional coding as an inner coding. Hence, there is more flexibility and more room for performance improvement with the interleaved convolutional coding system.

Acknowledgment

The author would like to thank Prof. J. K. Omura of U.C.L.A. and Dr. J. R. Lesh of JPL for their encouragement and helpful discussions.

References

1. Pierce, J. R., "Optical Channels: Practical Limits With Photon Counting," *IEEE Trans. Comm.*, Vol. COM-16, pp. 1819-1821, Dec. 1978.
2. Davidson, F., "Direct-Detection Optical Communication with Color Coded Pulse Position Modulation Signaling," submitted to *IEEE Trans. Commun.*
3. McEliece, R. J., "Practical Codes for Photon Communication," *IEEE Trans. Infor. Theory*, Vol. IT-27, pp. 393-408, July 1981.
4. Massey, J. L., "Capacity, Cutoff Rate, and Coding for a Direct-Detection Optical Channel," *IEEE Trans. Commun.*, Vol. COM-29, pp. 1615-1621, Nov. 1981.
5. Lesh, J. R., J. Katz, H. H. Tan, and D. Zwillinger, "2.5 Bits/Detected Photon Demonstration Program: Analysis and Phase I Results," *TDA Progress Report 42-66*. Jet Propulsion Laboratory, Pasadena, Calif., pp. 115-132, Dec. 1981.
6. Tan, H. H., "A Statistical Model of the Photomultiplier Gain Process With Applications to Optical Pulse Detection," *TDA Progress Report 42-68*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 55-67, April 1982.
7. Berlekamp, E. R., "The Technology of Error-Correcting Codes," *Proceedings of IEEE*, Vol. 68, pp. 564-593, May 1980.
8. Omura, J. K., and M. K. Simon, *Modulation/Demodulation Techniques for Satellite Communications, Part IV. Appendices*, JPL Publication 81-73, Jet Propulsion Laboratory, Pasadena, Calif., Nov. 1981.
9. Odenwalder, J. P., *Optimal Decoding of Convolutional Codes*, Ph.D. Thesis, School of Engineering and Applied Sciences, U.C.L.A., Los Angeles, Calif., 1970.
10. Yasuda, Y., K. Kashiki, and Y. Hirata, "High-Rate Punctured Convolutional Codes for Soft Decision Viterbi Decoding," *IEEE Trans. Commun.*, Vol. COM-32, pp. 315-319, March 1984.

Table 1. An example of M-ary symbol to L binary symbols mapping and an example set of observation values

m	$u_1\ u_2\ u_3$	y_m
0	0 0 0	3.11
1	0 0 1	2.72
2	0 1 0	1.53
3	0 1 1	1.18
4	1 0 0	6.67
5	1 0 1	4.06
6	1 1 0	1.74
7	1 1 1	0.98

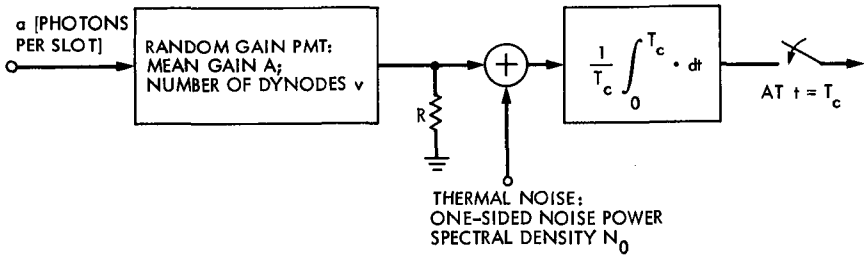


Fig. 1. A model for an optical receiver employing a random-gain photomultiplier tube

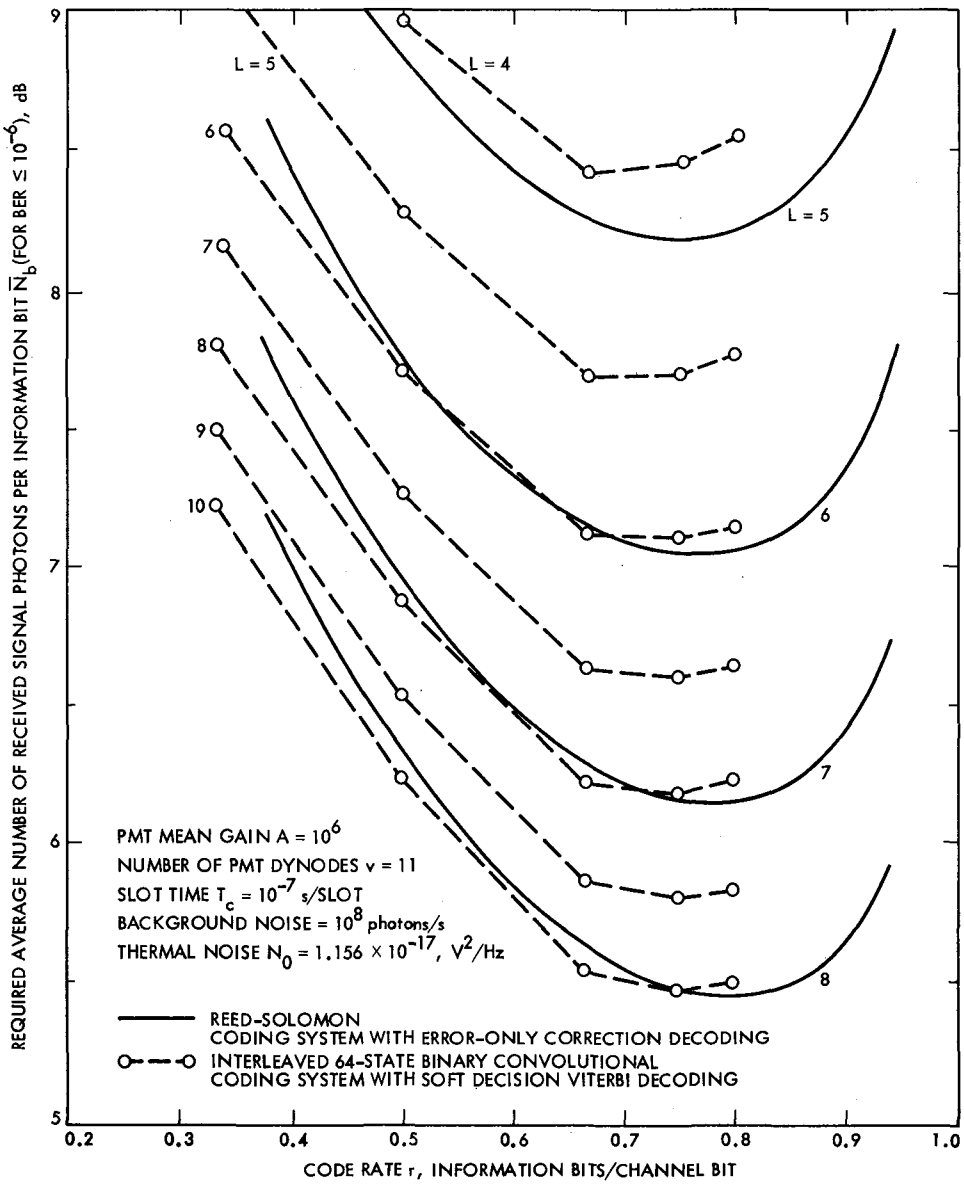


Fig. 2. Performances of coding systems for $M(= 2^L)$-ary orthogonal input noisy optical channel

A Very Efficient Transfer Function Bounding Technique on Bit Error Rate for Viterbi Decoded, Rate 1/N Convolutional Codes

P. J. Lee
Communications Systems Research Section

For rate 1/N convolutional codes, a recursive algorithm for finding the transfer function bound on bit error rate (BER) at the output of a Viterbi decoder is described. This technique is very fast and requires very little storage since all the unnecessary operations are eliminated. Using this technique, we find and plot bounds on the BER performance of known codes of rate 1/2 with $K \leq 18$, rate 1/3 with $K \leq 16$, and rate 1/4 with $K \leq 14$. When more than one reported code with the same parameters is known, we select the code that minimizes the required signal-to-noise ratio for a desired bit error rate of 10^{-6}. This criterion of determining goodness of a code had previously been found to be more useful than the maximum free distance criterion and was used in the code search procedures of very short constraint length codes. This very efficient technique can also be used for searches of longer constraint length codes.

I. Introduction

The best decoding method for convolutional codes is maximum-likelihood (ML) decoding (often called Viterbi decoding) (Refs. 1 and 2), which is considered to be practical only for "short" constraint length codes. For longer constraint length codes sequential decoding is often employed (Ref. 2). However, due to rapidly developing hardware technologies, the length which is considered to be "short" has been increasing. Also, the bit error rate (BER) with sequential decoding may be lower bounded by the BER with ML decoding. Hence, finding the performance of Viterbi-decoded long constraint length codes is useful, even if construction of Viterbi decoders for such length codes is impractical with today's technologies.

The BER at the Viterbi decoder output is well bounded by the well-known transfer function bound (Refs. 1 and 2). For this bound, matrix inversion is required (Refs. 2 and 3). Accordingly, a large amount of computer storage and a substantial amount of computing time have been required. In this report, for rate $1/N$ convolutional codes, we present a recursive algorithm for finding the transfer function bound. By eliminating all the unnecessary computations (e.g., multiplication-by-zero, etc.), we can perform the matrix inversion with vector operations only. Hence, this technique is very fast computationally, and it furthermore requires only a very small amount of computer memory storage. After a brief review of preliminaries in the next section, the algorithm is described in Section III.

A large number of good rate $1/N$ convolutional codes have been found and reported (Refs. 4 to 7). In those code search procedures, maximum free distance (d_f) or the maximum d_f together with minimizing the first few weight spectral components (number of adversaries) were used for determining the

goodness of a code. More recently, in Ref. 8, we introduced a better criterion of "minimum required signal-to-noise ratio (SNR) for a given desired BER" and used this criterion for code searches of rate $1/N$ codes with very short constraint length ($K \leqslant 7$). For searches of longer constraint length codes, this technique will be used for finding BER performance bounds.

In the last section, we give the BER performance of some rate $1/N$ codes which require minimum SNR for desired BER of 10^{-6} among codes reported in Refs. 4 to 8. (For $K \geqslant 8$, we expect that there may be better codes with our criterion.) For the calculations, we assume binary antipodal signaling over the additive white Gaussian noise (AWGN) channel with no channel output quantization. Such codes are tabulated in Table 1. With a discussion on the accuracy of the bound, we conclude that, as compared to the series expansion approximation to the transfer function bound, our method is shown to be preferable in two aspects: (1) it gives uniform accuracy and (2) it is more suitable for comparison of the codes.

II. Preliminaries

In this section a general background for the transfer function bound is briefly reviewed, mainly to define necessary notation. A typical nonsystematic, constraint length K, rate $1/N$ convolutional encoder is shown in Fig. 1. The connection box with mod-2 adders is often represented by an $N \times K$ binary matrix **G**, which is called the code generator matrix. The n-th bit in t-th output vector y_n^t (see Fig. 1) for $n = 1, 2, \ldots, N$ and $t = 1, 2, \ldots$ is:

$$y_n^t = mod\left\{\sum_{k=0}^{K-1} \mathbf{G}(n, k) \cdot x^{t-k}, 2\right\} \tag{1}$$

where $mod\{a, b\}$ is the remainder when a is divided by b, $x^t \in \{0, 1\}$ is the encoder input sequence for $t = 1,2, \ldots$, and $x^t = 0$ for $t \leqslant 0$ by convention. The code rate r [information bits/channel bit] is $1/N$. The "present" state at time t, S^t, is defined as $S^t = (x^{t-K+1}, \ldots, x^{t-1})$, and we denote S^t by i, if

$$\sum_{k=1}^{K-1} x^{t-k}\, 2^{k-1} = i \tag{2}$$

Hence the state space is $\{0,1,2, \ldots, M-1\}$, where $M = 2^{K-1}$ is the size of the state space. Suppose S^t is i. Then by the definition of state, the next state j is given by:

$$j = x^t + mod\{2i, M/2\} \tag{3}$$

To find a transfer function, one often uses a state diagram where nodes represent states and directed branches represent state transitions. Let $\mathbf{W}(i,j)$, $i,j = 0,1, \ldots, M-1$, be the matrix representation of the branch metric on the directed branch from state i to state j. When such a transition exists, $\mathbf{W}(i,j)$ is given by the product of D to the power of the Hamming weight of the corresponding output vector and Z raised to the Hamming weight of the input bit, when a binary input channel is used. (D and Z are dummy variables. See Eq. (5).) While, $\mathbf{W}(i,j) = 0$ when there is no such transition. As an illustration, a $K = 4$, $r = 1/5$ convolutional encoder is shown in Fig. 2. For this code, **G** and **W** are given below and the state diagram is shown in Fig. 3.

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{W} = \left[\begin{array}{c|cccccccc} 1 & D^5Z & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & D^3 & D^2Z & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & D^2 & D^3Z & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & D^3 & D^2Z \\ D^5 & Z & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & D^2 & D^3Z & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & D^3 & D^2Z & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & D^2 & D^3Z \end{array}\right]$$

The transfer function $T(D,Z)$ can be represented by (Refs. 2 and 3)

$$T(D,Z) = \mathbf{b} \cdot (\mathbf{I} - \mathbf{A})^{-1} \cdot \mathbf{c} \tag{4}$$

where $(M - 1)$-dimensional row and column vectors **b** and **c** and $(M - 1) \times (M - 1)$ matrix **A** are portions of the $M \times M$ matrix **W**, such that $\mathbf{b}(j) = \mathbf{W}(0,j)$, $\mathbf{c}(i) = \mathbf{W}(i,0)$, and $\mathbf{A}(i,j) = \mathbf{W}(i,j)$, where $i,j = 1,2, \ldots, M - 1$. Here, **I** is the $(M - 1) \times (M - 1)$ identity matrix.

The BER at the output of an ML decoder is well upper-bounded by the following expression, called the "transfer function bound" (Refs. 1 to 3)

$$BER \leqslant C_o \cdot \frac{\partial}{\partial Z} T(D,Z)\bigg|_{D=D_o, Z=1} \tag{5}$$

where the coefficient C_o depends on the channel and code used, while the Bhattacharyya bound D_o depends only on the coding channel (everything between the encoder output and the decoder input, including decision rule or channel output quantization). For the AWGN channel with antipodal signaling and with no channel output quantization, C_o and D_o are given by (Refs. 1 and 2)

$$D_o = \exp(-E_s/N_o) \tag{6}$$

and

$$C_o = Q(\sqrt{2d_f E_s/N_o}) \cdot \exp(d_f E_s/N_o) \tag{7}$$

where E_s is the received signal energy per channel symbol which is related to the received signal energy per information bit E_b by, $E_s = r E_b$ ($r = 1/N$ in our case). N_o is the one-sided noise power spectral density,

$$Q(z) = \int_z^\infty \exp(-t^2/2) \cdot dt/\sqrt{2\pi}$$

and d_f is the free distance of the code.

III. Transfer Function Bound

When the matrix $(\mathbf{I} - \mathbf{A})$ has an inverse, the following holds:

$$(\mathbf{I} - \mathbf{A})^{-1} = \sum_{\ell=0}^{\infty} \mathbf{A}^\ell \tag{8}$$

or

$$T(D,Z) = \sum_{\ell=0}^{\infty} \mathbf{b} \cdot \mathbf{A}^\ell \cdot \mathbf{c} \tag{9}$$

Let

$$\mathbf{f}_\ell = \mathbf{b} \cdot \mathbf{A}^\ell \tag{10}$$

which can be found recursively as

$$\mathbf{f}_{\ell+1} = \mathbf{f}_\ell \cdot \mathbf{A}, \ell = 0,1,2,\ldots; \text{with } \mathbf{f}_o \triangleq \mathbf{b} \tag{11}$$

Hence

$$T(D,Z) = \sum_{\ell=0}^{\infty} \mathbf{f}_\ell \cdot \mathbf{c} \tag{12}$$

Let

$$\mathbf{g}_\ell = \partial \mathbf{f}_\ell / \partial Z \tag{13}$$

This also can be found recursively

$$\mathbf{g}_{\ell+1} = \mathbf{g}_\ell \cdot \mathbf{A} + \mathbf{f}_\ell \cdot \mathbf{A}', \ell = 0,1,\ldots; \text{with } \mathbf{g}_o = \mathbf{b}' \tag{14}$$

where $\mathbf{A}' = \partial \mathbf{A}/\partial Z$ and $\mathbf{b}' = \partial \mathbf{b}/\partial Z$.

Since $\partial \mathbf{c}/\partial Z = \mathbf{0}$, we have

$$\frac{\partial}{\partial Z} T(D,Z)\Big|_{D=D_o, Z=1} = \left\{\sum_{\ell=0}^{\infty} \mathbf{g}_\ell \cdot \mathbf{c}\right\}\Big|_{D=D_o, Z=1} \tag{15}$$

Equations (4) through (15) hold for any convolutional code. That is, for any rate b/N convolutional code, we can find the transfer function bound recursively, with a proper truncation of Eq. (15), by performing vector-by-matrix multiplications.

For rate $1/N$ codes, we can further reduce the required computer storage and computational effort. Notice that the matrix $\mathbf{W}$ has only $2M$ nonzero elements in special positions (see Eq. (3) and the example). That is, for the iterations of Eqs. (11) and (14), one often performs lots of multiplication-by-zero operations. By eliminating these unnecessary operations, we can reduce the computational burden substantially. Also we do not have to waste storage for the zero entries in the matrix $\mathbf{W}$. Instead of the $M \times M$ matrix $\mathbf{W}$, we may use the following two M-dimensional vectors;

$$\mathbf{u}(j) = \mathbf{W}([j/2], j)\Big|_{D=D_o, Z=1}$$

and

$$\mathbf{v}(j) = \mathbf{W}([j/2] + M/2, j)\Big|_{D=D_o, Z=1}, j = 0,1,\ldots,M-1 \tag{16}$$

where $[a]$ denotes the integer part of a. Let

$$\tilde{\mathbf{f}}_\ell = \mathbf{f}_\ell\Big|_{D=D_o, Z=1} \text{ and } \tilde{\mathbf{g}}_\ell = \mathbf{g}_\ell\Big|_{D=D_o, Z=1} \tag{17}$$

Then we can iterate $\tilde{\mathbf{f}}_\ell$ and $\tilde{\mathbf{g}}_\ell$ for $\ell = 1,2,\ldots,$ with $\tilde{\mathbf{f}}_o = \tilde{\mathbf{g}}_o \triangleq (\mathbf{u}(1),0,0,\ldots,0)$, as;

$$\begin{aligned} \tilde{\mathbf{f}}_{\ell+1}(1) &= \tilde{\mathbf{f}}_\ell(M/2) \cdot \mathbf{v}(1) \\ \tilde{\mathbf{g}}_{\ell+1}(1) &= \tilde{\mathbf{g}}_\ell(M/2) \cdot \mathbf{v}(1) + \tilde{\mathbf{f}}_{\ell+1}(1) \end{aligned} \tag{18}$$

and for $j = 2,3,\ldots,M-1$,

$$\tilde{\mathbf{f}}_{\ell+1}(j) = \tilde{\mathbf{f}}_{\ell}([j/2]) \cdot \mathbf{u}(j) + \tilde{\mathbf{f}}_{\ell}([j/2] + M/2) \cdot \mathbf{v}(j)$$

$$\tilde{\mathbf{g}}_{\ell+1}(j) = \tilde{\mathbf{g}}_{\ell}([j/2]) \cdot \mathbf{u}(j) + \tilde{\mathbf{g}}_{\ell}([j/2] + M/2) \cdot \mathbf{v}(j)$$

$$+ \tilde{\mathbf{f}}_{\ell+1}(j) \cdot \mathrm{mod}(j,2) \qquad (19)$$

Finally, for the transfer function bound, we have a recursive solution which requires only vector operations;

$$\left.\frac{\partial}{\partial Z} T(D,Z)\right|_{D=D_o, Z=1} = \left\{\sum_{\ell=0}^{\infty} \tilde{\mathbf{g}}_{\ell}(M/2)\right\} \cdot \mathbf{v}(0) \qquad (20)$$

Note that we need to truncate (20) at some depth. One may choose the stopping number L such that

$$\tilde{\mathbf{g}}_{L}(M/2) < 10^{-6} \cdot \left\{\sum_{\ell=0}^{L} \tilde{\mathbf{g}}_{\ell}(M/2)\right\} \qquad (21)$$

This gives 4 or more digits of accuracy for most cases of interest. Also notice that if $\mathbf{G}(n,K) = 1$ for all $n = 1,2,\ldots,N$, then we need only one of $\mathbf{u}$ or $\mathbf{v}$, since $\mathbf{u}(j) \cdot \mathbf{v}(j) = D^N$ for all j. (The same is true if $\mathbf{G}(n,1)$ are all n, by redefining the states in reverse order.)

IV. Applications, Discussion, and Conclusions

In the previous section we presented a technique for finding the transfer function bound using only vectors. Hence we can apply this technique for rather long constraint length codes very efficiently. This technique was used to compute the performance of reported codes in Refs. 4 through 8 for $r = 1/2$ codes with $K \leqslant 18$, $r = 1/3$ codes with $K \leqslant 16$, and $r = 1/4$ codes with $K \leqslant 14$. With a given K and r, we picked the best code, using a criterion of minimizing required E_b/N_o for a given desired BER of 10^{-6}. Such codes are tabulated in Table 1 and their performances are shown in Figs. 4 through 6. The vertical lines in those figures are the computational cutoff rate limit for the corresponding code rate ($E_b/N_o = -1/r \cdot \ln\{2^{1-r} - 1\}$). Note that the transfer function bound becomes loose as the operating SNR approaches the cutoff rate limit.

Notice that the code which is the best by our criterion may not have the maximum free distance. The transfer function bound itself is often represented by a series expansion as (Refs. 1 and 2)

$$\left.\frac{\partial}{\partial Z} T(D,Z)\right|_{D=D_o, Z=1} = \sum_{i=d_f}^{\infty} a_i D_o^i \qquad (22)$$

The first term of Eq. (22) may be used to determine the code performance if the value of operating D_o is extremely small. However, the value of operating D_o may not be so small in practice. Hence, rather than a single term approximation, a several-term truncated version of Eq. (22) is often used as an approximation to the transfer function bound. The coefficients a_i's (often called "weight spectra" or "number of bit errors in the adversaries") are needed for such approximations. For example, in Ref. 4, Odenwalder found the first 8 a_i's for his own codes, while in Ref. 9 Conan found the first 18 a_i's for the codes in Refs. 4 and 5. The technique of reducing required storage and computational effort described in the previous section can be applied for finding the a_i's also. However, note that the number of required terms for a good approximation varies with the value of operating D_o and the a_i's themselves. That is, a larger number of terms is required when the operating signal-to-noise ratio is small, when the code rate is low, and/or when the values of a_i's are large. As an illustration, in Fig. 7, 8-term (dotted lines) and 18-term (dashed lines) approximations are compared with our results (solid lines) for six codes considered in Ref. 9 with $r = 1/2$, $1/3$, and $1/4$, and $K = 7$ and 11. Notice that our method of finding the transfer function bound with the truncation rule (21) provides "uniform accuracy" for all cases considered.

Finding a_i's and using them for the performance evaluation of a code is very useful, if enough terms are provided for an accurate approximation. However, comparing two codes using those a_i's may not be practical, since vector comparison is not trivial. For example, consider the (7,1/3) case. Although the criterion for a good code in the code search in Ref. 4 was the maximum d_f, $d_f = 15$ codes were overlooked and a $d_f = 14$ code was found. Its transfer function bound is approximated as

$$\begin{aligned}\left.\frac{\partial}{\partial Z} T(D,Z)\right|_{Z=1} &= D^{14} + 20D^{16} + 53D^{18} + 184D^{20} + 555D^{22} \\ &\quad + 1961D^{24} + 6384D^{26} + 20655D^{28} \\ &\quad + 64598D^{30} + 203027D^{32} + 631873D^{34} \\ &\quad + 1958874D^{36} + 6028601D^{38} \\ &\quad + 18460857D^{40} + \ldots\end{aligned}$$

Later, a $d_f = 15$ code was found in Ref. 5 whose transfer function bound is approximated as

$$\begin{aligned}\left.\frac{\partial}{\partial Z} T(D,Z)\right|_{Z=1} &= 11D^{15} + 16D^{16} + 19D^{17} + 28D^{18} + 55D^{19} \\ &\quad + 96D^{20} + 169D^{21} + 338D^{22} + 636D^{23}\end{aligned}$$

$$+ 1276D^{24} + 2172D^{25} + 3628D^{26}$$
$$+ 6580D^{27} + 12048D^{28} + 20820D^{29}$$
$$+ 36358D^{30} + 65009D^{31} + 115368D^{32}$$
$$+ 204997D^{33} + 356650D^{34} + 622913D^{35}$$
$$+ 1097466D^{36} + 1924564D^{37}$$
$$+ 3356610D^{38} + 5848017D^{39}$$
$$+ 10215732D^{40} + 17821463D^{41} + \ldots$$

With the values of d_f and a_i's only, we cannot compare the two codes. The proper way of comparing them is substituting the actual value of D_o for D and comparing the summations, which is nothing but the evaluation of the transfer function bound.

In conclusion, we have described a technique for calculating the transfer function bound on the BER at the Viterbi decoder output which requires only vector operations for the matrix inversion. Using this technique on previously reported codes of some selected code rates and constraint lengths, we determined codes which require minimum SNR for a given desired BER of 10^{-6}, and provided their BER performance curves. As compared to the series expansion approximation method, our technique is shown to be preferable, since it gives a better approximation of the actual transfer function bound and it can be used directly for the comparison of codes.

References

1. A. J. Viterbi, "Convolutional Codes and Their Performance in Communication Systems," *IEEE Trans. Commun. Tech.*, Vol. COM-19, pp. 751-772, October 1971.
2. A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill Book Co., New York, NY, 1979.
3. J. K. Omura and M. K. Simon, *Modulation/Demodulation Techniques for Satellite Communications, Part IV. Appendices*, JPL Publication 81-73, Jet Propulsion Laboratory, Pasadena, CA, Nov. 1981.
4. J. P. Odenwalder, *Optimal Decoding of Convolutional Codes*, Ph. D. Thesis, Dept. Sys. Sci., Univ. Calif., L.A., Los Angeles, CA, 1970.
5. K. L. Larsen, "Short Convolutional Codes with Maximal Free Distance for Rate 1/2, 1/3, and 1/4," *IEEE Trans. Infor. Theory*, Vol. IT-19, pp. 371-372, May 1973.
6. S. J. Cury, *Selection of Convolutional Codes Having Large Free Distance*, Ph. D. Thesis, Dept. Sys. Sci., Univ. Calif., L.A., Los Angeles, CA, 1971.
7. R. Johannesson and E. Paaske, "Further Results on Binary Convolutional Codes with an Optimal Distance Profile," *IEEE Trans. Infor. Theory*, Vol. IT-24, pp. 264-268, March 1978.
8. P. J. Lee, "New Short Constraint Length, Rate $1/N$ Convolutional Codes Which Minimize Required E_b/N_o for Given Bit Error Rate," *JPL TDA Progress Report* 42-77, pp. 41-56, May 1984.
9. J. Conan, "The Weight Spectra of Some Short Low-Rate Convolutional Codes," *IEEE Trans. Commun.*, Vol. COM-32, pp 1050-1053, Sept. 1984.

Table 1. Some of the best known convolutional codes, which require minimum E_b/N_o for desired BER of 10^{-6}

(K,r)	d_f	Found in Ref.	Code Generator G in octal			
(3, 1/2)	5	4	7	5		
(4, 1/2)	6	4	17	15		
(5, 1/2)	7	4	35	23		
(6, 1/2)	8	8	77	45		
(7, 1/2)	10	4	171	133		
(8, 1/2)	10	4	247	371		
(9, 1/2)	12	4	561	753		
(10, 1/2)	12	7	1755	1363		
(11, 1/2)	14	7	3645	2671		
(12, 1/2)	15	7	7173	5261		
(13, 1/2)	16	7	12767	16461		
(14, 1/2)	16	7	22555	37457		
(15, 1/2)	18	7	63121	55367		
(16, 1/2)	19	7	111653	145665		
(17, 1/2)	20	7	347241	246277		
(18, 1/2)	20	7	506477	673711		
(3, 1/3)	8	4	7	7	5	
(4, 1/3)	10	4	17	15	13	
(5, 1/3)	12	4	37	33	25	
(6, 1/3)	13	4	75	53	47	
(7, 1/3)	14	4	171	145	133	
(8, 1/3)	16	4	225	331	367	
(9, 1/3)	18	5	557	663	711	
(10, 1/3)	19	6	1765	1631	1327	
(11, 1/3)	22	5	2353	2671	3175	
(12, 1/3)	24	5	4767	5723	6265	
(13, 1/3)	24	5	10533	10675	17661	
(14, 1/3)	25	6	37515	33457	20553	
(15, 1/3)	26	6	77233	67175	41327	
(16, 1/3)	28	6	172465	156371	102657	
(3, 1/4)	10	8	7	7	5	5
(4, 1/4)	12	8	17	15	13	11
(5, 1/4)	15	8	37	35	25	23
(6, 1/4)	18	8	77	73	55	45
(7, 1/4)	20	8	175	151	133	117
(8, 1/4)	22	5	235	275	313	357
(9, 1/4)	23	5	463	535	733	745
(10, 1/4)	27	5	1117	1365	1633	1653
(11, 1/4)	29	5	2327	2353	2671	3175
(12, 1/4)	32	5	4767	5723	6265	7455
(13, 1/4)	33	5	11145	12477	15573	16727
(14, 1/4)	36	5	21113	23175	35527	35537

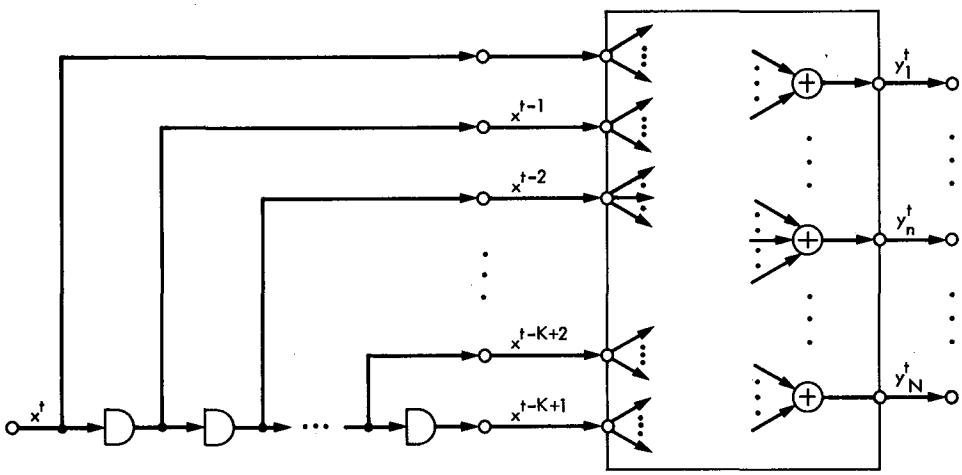


Fig. 1. A nonsystematic, constraint length *K*, rate 1/*N* convolutional encoder structure

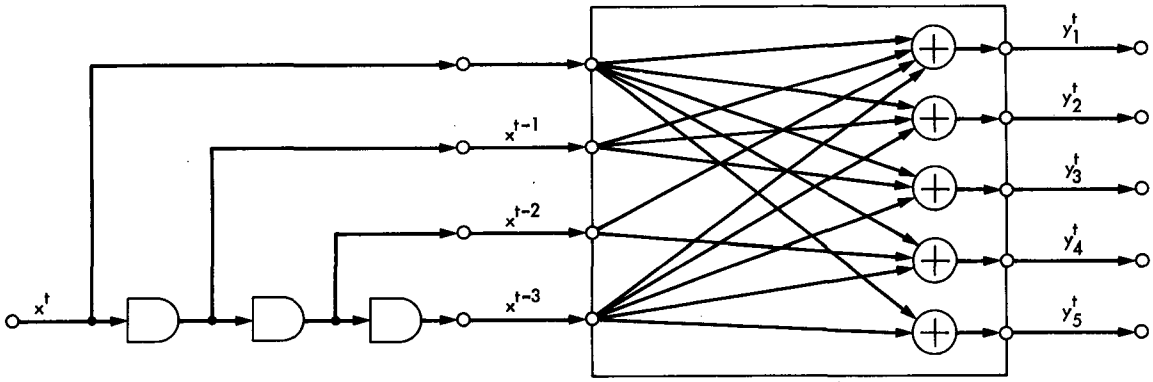


Fig. 2. A (4,1/5) convolutional encoder structure

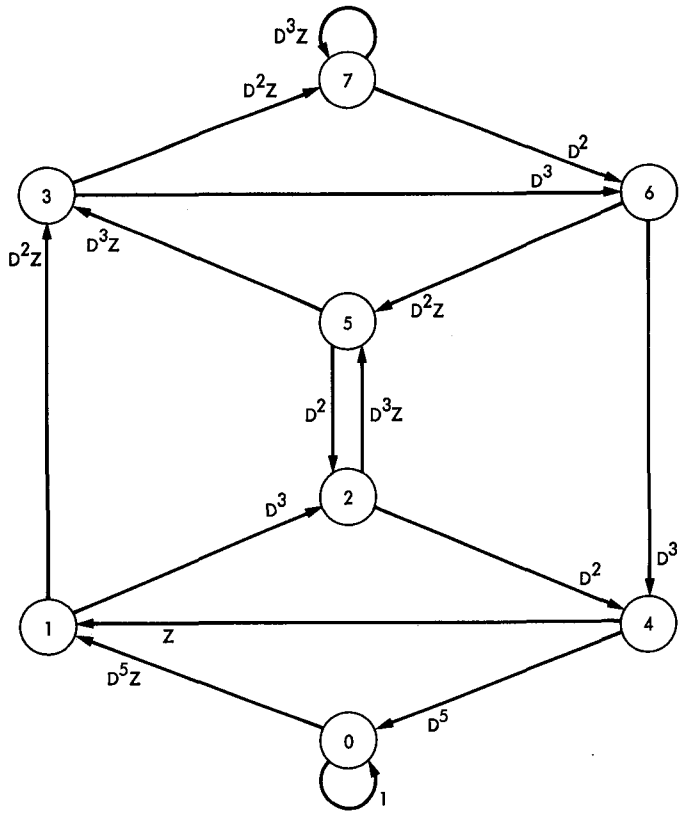


Fig. 3. The state diagram of the encoder in Fig. 2

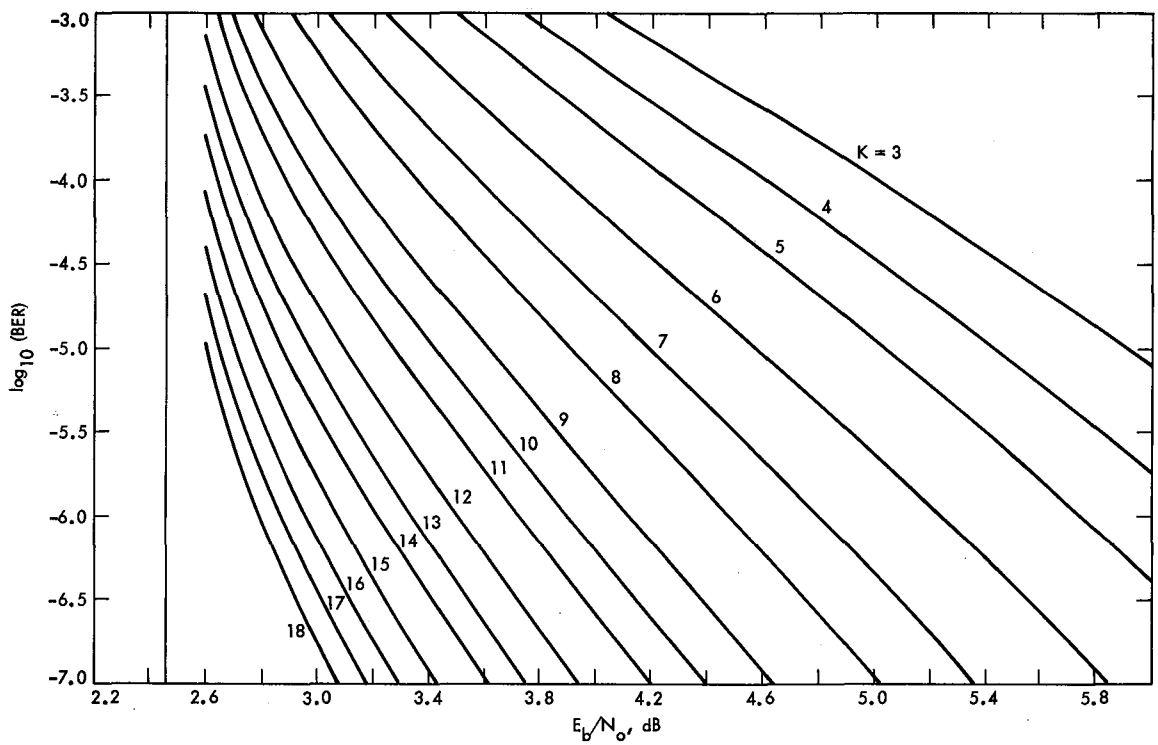


Fig. 4. Transfer function bounds on BER at the output of Viterbi decoder with best known rate 1/2 codes

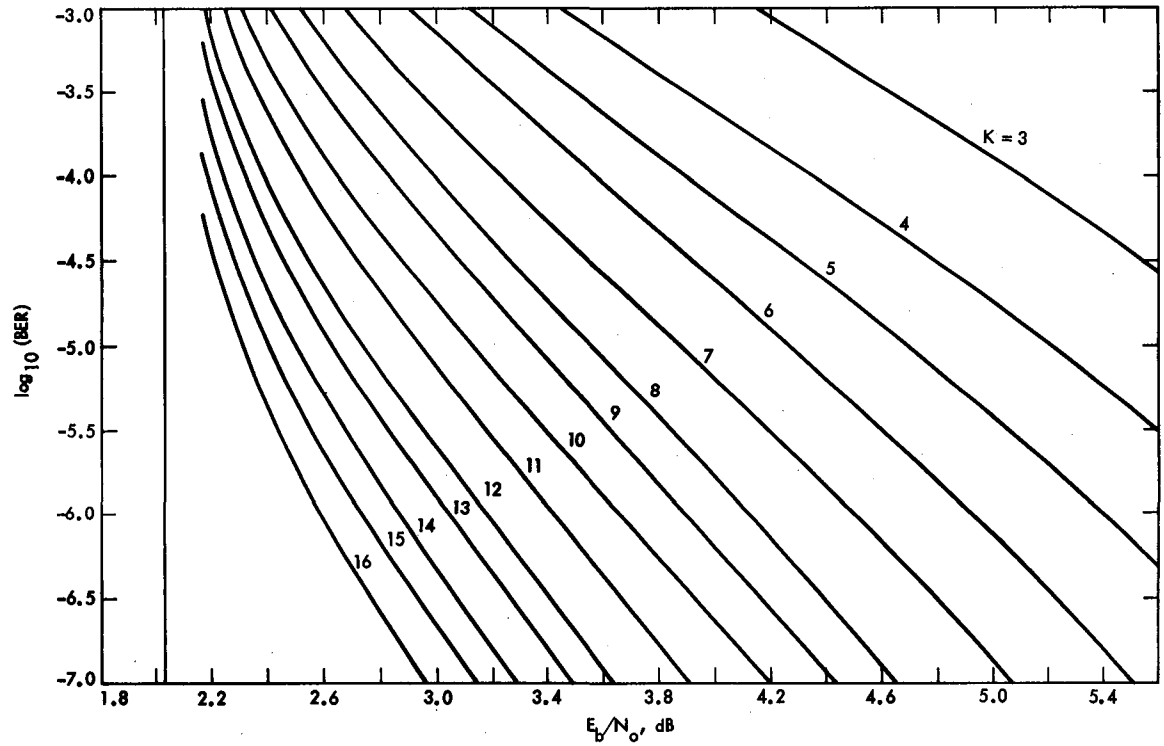


Fig. 5. Transfer function bounds on BER at the output of Viterbi decoder with best known rate 1/3 codes

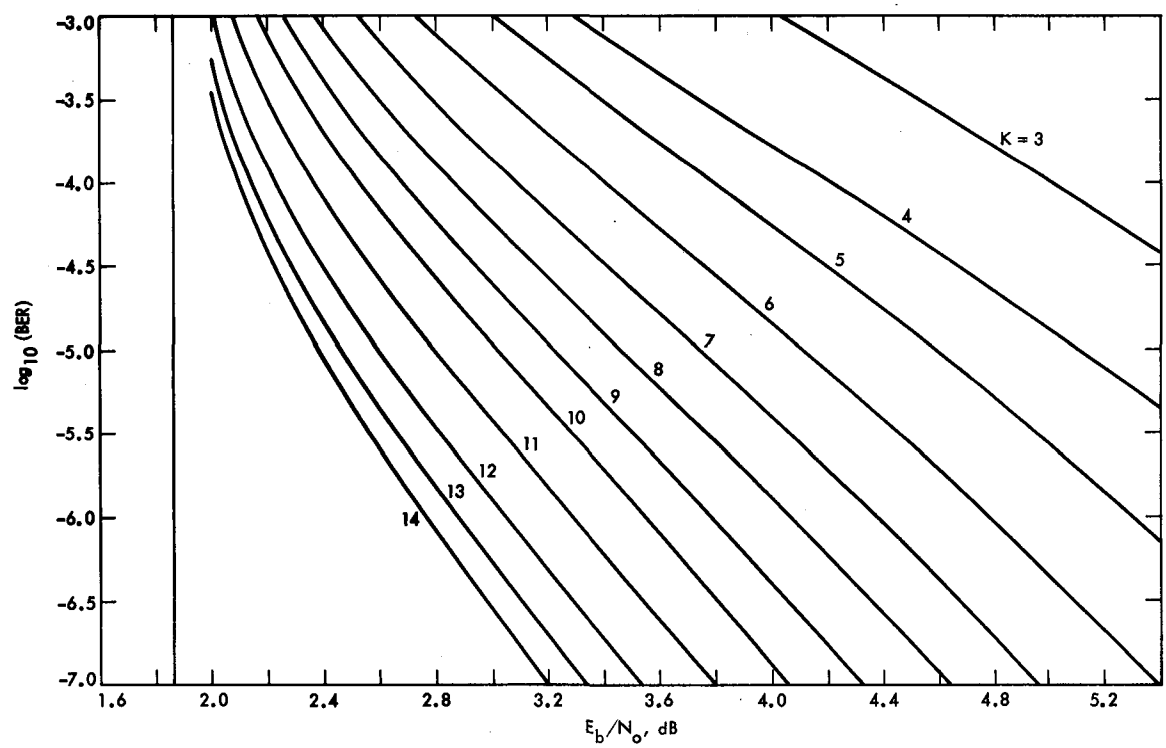


Fig. 6. Transfer function bounds on BER at the output of Viterbi decoder with best known rate 1/4 codes

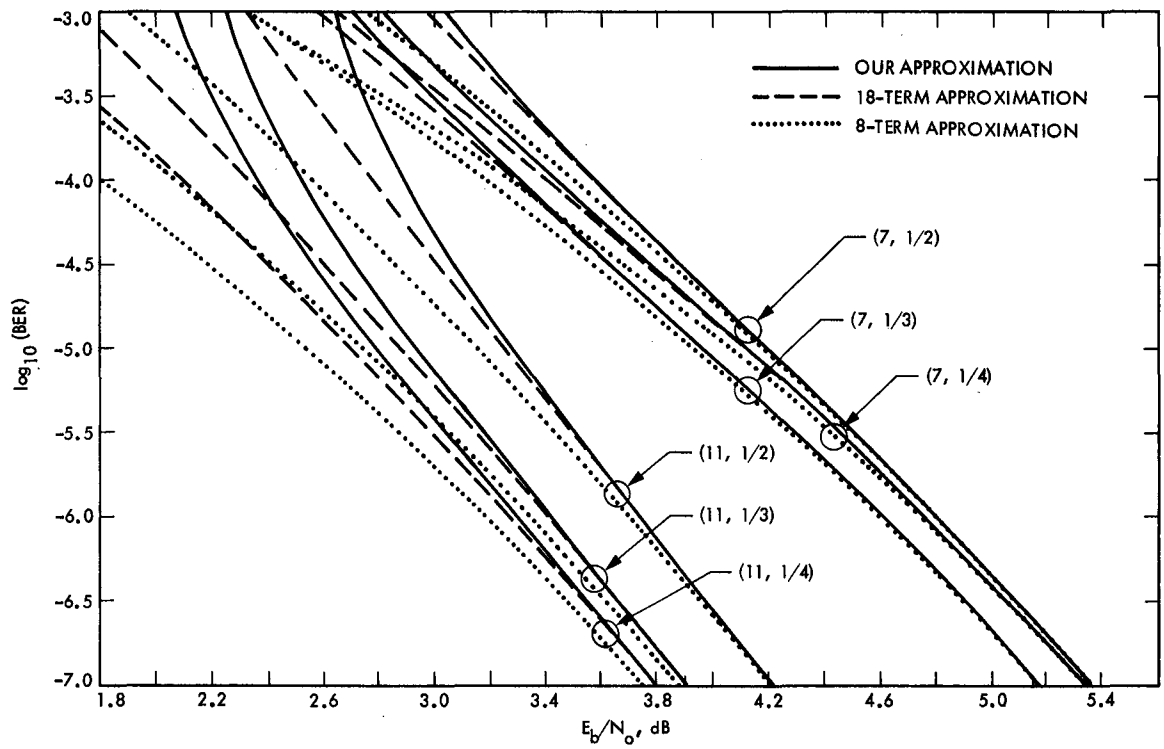


Fig. 7. Comparisons of approximations on transfer function bounds

Sequential Syndrome Decoding of Convolutional Codes

I. S. Reed
University of Southern California

T. K. Truong
Communications Systems Research Section

This paper reviews previous studies (Refs. 1 and 2) of the algebraic structure of convolutional codes and extends those studies to apply to sequential syndrome decoding. These concepts are then used to realize by example actual sequential decoding, using the stack algorithm.

I. Introduction

In Ref. 1, a general technique was developed for finding all solutions of the syndrome equation of noncatastrophic (n, k) convolution codes (CCs). The solutions of the syndrome equation constitute the set of all possible error sequences that might have been made in transmission. In general, these solutions of the syndrome equation are either graphed on a error tree or on its more compact equivalent, an error trellis. In Ref. 2, the identity of the mathematical concept of Vinck, de Paepe, and Schalkwijk (VPS) (Ref. 3) is first generalized to all basic encoders and proved rigorously by the general technique developed in Ref. 1. This identity represents a canonical solution of the syndrome equation for all (n, k) convolutional codes with a basic encoder.

In this paper, the Fano metric for use in sequential decoding is modified so that it can be utilized to sequentially find the minimum weight error sequence in the set (or coset) of all solutions of the syndrome equation. To accomplish this, the stack algorithm is used primarily to expose the concepts of sequential syndrome decoding. Sequential syndrome decoding of CCs is illustrated in detail by an example. However, most of the present development applies as well to a Fano-like sequential syndrome decoding algorithm.

II. Fundamentals of Syndrome Decoding of Convolutional Codes

This section provides a brief review of the concepts of a CC needed for the systematic construction of an error tree or trellis of a sequential syndrome decoding. First, the input to an (n, k) CC can be represented as the D-transform

$$x(D) = \sum_{j=0}^{\infty} \mathbf{x}_j D^j \tag{1}$$

of the sequence $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \cdots$, of k-vectors of form $\mathbf{x}_j = [x_{1j}, x_{2j}, \cdots, x_{kj}]$, where x_{ij} belong to a Galois field, F, here restricted to the binary field of two elements. Similarly, the ouptut of an (n, k) CC is the D-transform

$$y(D) = \sum_{j=0}^{\infty} \mathbf{y}_j D^j \tag{2}$$

where $\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, \cdots$, constitutes a sequence of n-vectors $\mathbf{y}_j = [y_{1j}, y_{2j}, \cdots, y_{nj}]$, with $y_{ij} \in F$.

The input and output of a convolutional code are linearly related in terms of the operations of the symbol field, F. As a consequence, $y(D)$ in Eq. (2) is related, in general, to $x(D)$ in Eq. (1) by

$$y(D) = x(D)\mathbf{G(D)} \tag{3}$$

where $\mathbf{G(D)}$ is a $k \times n$ matrix of formal power series in D over the symbol field, F.

If the elements $g_{ij}(D)$ of $k \times n$ matrix

$$\mathbf{G(D)} = \begin{bmatrix} \mathbf{g}_{11}(\mathbf{D}), \cdots, \mathbf{g}_{1n}(\mathbf{D}) \\ \vdots \qquad\qquad \vdots \\ \mathbf{g}_{k1}(\mathbf{D}), \cdots, \mathbf{g}_{kn}(\mathbf{D}) \end{bmatrix} \tag{4}$$

are restricted to be polynomials, $\mathbf{G(D)}$ is the generating matrix of some (n, k) convolutional code. The maximum degree, n, of the polynomial elements of $\mathbf{G(D)}$ is called the memory delay of the code. Usually, the constraint length of the code is defined to be $K = n + 1$.

The polynomial elements of the generating matrix $\mathbf{G(D)}$ belong to the ring, $F[D]$, of all polynomials in D over the finite field, F.

$$\mathbf{G(D)} = \mathbf{A(D)}[\Gamma_k, 0]\,\mathbf{B(D)} \tag{5}$$

To avoid what is called "catastrophic error propagation," Massey and Sain (Ref. 4) proved that the right inverse $\mathbf{G}^{-1}$ of the generating matrix $\mathbf{G}$ must be feedback-free. Forney (Ref. 5) defined a *basic* encoder to be a CC that has a feedback-free inverse $\mathbf{G}^{-1}$ of generating matrix $\mathbf{G}$. Also, Forney in Ref. 5 (see also Ref. 1) showed that only basic encoders with Smith normal form

$$\mathbf{G(D)} = \mathbf{A}[I_k, 0]\,\mathbf{B} \tag{6}$$

will be considered. Where $\mathbf{A} = \mathbf{A(D)}$ is a $k \times k$ matrix with elements in $F[D]$, $\mathbf{B} = \mathbf{B(D)}$ is an $n \times n$ matrix with elements in $F[D]$, and I_k is a $k \times k$ diagonal matrix.

After transmission over a possible noisy channel, let

$$z(D) = y(D) + e(D) \tag{7}$$

be the D-transform of the received coded message, where $e(D)$ is the D-transform of the error sequence. A parity-check matrix, $\mathbf{H} = \mathbf{H(D)}$, of the generation matrix, $\mathbf{G(D)}$, is any full-rank $(n - k) \times n$ matrix with elements in $F[D]$ such that

$$\mathbf{G(D)H}^T\mathbf{(D)} = 0 \tag{8}$$

To find the parity-check matrix, $\mathbf{H}$, associated with $\mathbf{G}$, the method of Forney (Ref. 5) is used. To accomplish this, partition matrix $\mathbf{B}$ of the Smith normal form in Eq. (6) and its inverse, $\mathbf{B}^{-1}$, in the following manner. Let

$$\mathbf{G} = [\mathbf{B}_1, \mathbf{B}_2]^T \tag{9}$$

where T denotes matrix transpose; $\mathbf{B}_1$ and $\mathbf{B}_2$ are the first k rows and the last $(n - k)$ rows of $\mathbf{B}$, respectively.

Similarly, let

$$\mathbf{B}^{-1} = [\bar{\mathbf{B}}_1, \bar{\mathbf{B}}_2] \tag{10}$$

where $\bar{\mathbf{B}}_1$ and $\bar{\mathbf{B}}_2$ are the first k columns and the last $(n - k)$ columns of $\mathbf{B}^{-1}$, respectively. Then, since $\mathbf{B}\mathbf{B}^{-1} = \mathbf{I}_n$, the following identities hold:

$$\begin{aligned} \mathbf{B}_1\bar{\mathbf{B}}_1 &= \mathbf{I}_k, & \mathbf{B}_1\bar{\mathbf{B}}_2 &= 0 \\ \mathbf{B}_2\bar{\mathbf{B}}_1 &= 0, & \mathbf{B}_2\bar{\mathbf{B}}_2 &= \mathbf{I}_{n-k} \end{aligned} \tag{11}$$

In terms of the above matrix partitions, the Forney parity-check matrix is defined by

$$\mathbf{H} = \bar{\mathbf{B}}_2^T \tag{12}$$

where "T" denotes matrix transpose. That $\mathbf{H}$, defined by Eq. (12), is a parity-check matrix satisfying Eq. (8) and is verified by substituting for $\mathbf{G}$ its Smith normal form and by partitioning $\mathbf{B}$ as given in Eq. (9) (see Ref. 1 for more details).

The syndrome $s(D)$ of the received sequence, corresponding to $z(D)$ in Eq. (7), is computed by

$$s = s(D) = z(D)\mathbf{H}^T\mathbf{(D)} \tag{13}$$

where $\mathbf{H(D)}$ is the parity-check matrix in Eq. (12). A substitution of $z(D)$, which appears in Eq. (7), into Eq. (13) yields

$$s(D) = e(D)\mathbf{H}^T\mathbf{(D)} \tag{14}$$

as the syndrome, computed by Eq. (13), but now in terms of only $e(D)$, the error sequence. This result shows that the

syndrome $s(D)$ is completely independent of the transmitted code sequence, $y(D)$.

Syndrome decoding of convolutional codes depends first, as it does of block codes, on solving the syndrome equation, Eq. (14), for all possible error sequence $e(D) \in F[D]$ that might have given rise to the syndrome $s(D)$ as computed by Eq. (17). Next, that error sequence, $\hat{e}(D)$, is chosen from that set of solutions of Eq. (14) that maximizes the likelihood of being close to the actual error sequence.

The general solution of the syndrome equation, Eq. (14), is given in Eq. (23) of Ref. 2. That is,

$$e = t\mathbf{G} + s\mathbf{B}_2 \tag{15}$$

where $\mathbf{B} = (\mathbf{H}^{-1})^T$, $\mathbf{G}$ is the $k \times n$ generating matrix, $\mathbf{B}_2$ is computed by Eq. (9), s is the $(n-k)$ component computed by Eq. (13), and t is an arbitrary k vector with elements in $F[D]$.

In the next section, the solution $e(D)$ of the syndrome equation, Eq. (14), is shown to graph on a rooted tree or trellis as a function of all possible binary sequences corresponding to the arbitrary D-transform $t(D)$. Such an error tree or trellis is used then to illustrate the sequential syndrome decoding algorithm for CCs using a modified Fano metric, μ_F, developed in Appendix A.

Now let $\hat{e} = \hat{e}(D)$ be the D-transform of the error sequence found by maximizing μ_F over all subsequences of the error trellis generated by the particular sequential syndrome decoding algorithm developed in Section III. Also, let $\hat{t} = \hat{t}(D)$ be the D-transform of the binary sequence path in the error trellis along which sequence $\hat{e}$ was found. By Eq. (23), $\hat{t}$ and $\hat{e}$ are related, in fact, by

$$\hat{e} = \hat{t}\mathbf{G} + s\mathbf{B}_2 \tag{16}$$

Next, note, by the Smith normal form in Eq. (6) of a generating matrix $\mathbf{G}$, that

$$\mathbf{G}^{-1} = \mathbf{B}^{-1}\begin{bmatrix}\mathbf{I}_k\\ 0\end{bmatrix}\mathbf{A}^{-1} \tag{17}$$

is the *right* inverse of $\mathbf{G}$. Hence, multiplying both sides of Eq. (16) yields, by Eq. (11), the identity,

$$\begin{aligned}\hat{e}\mathbf{G}^{-1} &= \hat{t} + s \cdot \mathbf{B}_2 \cdot \mathbf{G}^{-1}\\ &= \hat{t} + s\mathbf{B}_2[\bar{\mathbf{B}}_1, \bar{\mathbf{B}}_2]\begin{bmatrix}\mathbf{I}_k\\ 0\end{bmatrix}\mathbf{A}^{-1}\\ &= \hat{t} + s[0, \mathbf{I}_{n-k}]\begin{matrix}\mathbf{I}_k\\ 0\end{matrix}\mathbf{A}^{-1} = \hat{t}\end{aligned} \tag{18}$$

between $\hat{e}$ and $\hat{t}$.

The sequential syndrome decoding algorithm produces the D-transform $\hat{e}$ of the most likely estimate of the actual error sequence. Hence, by Eq. (7), a subtraction of $\hat{e}$ from z yields an estimate, $\hat{y}$, of the actual transmitted sequence y. But, $\hat{y}$, in turn, is generated by an estimated message sequence $\hat{x}$. That is,

$$\hat{y} = \hat{x}\mathbf{G} = z + \hat{e} \tag{19}$$

Multiplying both sides of the latter equality by $\mathbf{G}^{-1}$ produces, by Eq. (18),

$$\hat{x} = z\mathbf{G}^{-1} + \hat{e}\mathbf{G}^{-1} = z\mathbf{G}^{-1} + \hat{t} \tag{20}$$

This identity shows that $\hat{t}$, computed by the sequential syndrome algorithm, is a *correction factor* to the standard technique for recovering the message from its coded form in the noiseless case.

III. Sequential Syndrome Decoding with a Stack Algorithm

In this section, the details of a sequential syndrome decoding algorithm are developed by an example of the basic encoder. This encoder is the same as the (3, 1) CC described (Ref. 6, Chapter 12) in the development of the standard stack algorithm for the sequential decoding of convolutional codes. For expository and comparative purposes, some of the parameters in the example (Ref. 6, Chapter 12) are used also in this paper.

To illustrate a sequential syndrome decoding with a stack algorithm, consider the generating matrix

$$\mathbf{G}(\mathbf{D}) = [1 + \mathbf{D}, 1 + \mathbf{D}^2, 1 + \mathbf{D} + \mathbf{D}^2] \tag{21}$$

of a (3, 1) CC. Using elementary column transformations, the Smith normal form of $\mathbf{G}$ in Eq. (21) is

$$\mathbf{G} = [1\ 0\ 0]\,\mathbf{B}(\mathbf{D}) \tag{22a}$$

where

$$\mathbf{B}(\mathbf{D}) = \begin{bmatrix}1 + \mathbf{D}, & 1 + \mathbf{D}^2, & 1 + \mathbf{D} + \mathbf{D}^2\\ 0 & 1 & 1\\ 1 & 0 & 1\end{bmatrix} \tag{22b}$$

With the same elementary transformations needed to obtain Eq. (22a) applied in reverse order, the inverse of $\mathbf{B(D)}$ is obtained as

$$\mathbf{B}^{-1}(\mathbf{D}) = \begin{bmatrix} 1, & 1+\mathbf{D}^2, & \mathbf{D} \\ 1, & \mathbf{D}^2, & 1+\mathbf{D} \\ 1, & 1+\mathbf{D}^2, & 1+\mathbf{D} \end{bmatrix} \tag{23}$$

Also from Eqs. (22b) and (9), one observes

$$\mathbf{B}_2 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Hence, the solution of the syndrome equation, Eq. (14), for this example is, by Eq. (15),

$$e = [e_1, e_2, e_3] = t[1+\mathbf{D}, 1+\mathbf{D}^2, 1+\mathbf{D}+\mathbf{D}^2] + [s_1, s_2]\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \tag{24}$$

where $t \in \mathbf{F[D]}$ and the syndrome $s = [s_1, s_2]$ is computed from Eqs. (13), (12), and (23) by

$$s = [s_1, s_2] = z\mathbf{H}^T = [z_1, z_2, z_3]\begin{bmatrix} 1+\mathbf{D}^2, & \mathbf{D}^2, & 1+\mathbf{D}^2 \\ \mathbf{D}, & 1+\mathbf{D}, & 1+\mathbf{D} \end{bmatrix}^T \tag{25}$$

Thus, by Eqs. (24) and (25), the two components of the syndrome are

$$\begin{aligned} s_1 &= (1+\mathbf{D}^2)z_1 + \mathbf{D}^2 z_2 + (1+\mathbf{D}^2)z_3 \\ s_2 &= \mathbf{D}z_1 + (1+\mathbf{D})z_2 + (1+\mathbf{D})z_3 \end{aligned} \tag{26}$$

and the three components of the solution of the syndrome equation are

$$\begin{aligned} e_1 &= (1+\mathbf{D})t + s_2 \\ e_2 &= (1+\mathbf{D}^2)t + s_1 \\ e_3 &= (1+\mathbf{D}+\mathbf{D}^2)t + s_1 + s_2 \end{aligned} \tag{27}$$

For this example, let the input sequence be

$$x = [1\ 1\ 0\ 0\ 1] \tag{28a}$$

corresponding to its D-transform.

$$x = 1 + \mathbf{D} + \mathbf{D}^4 \tag{28b}$$

By Eqs. (3) and (21), the transmitted codeword is

$$y = [y_1, y_2, y_3] = [(1+\mathbf{D})x, (1+\mathbf{D}^2)x, (1+\mathbf{D}+\mathbf{D}^2)x] \tag{29}$$

The encoded transmission is obtained by substituting Eq. (28b) into Eq. (29) with the following calculations:

$$\begin{array}{lllllll} x: & 1 & 1 & 0 & 0 & 1 & \\ \mathbf{D}x: & & 1 & 1 & 0 & 0 & 1 \\ \hline y_1: & 1 & 0 & 1 & 1 & 1 & 1 \end{array} \qquad \begin{array}{lllllll} x: & 1 & 1 & 0 & 0 & 1 & \\ \mathbf{D}^2x: & & & 1 & 1 & 0 & 0\ 1 \\ \hline y_2: & 1 & 1 & 1 & 1 & 1 & 0\ 1 \end{array}$$

and

$$\begin{array}{lllllll} x: & 1 & 1 & 0 & 0 & 1 & \\ \mathbf{D}x: & & 1 & 1 & 0 & 0 & 1 \\ \mathbf{D}^2x: & & & 1 & 1 & 0 & 0\ 1 \\ \hline y_3: & 1 & 0 & 0 & 1 & 1 & 1\ 1 \end{array}$$

Hence,

$$\begin{aligned} y = {} & [1\ 1\ 1] + [0\ 1\ 0]\mathbf{D} + [1\ 1\ 0]\mathbf{D}^2 \\ & + [0\ 1\ 1]\mathbf{D}^3 + [1\ 1\ 1]\mathbf{D}^4 + [1\ 0\ 1]\mathbf{D}^5 \\ & + [0\ 1\ 1]\mathbf{D}^6 \end{aligned} \tag{30a}$$

is the D-transform of the encoded sequence

$$y = [1\ 1\ 1, 0\ 1\ 0, 1\ 1\ 0, 0\ 1\ 1, 1\ 1\ 1, 1\ 0\ 1, 0\ 1\ 1] \tag{30b}$$

Note that there is a one-to-one correspondence between sequences and their D-transforms. The final encoded sequence is obtained by multiplexing sequences y_1, y_2, and y_3. This example of encoding illustrates the point that a D-transform x operated on by $\mathbf{D}^k$, i.e., $\mathbf{D}^k x$, can be used interchangeably with sequence x, shifted right k times.

The circuit diagram for the (3, 1) CC encoder in Eq. (29) is shown in Fig. 1. This circuit is a linear finite-state machine, where the states of the machine are the four binary states of register (R_1, R_2). The state diagram of this machine is given in Fig. 2, where a dashed line corresponds to the input $x = 1$ and a bold line corresponds to the input $x = 0$. Finally, Fig. 3 is a graph of all possible inputs, next-state nodes, and outputs of

this state diagram as a function of discrete time. This is the trellis diagram of the (3, 1) CC encoder where the initial state is (0 0).

Let the received D-transform of code be after transmission of y in Eq. (30a),

$$z = [z_1, z_2, z_3] = [1\ 1\ 0] + [1\ 1\ 0]\mathbf{D} + [1\ 1\ 0]\mathbf{D}^2 + [1\ 1\ 1]\mathbf{D}^3 + [0\ 1\ 1]\mathbf{D}^4 + [1\ 0\ 1]\mathbf{D}^5 + [0\ 0\ 1]\mathbf{D}^6 \quad (31a)$$

so that

$$z_1 = [1\ 1\ 1\ 1\ 0\ 1\ 0], \qquad z_2 = [1\ 1\ 1\ 1\ 1\ 0\ 0] \quad (31b)$$

and

$$z_3 = [0\ 0\ 0\ 1\ 1\ 1\ 1] \quad (31b)$$

The syndrome $s = [s_1, s_2]$ for this received convolutional code is given in Eq. (26). Using Eq. (31b) in Eq. (26) yields

$$\begin{array}{rl}
z_1: & 1\ 1\ 1\ 1\ 0\ 1\ 0 \\
\mathbf{D}^2 z_1: & \quad 1\ 1\ 1\ 1\ 0\ 1\ 0 \\
\mathbf{D}^2 z_2: & \quad 1\ 1\ 1\ 1\ 1\ 0\ 0 \\
z_3: & 0\ 0\ 0\ 1\ 1\ 1\ 1 \\
\mathbf{D}^2 z_3: & \quad 0\ 0\ 0\ 1\ 1\ 1\ 1 \\
\hline
s_1 = & [1\ 1\ 1\ 0\ 1\ 1\ 1\ 0\ 1]
\end{array}$$

$$\begin{array}{rl}
\mathbf{D}z_1: & \quad 1\ 1\ 1\ 1\ 0\ 1\ 0 \\
z_2: & 1\ 1\ 1\ 1\ 1\ 0\ 0 \\
\mathbf{D}z_2: & \quad 1\ 1\ 1\ 1\ 1\ 0\ 0 \\
z_3: & 0\ 0\ 0\ 1\ 1\ 1\ 1 \\
\mathbf{D}z_3: & \quad 0\ 0\ 0\ 1\ 1\ 1\ 1 \\
\hline
s_2 = & [1\ 1\ 1\ 0\ 1\ 1\ 1\ 1\ 0]
\end{array}$$

Hence, the D-transform of syndrome is

$$s = [s_1, s_2] = [1\ 1] + [1\ 1]\mathbf{D} + [1\ 1]\mathbf{D}^2 + [1\ 1]\mathbf{D}^4 + [1\ 1]\mathbf{D}^5 + [1\ 1]\mathbf{D}^6 + [0\ 1]\mathbf{D}^7 + [1\ 0]\mathbf{D}^8 \quad (32a)$$

or, as a sequence, the syndrome is

$$s = [1\ 1, 1\ 1, 1\ 1, 0\ 0, 1\ 1, 1\ 1, 1\ 1, 0\ 1, 1\ 0] \quad (32b)$$

The syndrome $s(D)$ for this example is given above, in Eq. (32). Given $s(D)$, an error trellis or tree can now be constructed from the solution in Eq. (27) of the syndrome equation, Eq. (32). The error trellis for the solution, Eq. (27), of the syndrome equation is shown in Fig. 4, where $s = [s_1, s_2]$ is the syndrome sequence found in Eq. (32). By Eq. (27) and Fig. 4, the error trellis is constructed exactly like the trellis for the encoder in Fig. 3, except that input is $t(D)$ instead of message $x(D)$ and the output is the code $t\mathbf{G}$ *plus a correction factor*,

$$S\mathbf{B}_2 = [s_1, s_2, s_1 + s_2]$$

instead of the code only.

To illustrate how $[e_1, e_2, e_3]$ is computed at each branch of the trellis, suppose one is at node or state $b = 10$ at time 1 and let $t = 0$. The output at this branch of the code trellis in Fig. 3 is [1 0 1]. The syndrome at this branch is $[s_1, s_2] =$ [1 1], so that the correction factor is

$$[s_2, s_1, s_1 + s_2] = [1\ 1\ 0]$$

Hence, the output of the error trellis at this branch is

$$[e_1, e_2, e_3] = [1\ 0\ 1] + [1\ 1\ 0] = [0\ 1\ 1]$$

All other outputs of the error trellis are computed in precisely the same manner.

The error trellis in Fig. 4 will be used next to demonstrate sequential syndrome decoding using the stack algorithm. The metric used for this purpose is the modification of the Fano metric developed in Appendix A. Let $P(e_{ij}) = Prob[E_{ij} = e_{ij}]$ and $P(Z_{ij}) = Prob[Z_{ij} = z_{ij}]$ be, respectively, the probability that E_{ij}, the bit of the i-th coordinate of the error sequence e at time j, equals e_{ij} (0 or 1) and the probability that Z_{ij}, the bit of the i-th coordinate of the received sequence z at time j, equals z_{ij} (0 or 1). By Eq. (A-16), the Fano metric for some received sequence of an (n, k) CC of depth L is

$$\mu_F = \sum_{i=1}^{L}\sum_{j=1}^{n} \lambda(e_{ij}) \quad (33)$$

where

$$\lambda(e_{ij}) = \log_2 \frac{P(e_{ij})}{P(z_{ij})} - R \quad (34)$$

is the incremental change in the Fano metric at time stage i and coordinate j. The rate for this example in Eq. (34) is one-third. Also, let $P(e = 1) = 0.1$ and $P(e = 0) = 0.9$. Since

$$P(z_{ij}) = P(z_{ij}|\mu_{ij} = 0)P(\mu_{ij} = 0) + P(z_{ij}|\mu_{ij} = 1)P(\mu_{ij} = 0)$$

$$= \tfrac{1}{2}$$

for $z_{ij} = (0, 1)$, Eq. (34) becomes $\lambda(e_{ij}) = \log_2(2 \times 0.9) - 1/3 = 0.52$ if $e = 0$ and $\lambda(e_{ij}) = \log_2(2 \times 0.1) - 1/3 = 2.65$ if $e_{ij} = 1$. For simplicity, this increment of the Fano metric is often scaled so that its values are approximately integers. For the present case, if $\sigma(e_{ij}) = \lambda(e)/0.52$ then approximately

$$\sigma(e_{ij}) = 1 \quad \text{if} \quad e_{ij} = 0$$

$$= -5 \quad \text{if} \quad e_{ij} = 1 \qquad (35)$$

is the scaled version of the modification of Fano's metric for this example.

The stack algorithm for sequential syndrome decoding assumes that paths in the error trellis with their associated metrics are stored in a stack. The entries in the stack are ordered by the Fano metric. The path with the largest metric is placed on top and all other paths are listed in the order of decreasing metric. In detail, the stack algorithm for sequential syndrome decoding has the following steps:

(1) Load stack with starting node or state of error trellis or tree; the starting metric is zero.

(2) Compute metric of successors of path in the top of the stack.

(3) Delete path in the top of the stack.

(4) Insert the new paths in the stack, and resort paths in the stack in the order of decreasing metric.

(5) If the path at the top of stack ends at a terminating node, stop; otherwise, return to Step 2, above.

Next, the stack algorithm is illustrated below for the error trellis in Fig. 4.

In the following example of sequential syndrome decoding using the stack algorithm, each entry of a stack contains three times: a partial path in the error trellis, the path metric, and the Hamming weight of path. With this stack structure, the steps of the stack algorithm for the error trellis in Fig. 4 and the metric in Eq. (35) appear as follows:

Step 1

1 (-3) (1)
0 (-9) (2)

Step 2

1 1 (- 6) (2)
0 (- 9) (2)
1 0 (-12) (3)

Step 3

1 1 0 (- 3) (2)
0 (- 9) (2)
1 0 (-12) (3)
1 1 1 (-21) (5)

Step 4

1 1 0 1 (- 6) (3)
0 (- 9) (2)
1 0 (-12) (3)
1 1 0 0 (-12) (4)
1 1 1 (-21) (5)

Step 5

1 1 0 1 1 (- 9) (5)
0 (- 9) (2)
1 0 (-12) (3)
1 1 0 0 (-12) (4)
1 1 0 1 0 (-15) (6)
1 1 1 (-21) (5)

Step 6

1 1 0 1 1 0 (- 6) (4)
0 (- 9) (2)
1 0 (-12) (3)
1 1 0 0 (-12) (4)
1 1 0 1 0 (-15) (6)
1 1 1 (-21) (5)
1 1 0 1 1 1 (-24) (7)

Step 7

1 1 0 1 1 0 1 (- 9) (5)
0 (- 9) (2)
1 0 (-12) (3)
1 1 0 0 (-12) (4)
1 1 0 1 1 0 0 (-15) (6)
1 1 0 1 0 (-15) (6)
1 1 1 (-21) (5)
1 1 0 1 1 1 (-24) (7)

Step 8

1 1 0 1 1 0 1 0 (- 6) (5)
0 (- 9) (2)
1 0 (-12) (3)
1 1 0 0 (-12) (4)
1 1 0 1 1 0 0 (-15) (6)
1 1 0 1 0 (-15) (6)
1 1 1 (-21) (5)
1 1 0 1 1 0 1 1 (-24) (8)
1 1 0 1 1 1 (-24) (7)

Step 9

1 1 0 1 1 0 1 0 0 (- 3) (5)
0 (- 9) (2)
1 0 (-12) (3)
1 1 0 0 (-12) (4)
1 1 0 1 1 0 0 (-15) (6)
1 1 0 1 0 (-15) (6)
1 1 0 1 1 0 1 0 1 (-21) (8)
1 1 1 (-21) (5)
1 1 0 1 1 0 1 1 (-24) (8)
1 1 0 1 1 1 (-24) (7)

The above stack algorithm shows that the best estimate of path t is

$$\hat{t} = [1\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 0] \qquad (36)$$

The error sequence along this path is the estimate

$$\hat{e} = [0\ 0\ 1, 1\ 0\ 0, 0\ 0\ 0, 1\ 0\ 0,$$
$$1\ 0\ 0, 0\ 0\ 0, 0\ 1\ 0, 0\ 0\ 0, 0\ 0\ 0] \quad (37)$$

If $\hat{e}$ were added to sequence z in Eq. (31a), the result would be the original coded sequence y in Eq. (30). However, it is more efficient to recover an estimate $\hat{x}$ of the original message by using Eq. (20).

The right inverse of the generating matrix for the example in Eq. (21) is, by Eqs. (22), (23), and (17)

$$\mathbf{G}^{-1} = \begin{bmatrix} 1, & 1+\mathbf{D}^2, & \mathbf{D} \\ 1, & \mathbf{D}^2, & 1+\mathbf{D} \\ 1, & 1+\mathbf{D}^2, & 1+\mathbf{D} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad (38)$$

Hence, the estimate $\hat{x}$ of the original message is, by Eq. (20), for the example

$$\hat{x} = [z_1, z_2, z_3] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \hat{t}$$

$$= z_1 + z_2 + z_3 + t \quad (39)$$

where $\hat{t}$ is the message correction factor obtained from Eq. (36) and the sequential syndrome decoding algorithm is developed below.

A substitution of Eqs. (31) and (36) into Eq. (39) yields the computation of x as follows:

$$\begin{array}{rl} z_1\text{:} & 1\ 1\ 1\ 1\ 0\ 1\ 0 \\ z_2\text{:} & 1\ 1\ 1\ 1\ 1\ 0\ 0 \\ z_3\text{:} & 0\ 0\ 0\ 1\ 1\ 1\ 1 \\ t\text{:} & 1\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 0 \\ \hline x = & [1\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ 0] \end{array}$$

Hence,

$$\hat{x} = 1 + \mathbf{D} + \mathbf{D}^4 \quad (40)$$

is the estimate of the original message x. Since this agrees with Eq. (28), the sequential syndrome decoding algorithm correctly decoded z as given in Eq. (31).

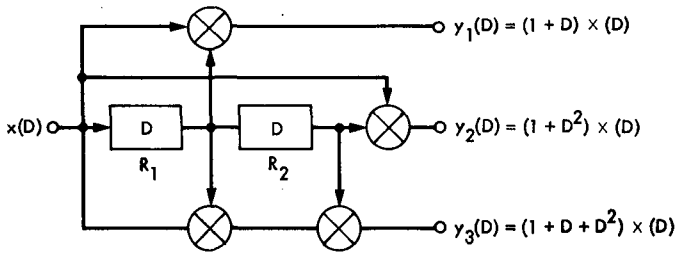


Fig. 1. Encoder circuit for (3, 1) convolutional code

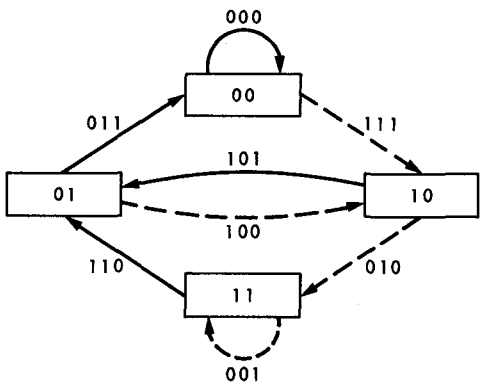


Fig. 2. State diagram for (3, 1) convolutional code (dashed lines correspond to x = 1; solid lines correspond to x = 0)

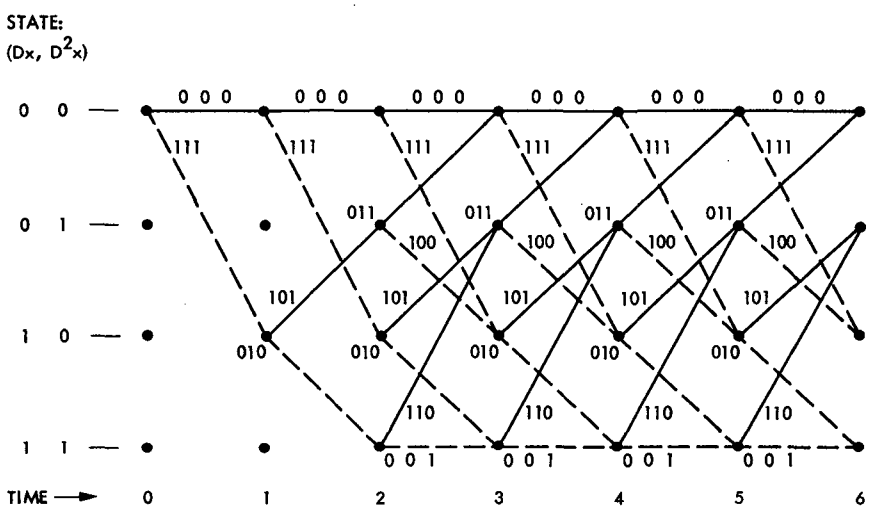


Fig. 3. Trellis diagram for encoder (3, 1) convolutional code (dashed lines correspond to x = 1; solid lines correspond to x = 0)

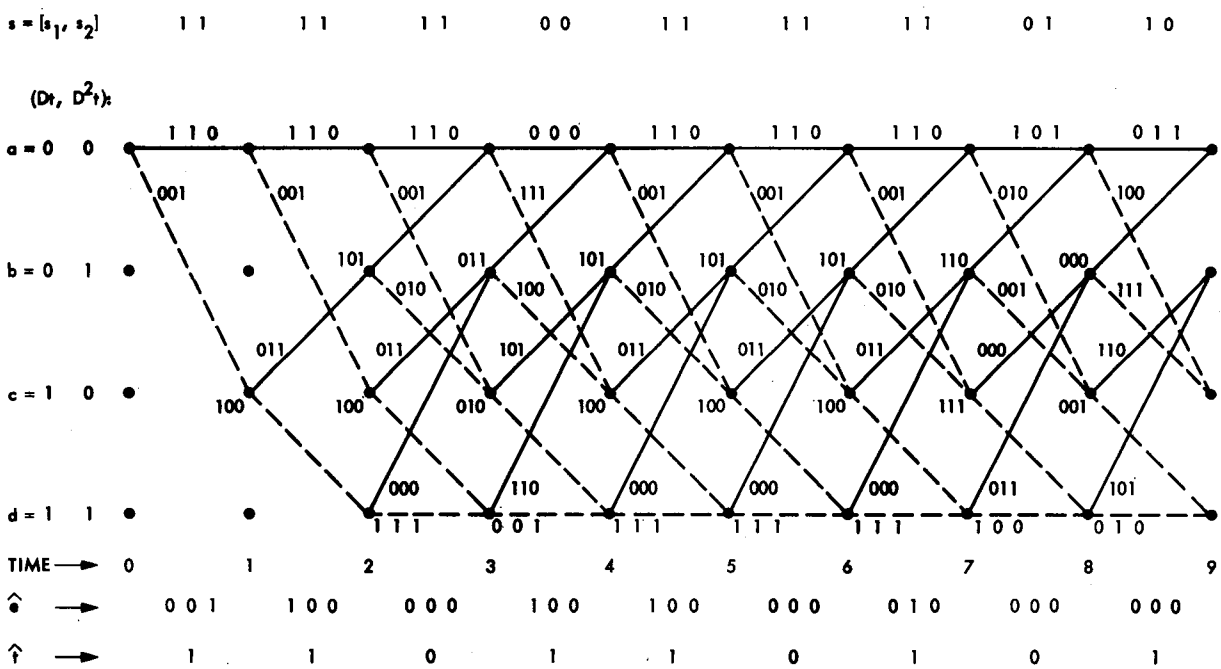


Fig. 4. Error trellis of $e = [e_1, e_2, e_3]$ as a function of t, state (Dt, D^2t), syndrome s, and stage k. Estimates of error sequence $\hat{e}$ and path $\hat{t}$ are shown below the trellis

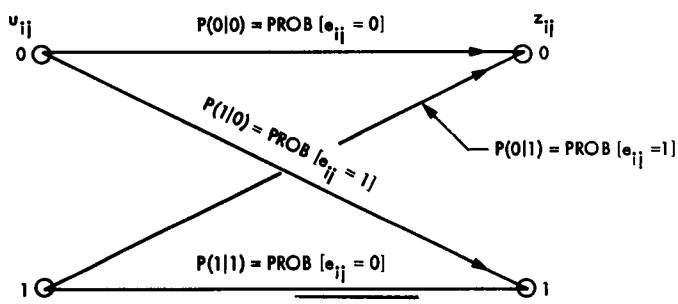


Fig. 5. Relationship of transition and error probabilities

Appendix A

The Modified Fano Metric for Sequential Syndrome Decoding of Binary Convolutional Codes

In Section II, the general technique is developed for finding all solutions of the syndrome equation of all noncatastrophic (n, k) CCs. In general, these solutions can be graphed on an error tree in precisely the same manner that these solutions were shown to graph on a trellis diagram. In fact, the trellis is just another way of representing the *graph* of a linear tree and vice versa.

The problem of a sequential syndrome decoding algorithm is to find, sequentially, at some desired tree depth L, the path or error sequence that has minimum Hamming weight. At any stage of such a sequential syndrome decoding algorithm, one must utilize a weight or metric to order the different partial paths the algorithm has already generated and considered in the error tree. In this appendix, the standard Fano metric is modified into a metric suitable for the syndrome decoding of convolutional codes.

To find the Fano metric, consider the code tree of a binary (n, k) CC of depth L. The transmitted code y_L is a code of L blocks of n binary bits of form

$$y_L = (y_{11}, y_{12}, \cdots, y_{1n}, y_{21}, y_{22}, \cdots, y_{2n}, \cdots, y_{L1}, y_{L2}, \cdots, y_{Ln})$$

where $x_{ij} = 0$ or 1 for $i = 1, 2, \cdots, L$ and $j = 1, 2, \cdots, n$. If such a code is sent through a binary symmetric channel, the received code z_L is of the same form as the transmitted code, namely,

$$z_L = (z_{11}, z_{12}, \cdots, z_{1n}, z_{21}, z_{22}, \cdots, z_{2n}, \cdots, z_{L1}, z_{L2}, \cdots, z_{Ln}) \tag{A-1}$$

An algorithm of a sequential decoder examines a set S_M of, say M, subsequences or subcodes of the code tree of depth L. This set of subcodes has the form

$$S_M = \left\{u_{k_1}, u_{k_2}, \cdots, u_{k_M}\right\}$$

where the subsequence u_m of the (n, k) CC code tree is of the form

$$u_m = \left(u_{11}, u_{12}, \cdots, u_{1n}, u_{21}, u_{22}, \cdots, u_{2n}, \cdots, u_{k_m 1}, u_{k_m 2}, \cdots, u_{k_m n}\right) \tag{A-2}$$

where $m = 1, 2, \cdots, M$.

The problem of the decoder is to find the sequence u_m in the set S_M that is most likely to compare with the received codeword z_L. To accomplish this comparison process with probability methods, one imagines, following Massey (Ref. 7), that each one of the partial codes u_m in S_M of depth k_m, for $1 \leqslant k_m \leqslant L$, is randomly extended to a depth L in the code tree. Such an extension is called a random tail. Explicitly, the extended codes for $u_{\varrho m}$ have the form

$$\hat{y}_L = (u_m, t_m) = (\hat{y}_{11}, \hat{y}_{12}, \cdots, \hat{y}_{1n}, \hat{y}_{21}, \hat{y}_{22}, \cdots, \hat{y}_{2n}, \cdots, \hat{y}_{L1}, \hat{y}_{L2}, \cdots, \hat{y}_{Ln})$$

where the "tail" is the binary vector,

$$t_m = \left(t_{11}, t_{12}, \cdots, t_{1n}, t_{21}, t_{22}, \cdots, t_{2n}, \cdots, t_{L-k_m,1}, t_{L-k_m,2}, \cdots, t_{L-k_m,n}\right) \tag{A-3}$$

chosen randomly from the code tree. Let t_m be chosen independently of u_m and y_L, in accordance with some arbitrary probability distribution $P(t_m)$.

The joint probability of transmitting u_m, augmenting u_m with the random tail t_m, and receiving y_L is

$$\begin{aligned} P(\hat{y}_L, z_L) &= P(u_m, t_m, z_L) \\ &= P(u_m) P(t_m | u_m) P(z_L | u_m, t_m) \\ &= P(u_m) P(t_m) P(z_L | \hat{y}_L) \end{aligned} \tag{A-4}$$

The last equality follows from the independence of t_m and u_m. Assume that the channel is a discrete memoryless channel (DMC) (Ref. 8). Then,

$$P(z_L | \hat{y}_L) = \prod_{j=1}^{n} \prod_{i=1}^{L} P(z_{ij} | \hat{y}_{ij})$$

$$= \prod_{j=1}^{n} \left[\prod_{i=1}^{k_m} P(z_{ij}|u_{ij}) \prod_{\ell=1}^{L-k_m} P\left(z_{\ell+k_m,j} \,|\, t_{\ell j}\right) \right] \tag{A-5}$$

But, the independence of y_L and t_m implies

$$P\left(z_{\ell+k_m,j} \,|\, t_{\ell j}\right) = P\left(z_{\ell+k_m,j}\right) \tag{A-6}$$

Hence, substituting Eqs. (B-5) and (B-7) into Eq. (4) yields

$$P(u_m, t_m, z_L) = P(u_m)P(t_m)\prod_{j=1}^{n} \times \left[\prod_{i=1}^{k_m} P(z_{ij} \,|\, u_{ij}) \prod_{\ell=1}^{L-k_m} P\left(z_{\ell+k_m,j}\right) \right] \tag{A-7}$$

The marginal joint distribution of u_m and y_L is obtained by summing $P(u_m, t_m, z_L)$ in Eq. (A-7) over all possible random tails t_m. Since

$$\sum_{t_m} P(t_m) = 1$$

this distribution is

$$P(u_m, z_L) = P(u_m)\prod_{j=1}^{n} \times \left[\prod_{i=1}^{k_m} P(z_{ij} \,|\, u_{ij}) \prod_{\ell=1}^{L-k_m} P\left(z_{\ell+k_m,j}\right) \right] \tag{A-8}$$

where $P(u_{ij})$ is the probability of bit y_{ij} at the output of the binary symmetric channel.

To compute the probability $P(u_m)$ that the partial code u_m in set S_M was transmitted, assume that the information symbols are independent and equally likely, i.e., the probability an information symbol is one equals $P = 1/2$. By Eq. (A-2), the number of information symbols used to generate the partial codeword u_m is

$$N_m = Rnk_m = \left(\frac{k}{n}\right) nk_m = k \cdot k_m \tag{A-9}$$

since the dimensionless code rate is $R = k/n$. Let i_m be the number of ones in the information sequence that generated the partial codeword u_m. Then,

$$P(u_m) = p^{i_m} q^{N_m - i_m} = p^{i_m} q^{Rnk_m - i_m} = 2^{-Rnk_m} \tag{A-10}$$

By Eqs. (A-8) and (A-10), and the rules for conditional probability,

$$P(u_m|z_L) = \frac{2^{-Rnk_m} \prod_{j=1}^{n} \left[\prod_{i=1}^{k_m} P(z_{ij}|u_{ij}) \prod_{\ell=1}^{L-k_m} P\left(z_{\ell+k_m,j}\right) \right]}{P(z_L)} \tag{A-11}$$

is the conditional probability that the partial codeword u_m was transmitted, given that y_L was received. The minimum error decoding (MED) principle (Ref. 8) says to choose that path u_m in set S_M that maximizes $P(u_m|z_L)$ in Eq. (A-11). Since $P(y_L)$ is constant over this maximization, this is equivalent to the choice m in set $\{1, 2, \cdots, M\}$, which maximizes

$$\psi(m) = 2^{-Rnk_m} \prod_{j=1}^{m} \left[\prod_{i=1}^{k_m} P(z_{ij}|u_{ij}) \prod_{\ell=1}^{L-k_m} P\left(z_{\ell+k_m,j}\right) \right] \tag{A-12}$$

Finally, since the term

$$\prod_{i=1}^{n} \prod_{j=1}^{L} P(z_{ij})$$

is a constant with respect to the maximization of $\psi(m)$, this maximization is equivalent to the maximization with respect to m of the ratio

$$\phi(m) = \psi(m) \Big/ \prod_{j=1}^{n} \prod_{i=1}^{L} P(z_{ij})$$

or its logarithm,

$$\mu_F(m) = \sum_{i=1}^{k_m} \sum_{j=1}^{n} \left[\log \frac{P(z_{ij} \,|\, u_{ij})}{P(z_{ij})} - R \right] \tag{A-13}$$

$\mu_F(m)$ in Eq. (A-13) is the standard Fano metric for the sequential decoding of an (n, k) CC.

The above derivation of the Fano metric is more directly applicable to (n, k) CC than the classical proof due to Massey (Ref. 7). It is also somewhat more precise than the recent derivation given by Clark and Cain (Ref. 9), upon which the present derivation is based.

As in Eq. (A-2), let u_{ij} denote a particular bit of some subsequence u_m in S_M of possible transmitted messages in the code tree of an (n, k) CC, and let z_{ij} denote the corresponding received bit. Then, z_{ij} is related to bit u_{ij} by

$$z_{ij} = u_{ij} + e_{ij}$$

where, solving for e_{ij},

$$e_{ij} = z_{ij} + u_{ij} \tag{A-14}$$

is the difference, or error, between the possible transmitted bit u_{ij} and the received bit z_{ij}, and addition is modulo 2 or addition of the field $F_2 = \{0, 1\}$ of two elements.

But, by the Shannon transition diagram in Fig. 5, the conditional probabilities $P(z_{ij}|u_{ij})$ are in all cases given by

$$P(z_{ij}|u_{ij}) = \text{Prob}\,[e_{ij} = u_{ij} + z_{ij}] \equiv P(e_{ij}) \tag{A-15}$$

A substitution of Eq. (A-15) into Eq. (A-13) yields

$$\mu_F(m) = \sum_{i=1}^{k_m} \sum_{j=1}^{n} \left[\log \frac{P(e_{ij})}{P(z_{ij})} - R \right] \tag{A-16}$$

as the desired modification of the Fano metric.

References

1. Reed, I. S., T. K. Truong, "New Syndrome Decoding Techniques for the (n, k) Convolutional Codes," *Proceedings IEEE*, Vol. 131, Pt. F, No. 4, July 1984.
2. Reed, I. S., T. K. Truong, "Error Trellis Syndrome Decoding Techniques for Convolutional Codes," *TDA Progress Report 42-78*, Jet Propulsion Laboratory, Pasadena, California, April-June 1984.
3. Vinck, A. J., A. J. P. de Paepe, and J. P. M. Schalkwijk, "A Class of Binary Rate One-Half Convolutional Codes that Allows an Improved Stack Decoder," *IEEE Transactions on Information Theory*, IT-26, No. 4, 1980, pp. 389-392.
4. Massey, J. L., and M. K. Sain, "Inverses of Linear Circuits," *IEEE Transactions on Comput.*, C-17, 1968, pp. 330-337.
5. Forney, G. D., "Convolutional Codes: Algebraic Structures," *IEEE Transactions on Information Theory*, IT-6, 1970, pp. 720-738.
6. Lin, S., and D. J. Costello, Jr., *Error Control Coding*, Prentice-Hall, New Jersey, 1983.
7. Massey, J. L., "Variable-Length Codes and the Fano Metric," *IEEE Transactions on Information Theory*, IT-18, January 1972, pp. 190-198.
8. Gallager, R. G., *Information Theory and Reliable Communications*, John Wiley and Sons, Inc., New York, 1968.
9. Clark, G. C., Jr., and J. B. Cain, *Error-Correcting Codes for Digital Communications*, Plenum Press, New York, 1981.

VLSI Architectures for the Multiplication of Integers Modulo a Fermat Number

J. J. Chang
Electronic Parts Reliability Section

T. K. Truong
Communications Systems Research Section

I. S. Reed and I. S. Hsu
University of Southern California

Multiplication is central in the implementation of Fermat number transforms and other residue number algorithms. There is need for a good multiplication algorithm that can be realized easily on a VLSI chip. In this article, the Leibowitz multiplier is modified to realize multiplication in the ring of integers modulo a Fermat number. This new algorithm requires only a sequence of cyclic shifts and additions. The designs developed for this new multiplier are regular, simple, expandable, and, therefore, suitable for VLSI implementation.

I. Introduction

Fermat number transforms (FNTs) were developed to compute cyclic convolutions (Refs. 1 and 2) and to encode and decode a (256, 224) Reed-Solomon (R-S) code in $GF(F_3)$, where $F_3 = 2^{2^3} + 1 = 257$ (Ref. 3). Recently Leibowitz (Ref. 4) proposed the diminished -1 representation for binary arithmetic operations modulo F_t. Based on the ideas in Ref. 4, a new algorithm is developed to realize a multiplier over the ring of integers modulo a Fermat number. This algorithm requires only cyclic shifts and additions. An example, illustrating both the pipeline and systolic array aspects of this structure, is given for a multiplier in the field of $GF(2^{2^2} + 1)$.

II. Multiplication Modulo F_t

In this section, a new algorithm is developed for multiplication in the ring of integers modulo $F_t = 2^{2^t} + 1$. This new algorithm is illustrated by the example for $t = 2$. The same structure clearly extends to more general multiplication algorithms over F_t.

To perform efficiently, the binary arithmetic operations modulo F_t, the diminished -1 representation proposed by Leibowitz (Ref. 4), are used. Table 1 shows the correspondence of elements in $GF(2^{2^2} + 1)$ with their decimal equivalents in a normal binary representation and with their values in the

diminished -1 representation. In the Leibowitz diminished -1 representation, the most significant bit (MSB) can be viewed as the zero-detection bit.

To realize this new algorithm for multiplication, let A, B be two binary integers and $A - 1, B - 1$ their diminished -1 representations. It is now desired to perform a multiplication of the two positive numbers $A - 1$ and $B - 1$ in their diminished -1 representation. Note that if the MSB of either $A - 1$ or $B - 1$ is one, then the multiplication is inhibited, and the product is zero.

To accomplish this new multiplication process, one first translates $B - 1$ from the diminished -1 representation to the normal binary representation, B. Then the multiplication of $A - 1$ and B is performed as follows:

$$(A-1)\cdot B = (A-1)\left(\sum_{k=0}^{4} b_k 2^k\right) = \sum_{k=0}^{4} b_k 2^k \cdot (A-1) \tag{1}$$

Multiplying the diminished -1 numbers by the different powers of two in Eq. (1) so that the result is a diminished -1 number. This is achieved by the identity $2^k (A - 1) + (2^k - 1) = 2^k A - 1$ for $0 \leqslant k \leqslant 4$. Thus when considered as a sum of diminished numbers, Eq. (1) becomes

$$(A-1)B = \sum_{k=0}^{4} b_k \cdot (2^k A - 1) = \sum_{k=0}^{4} (d_k - 1) \tag{2}$$

where $d_k - 1 = b_k \cdot (2^k A - 1)$ for $0 \leqslant k \leqslant 4$.

For $d_k - 1$ with $0 \leqslant k \leqslant 4$ to be diminished -1 numbers, they must add correctly pairwise in accordance with the formula,

$$(d_i - 1) + (d_j - 1) + 1 = (d_i + d_j - 1) \tag{3}$$

It is readily verified that this holds in all four cases of (b_i, b_j). Hence the diminished -1 addition in Eq. (2) can be performed recursively in the following manner:

$$(A-1)\cdot B = ((((d_0 + d_1 - 1) + d_2 - 1) + d_3 - 1) + d_4 - 1) \tag{4}$$

A substitution of Eq. (3) into Eq. (4) yields

$$\begin{aligned}(A-1)\cdot B = {} & (((((d_0 - 1) + (d_1 - 1) + 1) \\ & + (d_2 - 1) + 1) + (d_3 - 1) + 1) + (d_4 - 1) + 1)\end{aligned} \tag{5}$$

where $d_k - 1 = b_k(2^k A - 1)$.

From Eq. (5), the following identity is obtained for the product $A \cdot B$ diminished by 1:

$$\begin{aligned}(A-1)\cdot B &= b_0 \cdot (A-1) + b_1 \cdot (2A-1) \\ &\quad + b_2 \cdot (2^2 A - 1) + b_3 \cdot (2^3 A - 1) \\ &\quad + b_4 \cdot (2^4 A - 1) + 4 \\ &= A \cdot B - S + 4 = (A \cdot B - 1) - S + 5\end{aligned} \tag{6}$$

where $S = b_0 + b_1 + b_2 + b_3 + b_4$.

It follows from Eq. (6) that $A \cdot B$ in diminished -1 notation is

$$(A \cdot B - 1) = (A-1)\cdot B - (4 - S) - 2 + 1 \tag{7}$$

Now let

$$D = 4 - S \tag{8}$$

where $0 \leqslant D \leqslant 4$. Since $2^4 + 1 \equiv 0 \bmod F_2$, one has $-D - 2 = 2^4 + 1 - D - 2 = (2^4 - D - 1) = \bar{D}$, the binary one's complement of D. Hence, by Eq. (7), the diminished -1 representation of $A \cdot B$ becomes

$$C = (A \cdot B - 1) = (A-1)\cdot B + \bar{D} + 1 \tag{9}$$

where $\bar{D}$ is the one's complement of D in Eq. (8). A substitution of Eq. (4) into Eq. (9) yields

$$\begin{aligned}(A \cdot B - 1) = {} & (((((\bar{D} + b_0 \cdot (2^0 A - 1) + 1) + b_1 \cdot (2^1 A - 1) \\ & + 1) + b_2 \cdot (2^2 A - 1) + 1) + b_3 \cdot (2^3 A - 1) \\ & + 1) + b_4 \cdot (2^4 A - 1) + 1)\end{aligned} \tag{10}$$

as the new multiplication algorithm.

Let $Co = \bar{D}$. Then the multiplication algorithm in Eq. (10) can be put into the following recursive form:

$$C_{k+1} = C_k + b_k \cdot (2^{k+1} \cdot A - 1) + 1 \text{ for } 0 \leqslant k \leqslant 4 \tag{11a}$$

If one successively computes C_{k+1} in Eq. (11a) for $0 \leqslant k \leqslant 4$, then the required result is obtained as follows:

$$C_5 = C_4 + b_4(2^4A - 1) + 1 = (A \cdot B - 1) = C \tag{11b}$$

Example 1: A recursive diminished -1 multiplication algorithm: Let $A - 1 = 0\ 1\ 0\ 1\ 0$, $B - 1 = 0\ 0\ 1\ 0\ 1$; compute $C = (A \cdot B - 1) = 0\ 1\ 0\ 1\ 0 \times 0\ 0\ 1\ 0\ 1$ modulo $2^4 + 1$.

To compute C, one first translates $B - 1$ to B. That is, $B = B - 1 + 1 = 0\ 0\ 1\ 0\ 1 + 1 = 0\ 0\ 1\ 1\ 0 = b_4\ b_3\ b_2\ b_1\ b_0$. From Eq. (10), the sequence of computations for $0\ 1\ 0\ 1\ 0 \times 0\ 0\ 1\ 0\ 1$ is then as follows:

0 1 1 0 1	$C_0 = \overline{D} = 0\ 1\ 1\ 0\ 1$
+0 0 0 0 0	$b_0 \cdot (2^0A - 1) = 0(0\ 1\ 0\ 1\ 0) = 0\ 0\ 0\ 0\ 0$
0 1 1 0 1	
+ ↘1	
0 1 1 1 0	$C_1 = C_0 + b_0 \cdot (2^0A - 1) + 1$
+0 0 1 0 0	$b_1 \cdot (2^1A - 1) = 1 \cdot (0\ 0\ 1\ 0\ 0) = 0\ 0\ 1\ 0\ 0$
1 0 0 1 0	
+ ↘0	
0 0 0 1 0	$C_2 = C_1 + b_1 \cdot (2^1A - 1) + 1$
+0 1 0 0 1	$b_2 \cdot (2^2A - 1) = 1 \cdot (0\ 1\ 0\ 0\ 1) = 0\ 1\ 0\ 0\ 1$
0 1 0 1 1	
+ ↘1	
0 1 1 0 0	$C_3 = C_2 + b_2 \cdot (2^2A - 1) + 1$
+0 0 0 0 0	$b_3 \cdot (2^3A - 1) = 0 \cdot (0\ 0\ 0\ 1\ 0) = 0\ 0\ 0\ 0\ 0$
0 1 1 0 0	
+ ↘1	
0 1 1 0 1	$C_4 = C_3 + b_3 \cdot (2^3A - 1) + 1$
+0 0 0 0 0	$b_4 \cdot (2^4A - 1) = 0 \cdot (0\ 0\ 1\ 0\ 1) = 0\ 0\ 0\ 0\ 0$
0 1 1 0 1	
↘1	
0 1 1 1 0	$C_5 = C_4 + b_4(2^4A - 1) + 1 = C$

Thus $C = 0\ 1\ 1\ 1\ 0$ is the desired result of $0\ 1\ 0\ 1\ 0$ times $0\ 0\ 1\ 0\ 1$, modulo $2^4 + 1$ in diminished -1 notation.

III. A VLSI Structure for Implementing Multiplication Modulo F_t

Example 1 of the new diminished -1 multiplication algorithm in the previous section shows that diminished -1 additions require the addition of the complement of an end around carry to its sum. A considerable speed improvement can be obtained by performing this operation simultaneously with the summation. A modified algorithm with this simultaneous addition is given for the previous example as follows:

Example 2: Modified recursive diminished -1 multiplication:

1 1 1 0 1	$C_0 = 1\ 0\ 0\ 0\ 0 + \overline{D} = 1\ 1\ 1\ 0\ 1$
0 0 0 0 0	$b_0 \cdot (2^0A - 1) = 0(0\ 1\ 0\ 1\ 0) = 0\ 0\ 0\ 0\ 0$
+ ↘0	
0 1 1 0 1	$C_1 = C_0 + b_0 \cdot (2^0A - 1) + 1$
0 0 1 0 0	$b_1(2^1A - 1) = 1 \cdot (0\ 0\ 1\ 0\ 0) = 0\ 0\ 1\ 0\ 0$
+ ↘1	
1 0 0 1 0	$C_2 = C_1 + b_1(2^1A - 1) + 1$
0 1 0 0 1	$b_2(2^2A - 1) = 1 \cdot (0\ 1\ 0\ 0\ 1) = 0\ 1\ 0\ 0\ 1$
+ ↘0	
0 1 0 1 1	$C_3 = C_2 + b_2(2^2A - 1) + 1$
0 0 0 0 0	$b_3(2^3A - 1) = 0 \cdot (0\ 0\ 0\ 1\ 0) = 0\ 0\ 0\ 0\ 0$
+ ↘1	
0 1 1 0 0	$C_4 = C_3 + b_3(2^3A - 1) + 1$
0 0 0 0 0	$b_4(2^4A - 1) = 0 \cdot (0\ 0\ 1\ 0\ 1) = 0\ 0\ 0\ 0\ 0$
+ ↘1	
0 1 1 0 1	$C_5 = C_4 + b_4(2^4A - 1) + 1$
0 0 0 0 0	$0(2^5A - 1) = 0 \cdot (0\ 1\ 0\ 1\ 1) = 0\ 0\ 0\ 0\ 0$
+ ↘1	
0 1 1 1 0	$C = C_5 + 1$

A possible VLSI structure for Example 2 is presented in Fig. 1. In Fig. 1, A, B, and C are 4-bit, 6-bit, and 5-bit registers, respectively. Initially, registers A, B, and C contain the multiplicand in the diminished -1 representation, the multiplier in normal representation, and $2^4 + D$, respectively. At the very same moment, $C_{k+1} = C_k + b_k(2^kA - 1)$ is computed and

loaded into the C register. Simultaneously, the diminished -1 multiplication of $(A - 1)$ by 2 is performed first by a left cyclic shift of the four least-significant bits of the register A with the A_3-bit circulated into the first significant bit complemented. Also, at the same time, register B is shifted right by one bit. These operations are continued repetitively until the MSB of the register B is shifted out. The desired final result of 0 1 1 1 0 after 5 iterations is obtained in register C.

The layout of the structure in Fig. 1 has been completed with the use of the CAESAR design tool (Ref. 5). The final layout of the multiplication chip is shown in Fig. 2.

In the future, logic and circuit simulations will be performed. These will be followed by chip fabrication and final testing. The total number of transistors in this chip is about 300. The area of the chip is estimated to be about 12,000 λ^2.

References

1. Rader, C. M., "Discrete Convolution via Mersenne Transforms," *IEEE Trans. Computers*, Vol. C-21, No. 12, pp. 1269-1273, Dec. 1972.
2. Agarwal, R. C., and C. S. Burns, "Fast Convolution Using Fermat Number Transforms with Applications to Digital Filtering," *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol. ASSP-22, No. 2, pp. 87-97, April 1974.
3. Reed I. S., T. K. Truong, and L. R. Welch, "The Fast Decoding of Reed-Solomon Code Using Fermat Number Transforms," *IEEE Trans. Information Theory*, Vol. IT-24, No. 4, pp. 497-499, July 1978.
4. Leibowitz, L. M., "A Simplified Binary Arithmetic for the Fermat Number Transform," *IEEE Trans., Acoustic, Speech and Signal Processing*, Vol. ASSP-24, No. 5, pp. 356-359, Oct. 1976.
5. Ousterhout, J., *Editing VLSI Circuits with Caesar*, Computer Science Division, Electrical Engineering and Computer Sciences, University of California, Berkeley, April 21, 1982.

Table 1. The correspondence among decimal numbers, their values in the normal binary representation, and in the diminished −1 representation

Decimal number	Normal binary representation	Diminished −1 representation
0	0 0 0 0 0	1
1	0 0 0 0 1	2
2	0 0 0 1 0	3
3	0 0 0 1 1	4
4	0 0 1 0 0	5
5	0 0 1 0 1	6
6	0 0 1 1 0	7
7	0 0 1 1 1	8
8	0 1 0 0 0	9 (-8)
9 (-8)	0 1 0 0 1	10 (-7)
10 (-7)	0 1 0 1 0	11 (-6)
11 (-6)	0 1 0 1 1	12 (-5)
12 (-5)	0 1 1 0 0	13 (-4)
13 (-4)	0 1 1 0 1	14 (-3)
14 (-3)	0 1 1 1 0	15 (-2)
15 (-2)	0 1 1 1 1	16 (-1)
16 (-1)	1 0 0 0 0	0

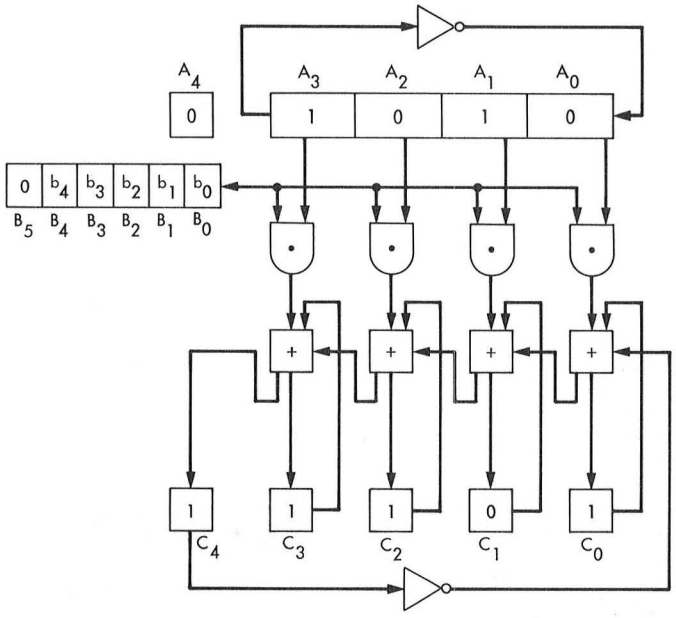


Fig. 1. The pipeline architecture for the implementation of multiplication modulo the Fermat number $2^4 + 1$ using diminished -1 number representations

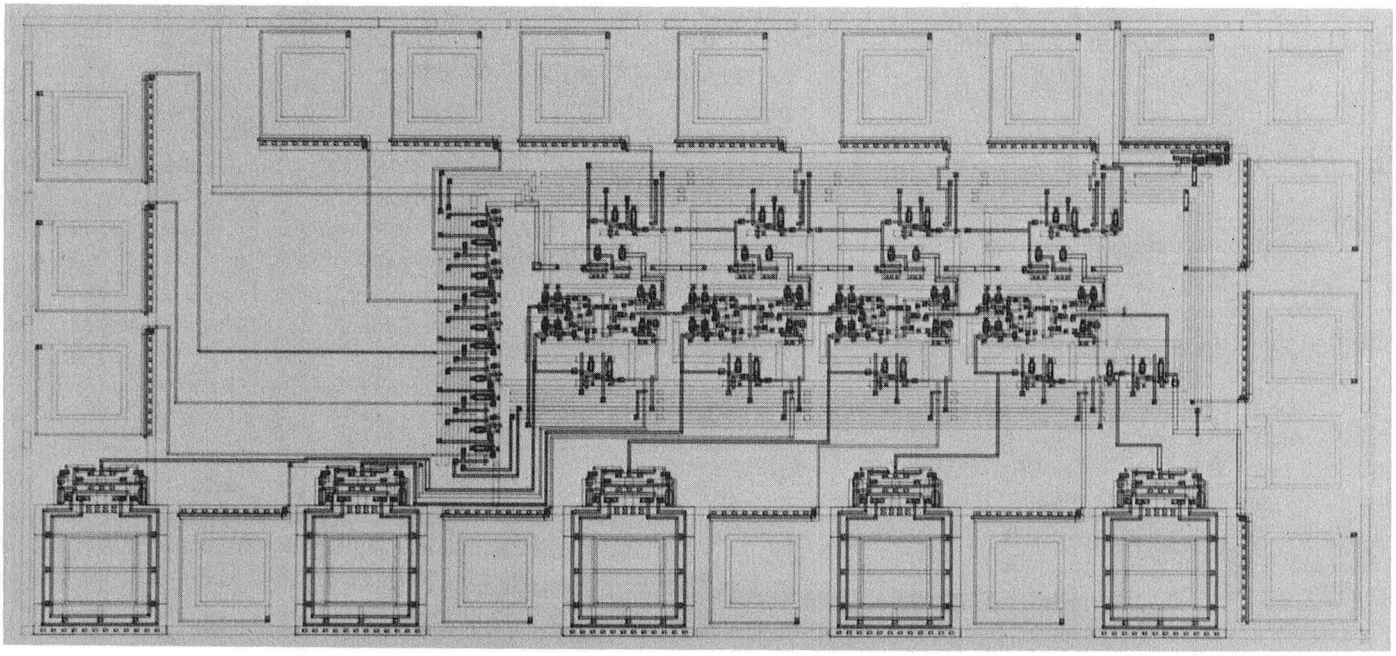


Fig. 2. The VLSI nmos technology layout for multiplication modulo the Fermat number $2^4 + 1$, using diminished -1 number representation

Programmable Digital Baud Integrators for the Radar High-Speed Data Acquisition System

K. H. Farazian and R. F. Jurgens
Communications Systems Research Section

As part of the planetary radar upgrade program to increase the resolution and speed of the radar system and in support of radio and radar astronomy, a new high-speed data acquisition system has been designed. Included in this article is a description of the programmable digital Baud Integrators used at the early stages of this system.

I. Introduction

With the advance of technology in LSI and VLSI, there exists a potential for the use of digital integrated systems in real-time processing areas that were previously implemented with analog circuitry. Digital circuits have become more attractive since they are less subject to distortion and interference than analog circuits, and can be easily regenerated so that noise does not accumulate in transmission. In addition, they have potential for extremely low error rates and high fidelity through error detection and correction, and they have greater reliability and can be mass produced. This article describes an all-digital technique for the Baud Integrators of the Radar High-Speed Data Acquisition System, a technique that avoids the inherent problems associated with analog systems such as the need for calibration and adjustment. The integration period of this system is selectable in 100-ns steps from 100 ns to 3276700 ns.

In the Radar High-Speed Data Acquisition System implemented to support the Deep Space Network program, the incoming IF signal is digitized immediately and all further processing is done digitally. Figure 1a shows the functional block diagram of the proposed radar system and Fig. 1b shows the front end of this system where the Baud Integrators are employed. The TRW TDC1025E1C analog-to-digital (A-D) converter boards produce 8-bit samples that are passed to the digital Baud Integrators, which are used as matched detectors for the range code elements. The Baud Integrators sum the samples for a preprogrammed number of samples without producing any overflow. The result is rounded to the most significant 8 bits. Of these 8 bits, four are used as extended bits for future use and four are sent out for further processing. The number of samples, N, and the scaling factor are the two inputs that must be supplied to the Baud Integrators. These are controlled by the VAX 11/780 host computer via a Unibus DR11C interface.

Figure 2 shows that the Baud Integrator can be broken into two parts. The first part is the Prebaud Integrator System, which represents a finite impulse response (FIR) filter operating at a fixed 50-MHz clock rate. The second part is the Postbaud Integrator System that is clocked with the 10-MHz

drifted clock as are the following digital processors. Since the drifted clock is computed from the 50-MHz clock, its edges are stable at the 50-MHz clock rate.

II. Prebaud Integrator

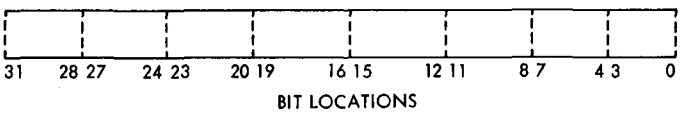
The function of the Prebaud Integrator is to add five successive samples and produce the correct result with no overflow every 20 ns to the Postbaud Integrator. The input to the Prebaud Integrator is the 8 bits of the TRW A-D converter clocked at 50 MHz. Refer to Fig. 3. The Prebaud Integrator, which has a FIR architecture associated with it, employs buffered adders that can operate up to a maximum of 50 MHz. The maximum clock frequency of this system is restricted only by the speed of the adders. In this system, the Fairchild Fast™ series 12-bit full adder with fast carry has been used. This procedure is faster than the 12-bit lookahead carry ALU, which is commonly used.

The 50-MHz input/output rate of the Prebaud Integrator not only significantly reduces the error introduced into the calculations by the shifting of the 10-MHz drifted clock of the Postbaud Integrator, but also eliminates the need for any "clear" or "reset" signal. That is, at most, only one sample will be deleted or used twice in the integration when the 20-ns corrections are made in the drifted clock. Also, being an iterative network, this configuration makes it easy to expand the number of bits or replace the adder modules with faster chips in the future.

The Prebaud Integrator can also be implemented using an Infinite Impulse Response (IIR) System as shown in Fig. 4. Here double-buffered accumulators are used to add five successive samples while the other branch is stable for unloading. Because of the high 50-MHz clock rate, the (A-D) converter output is split into two branches, each operating at a reduced rate of 25 MHz. Finally, the two branches are brought back together and added. Note that when one branch is adding three samples the other is adding two and vice versa. The output of this system will be at 10 MHz and must be resampled by the 10-MHz drifted clock of the Postbaud Integrator in a resampling circuit. The 10-MHz output and the resampling circuit of this IIR system increase the error introduced in the system significantly compared to the FIR system, making the IIR system less desirable. For example, the resampling involves deleting or using a 100-ns sum twice whenever the sample clock is corrected by 20 ns unless some more complicated scheme is employed. Refer to Fig. 5 for a graphic representation of this comparison. On the other hand, because each branch of the IIR system operates at the reduced rate of 25 MHz, this system is attractive for some applications where the chip count is important.

III. Postbaud Integrator

The Postbaud Integrator uses the double-buffered TRW TDC1043J3C[1] multiplier-accumulator chip, which adds N successive samples of the Prebaud Integrator in offset binary. Because of the reduced rate of the 10-MHz drifted clock, no branching is necessary. However, because of the large number of accumulations, the sum must be scaled so that the result is placed in the correct bit position for further processing. This is done by the multiplier portion of the TRW chip. Refer to Figs. 6 and 7 for block diagrams. The number of samples accumulated, N, is programmable along with the scaling factor associated with it by the VAX 11/780 host computer via a Unibus DR11C interface. N ranges from 2 to 2^{15} and is supplied in a two's-complement format via four binary counters. This reduces the number of chips required to input it in binary. The scale factor is written in binary and uses the equation, scale = $2^M/N$, where $M < 17$ depends on the number of inputs to the postbaud integrator. The following calculation assumes that the A-D converter is adjusted so that zero volt corresponds to an average value of 127.5 and that there are 11 output lines from the Pre- to the Postbaud Integrator. These 11 output lines will be the 11 most significant bits (MSBs) (out of 16 bits) X-input of the TRW chip, namely X5 to X15 where X0 to X7 are not used and therefore are pulled down. The scaling factor will be the 16-bit Y-input lines of the multiplier-accumulator. Depending on how many samples are added, the scaling factor is so chosen as to place the result in the MSB of the output of the TRW chip. These bits are selected to be P23 to P30 with P23 to P26 as the optional extended bits. Due to the accumulation in the Prebaud Integrator, samples of the (A-D) converter with the average of 127.5 now have the new average value of 5 × 127.5. With the correct scaling factor, N samples are accumulated and placed at the 8 MSBs of the 32 output bits of the multiplier-accumulator chip. This means that the fractional part starts at bit number 23.



Expressing 127.5 in the hexadecimal format and having 32 bits where the decimal point is just before bit 23, one can write

$$7F.7FFFFF < 127.5 = 7F.800000$$
$$= 7 \times 2^{28} + 15 \times 2^{24} + 8 \times 2^{20}$$
$$= 2.139 \times 10^9$$

[1] *TRW TDC1043 VLSI Multiplier-Accumulator Preliminary Information.* TRW Inc., La Jolla, Calif., 1983.

Then, to calculate the scaling factor, use the following equation:

$$N \times 2^5 \times 5 \times 127.5 \times \text{scale} = 2.139 \times 10^9$$

or

$$\text{scale} = 104852.9/N < 2^{17}/N = 131072/N$$

There is a factor of 2^5 because X5 to X15 are the only actual inputs from the Prebaud Integrator.

Therefore select

$$\text{scale} = 2^{15}/N$$

extracting the output at bit locations P23 to P30. For example, if $N = 2$ (minimum value of N, adding only two samples) then scale = 2^{14}, which implies that the input is shifted by 14 bit locations. With the input being X5 to X15, the result will appear at output bit locations P19 to P29. When two of these are added, the 8 MSBs of the 9-bit addition result are placed at output bits P23 to P30. Note that M could have been chosen to be 16. In this case, the 8 Most Significant Products (MSPs) of the TRW chip, namely P24 to P31, would have been the correct output to use for further processing of the result.

The maximum clock frequency of Postbaud Integrator is limited by the TRW chip, which can run up to a maximum of 11 MHz, as was verified in the laboratory.

The Baud Integrator can also be designed to operate in two's complement, which has some advantages and disadvantages relative to offset-binary. These are fully discussed in the next section. In the remainder of this section, however, two's-complement hardware implementation is compared to that of the offset-binary already discussed. The adders in the Prebaud Integrator stage and the TRW chip in the Postbaud Integrator stage both have the capability of operating in two's complement. However, to avoid overflow, the resolution of the system decreases. To increase the resolution to the same 8 bits as in the offset binary, more chips are required. The two's-complement operation of the system is achieved by extending the 8 output bits (0 to 7) of the Prebaud Integrator, with bit 7 being the sign bit, to 16 bits at the Postbaud Integrator input. This 16-bit input (0 to 15), with its bit 15 carrying the sign bit, is in turn extended to 35 bits internally by the TRW chip where the sign bit now is bit 34. Notice that with the TRW chip, three-state output drivers are provided for one 16-bit word, the MSP, and one 3-bit word, the extended product (XTP). The least-significant product is not available with the TDC1043J3C, but is held internally for use in accumulation. This would mean that in the addition of small numbers, barely any information, if any, regarding the result of the addition will show up at the external output. To overcome this problem, one can implement the process just described by the IIR system discussed in the previous section where the scaling of the samples and number of samples added must be programmable. This would require a more complex system with an increased chip count.

IV. Two's-Complement Operation Analysis

This section deals with the analysis of the Baud Integrator operation in two's complement. As is shown in the following, this scheme requires less memory space, which reduces the number of registers used in the system. On the other hand, it introduces some problems that make this system less desirable.

Assume that the incoming signals, n_i, to the Baud Integrator are independant random variables with the following characteristics:

$$E\{n_i\} = 0 \text{ (zero-volt average)}$$

and

$$E\{n_i^2\} = \sigma_n^2$$

where E is the expected value operator. Then the output of the Baud Integrator, which is the sum of these samples, will have the following characteristics:

with

$$S = \sum_{i=1}^{N} n_i$$

$$E\{S\} = 0$$

and

$$E\{S^2\} = N\sigma_n^2 = \sigma_S^2$$

or

$$\sigma_S = \sqrt{N}\sigma_n$$

This implies that the output register grows as $\sqrt{N}$ in the two's complement compared to N in the offset-binary. So a noticeable savings in memory space can be achieved for a large number of additions.

In spite of this fact, there are other factors that make this system less desirable. For one, the sign bit extension must be implemented; and for another, the final result must be scaled properly for further processing of the signal. These were fully discussed at the end of Section III. In addition, either system must be able to operate with test signals that do not have the statistical properties alluded to in the above discussion. This makes the two's-complement operation equally or more complex than the offset binary.

V. Computer Interface

The main interface to the VAX 11/780 computer for the Radar Baud Integrators is a Unibus DR11C (Ref. 1). The DR11C requires three address locations for operation on the Unibus. Address XXXO is DRCSR (DR11C Control and Status Register). The second address location can be used as a read/write port and is being used for write only. The third address location is for read only and is being used for the purpose of reading back the Baud-Integrator internal registers.

Figures 8 and 9 show the formats and functions of the baud internal control and status register (ICSR) and functional hold registers. The DR11C CSR bits CSR1 and CSR0 are used to control the mode of operation in read/write to DR11C addresses. With both CSR1 and CSR0 equal to zero, addresses 2 and 3 write or read from the ICSR in the Baud Integrator. With CSR1 = 0 and CSR0 = 1, addresses 2 and 3 write and read from the registers pointed to by the ICSR. The register pointer within the ICSR can be set to auto increment on read or write.

Figure 9 shows the two internal registers for controlling the Postbaud Integrator, setting N and the scaling factor.

VI. Summary and Conclusion

The Baud Integrator employs an all-digital system technology instead of the customary analog technology. It operates with two coherent input clocks running at 10 MHz and 50 MHz. The Prebaud Integrator sums five successive samples and the Postbaud Integrator accumulates up to 2^{15} samples of the Prebaud Integrator. Both N and the scaling factor are programmable via a Unibus DR11C interface from the VAX 11/780 host computer.

Acknowledgments

The authors wish to thank Stan Brokl for original interface design.

References

1. S. S. Brokl, "Polynomial Driven Time Base and PN Generator, *The Telecommunications and Data Acquisition Progress Report 42-75*, pp. 84-90, July-September 1983. Jet Propulsion Laboratory, Pasadena, Calif., November 15, 1983.

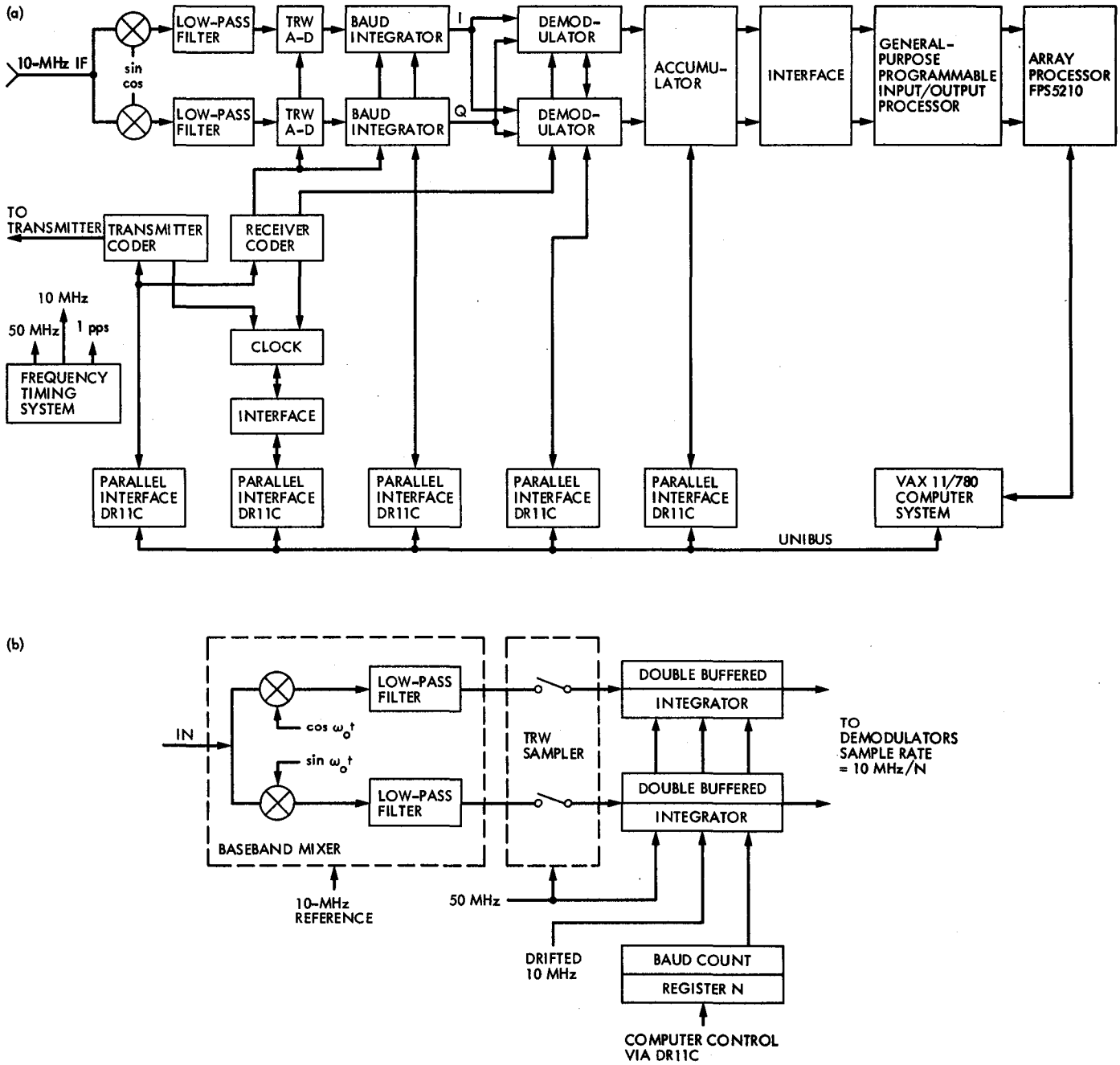


Fig. 1. High-resolution radar system: (a) functional block diagram; (b) front-end digital processing system

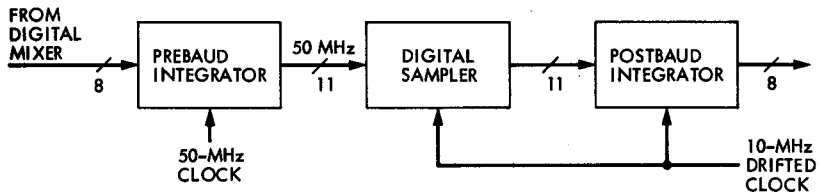


Fig. 2. Separation of the baud integrator into pre- and postbaud integrator with resampling in the center

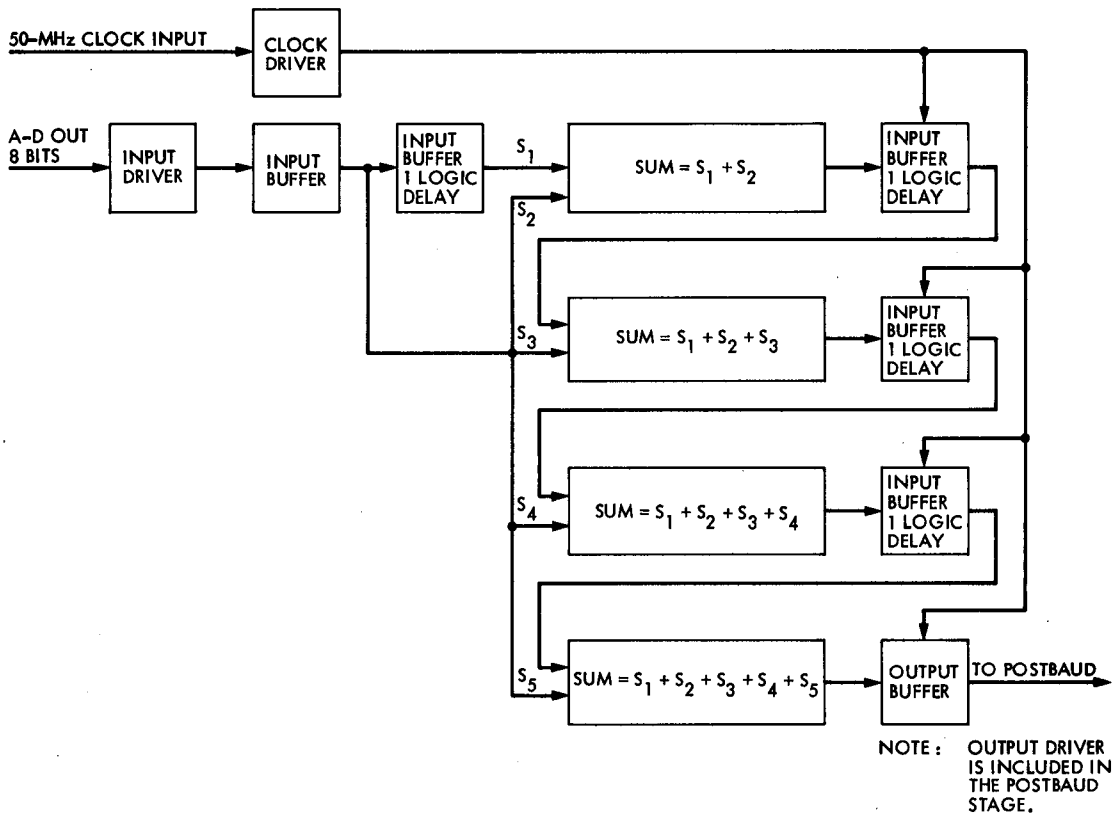


Fig. 3. Prebaud integrator design block diagram

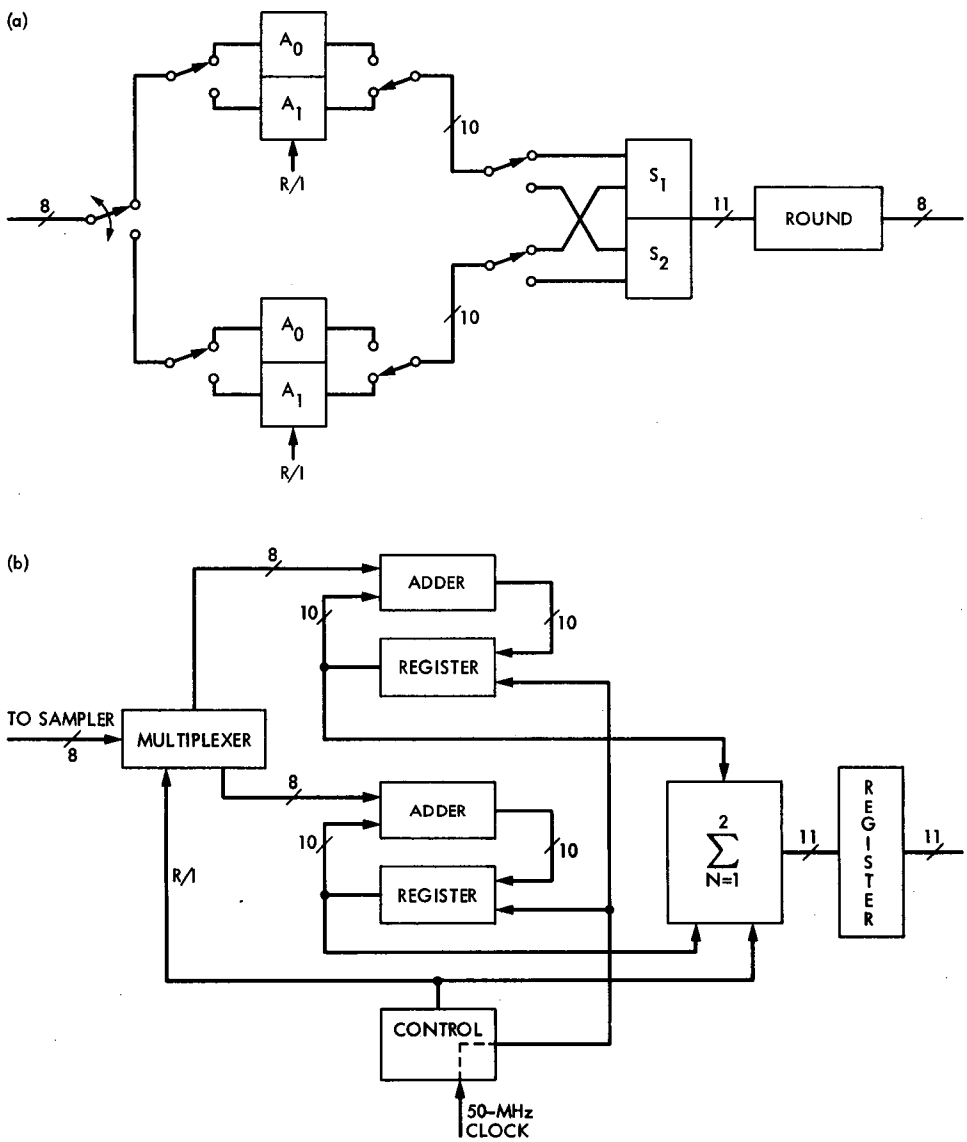


Fig. 4. Prebaud integrator with all clocks derived from the stationary 50 MHz: (a) using IIR system (integrate and dump); (b) double-buffered block diagram of (a)

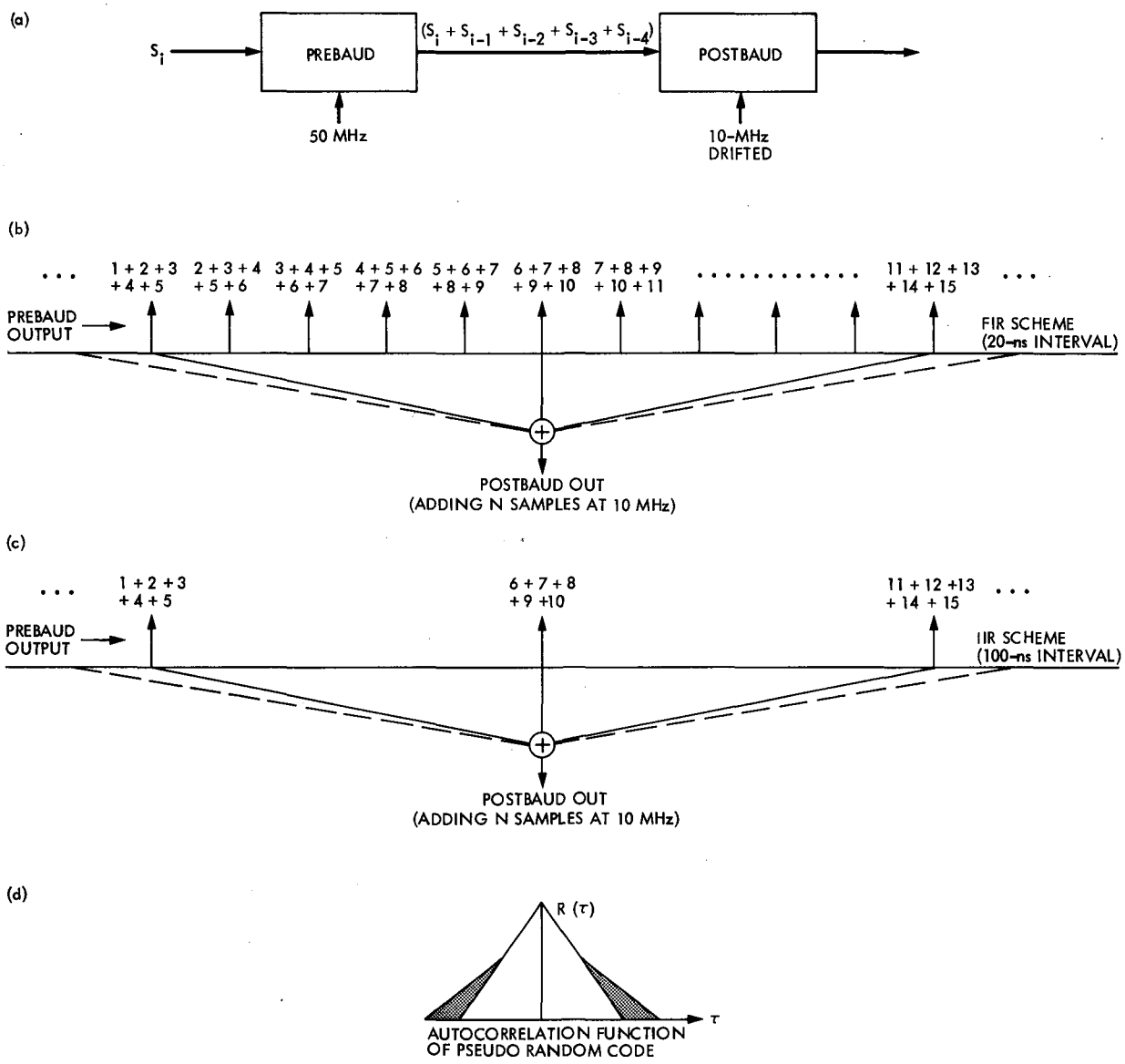


Fig. 5. Comparison of the FIR and IIR structures for the prebaud integrator design: (a) system block diagram; (b) FIR structure with an output rate of 50 MHz; (c) IIR structure with an output rate of 10 MHz; (d) autocorrelation function of pseudorandom code where the shaded area represents the error introduced with the shifting of the 10-MHz drifted clock. This error is reduced significantly in the FIR scheme.

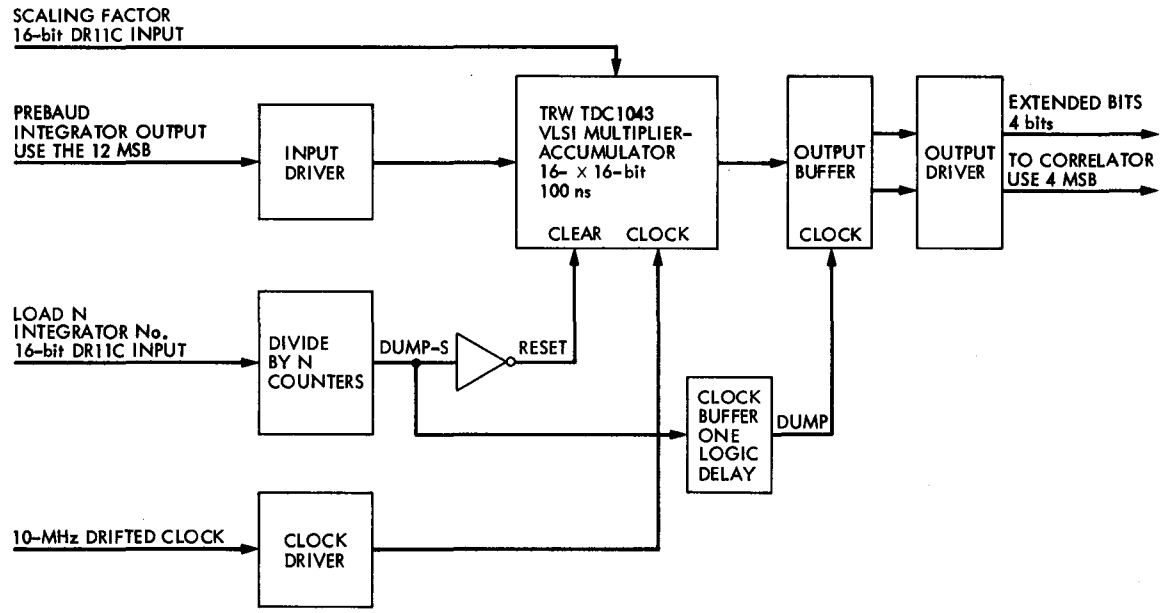


Fig. 6. Postbaud integrator design

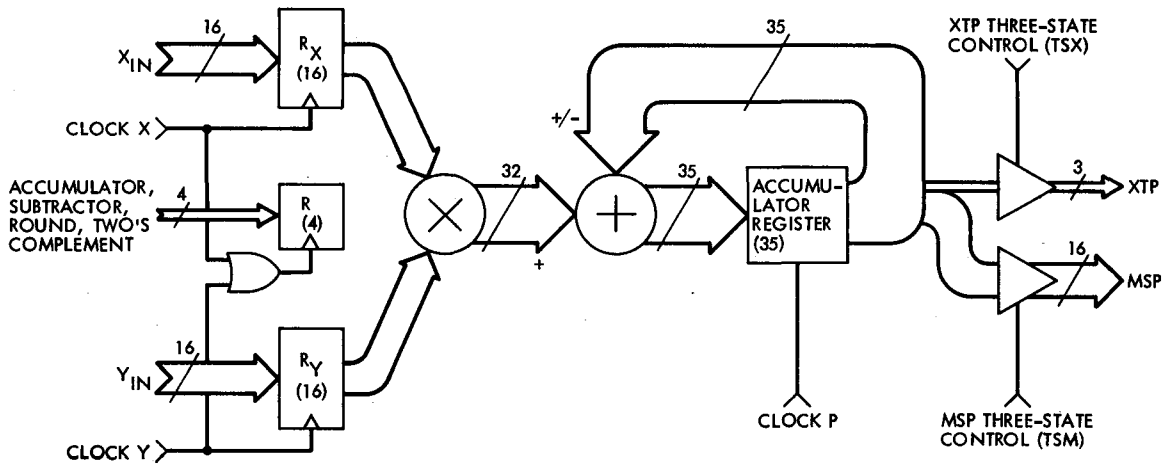
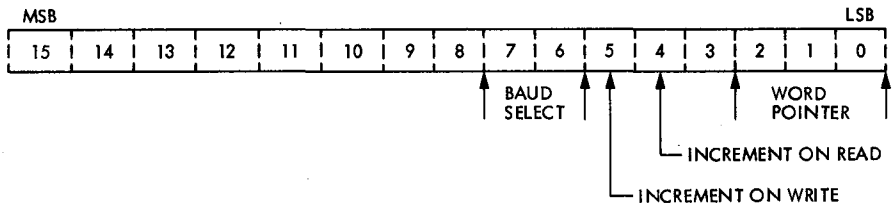


Fig. 7. Functional block diagram of the 10-MHz TRW TDC 1043J3C, 16- × 16-bit, VLSI multiplier-accumulator



WORD POINTER

000 = SCALE
001 = LOADN
010 = F
011 = F U U
100 = O T S
101 = R U E
110 = R
111 = E

BAUD INTEGRATOR SELECT

00 = B.I. 1
01 = B.I. 2
10 = B.I. 3
11 = B.I. 4

DR11C MODE CONTROL BITS

CSR1	CSR0	
0	0	= R/W ICSR
0	0	= R/W FUNCTION REGISTERS POINTED TO IN THE SELECT AND WORD POINTER
1	0	= FOR FUTURE USE
1	1	= FOR FUTURE USE

Fig. 8. Baud ICSR

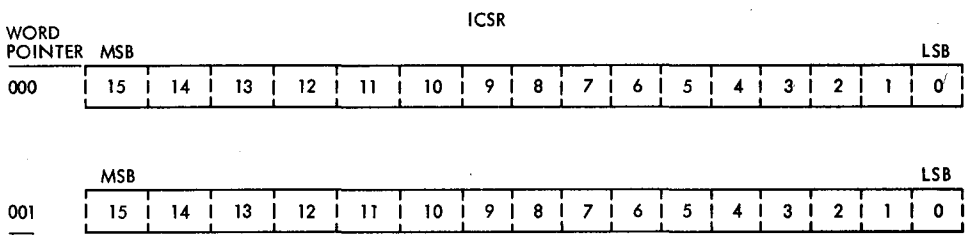


Fig. 9. Function register formats

Mark IVA DSN 26-Meter Subnet

D. D. Gordon
TDA Engineering Office

The Office of Space Tracking and Data Systems' Networks Consolidation Program (NCP), managed by the Jet Propulsion Laboratory (JPL), includes the implementation of a 26-meter Tracking and Communications Subnet as a part of the Mark IVA Deep Space Network (DSN). The incorporation of this subnet into the DSN will contribute to the NCP goal of consolidating the two NASA ground tracking networks into one tracking network to be operated for NASA by JPL. The ground networks to be consolidated are the Ground Spaceflight Tracking and Data Network operated by the Goddard Space Flight Center, and the DSN operated by JPL.

The 26-meter Tracking and Communication Subnet has been designed to provide the capability to support, at each Deep Space Communication Complex, the tracking and data communication requirements of the earth-orbital missions that cannot be supported by the Tracking and Data Relay Satellite System when it becomes operational.

Previous articles in the TDA Progress Report have traced the history and progress of the NCP's planning and budgeting activities. This article deals with the implementation activities associated with the 26-meter Tracking and Communications Subnet and its planned capabilities as an element of the consolidated DSN.

I. Introduction

Transfer of an operational 26-meter subnet from the Goddard Space Flight Center (GSFC) to JPL as an element of the Mark IVA DSN is scheduled to take place on February 1, 1985. This transfer contributes to the accomplishment of the NCP's objective to consolidate the capabilities of the two NASA ground tracking networks. Previous articles in the *TDA Progress Report* (Refs. 1, 2, and 3) have traced the history of this consolidation effort. The Mark IVA DSN 26-meter Subnet will be composed of a 26-meter, S-band, transmit and receive station located at each of the Deep Space Communication Complexes (DSCCs). The three stations in the Subnet are designed to provide spacecraft support consistent with the 26-meter link capabilities at a typical Ground Spaceflight Tracking and Data Network (GSTDN) station. Implementation of these stations will maximize the use of existing GSTDN hardware and software and will be configured to provide the minimum capabilities to support the needs of the 26-meter subnet mission set.

The NCP schedule requiring an operational 26-meter subnet was based on the planned operational date of the Tracking and Data Relay Satellite System (TDRSS) and the related phaseout of most of the GSTDN stations. However, the closure of GSTDN stations has been deferred, consistent with the recent

postponement of the TDRSS operational date. Therefore, the 26-meter subnet will serve in conjunction with the extended GSTDN to provide support to the TDRSS compatible missions, as well as the initially planned 26-meter Subnet mission set.

II. Mission Set

Definition of the missions requiring committed support by the 26-meter Subnet of the DSN has evolved to include a set of missions that can be divided into four classifications.

A. Extended Missions

Existing missions currently supported by the GSTDN that are not TDRSS compatible, but require support beyond the planned phasedown of the GSTDN.

(1) Nimbus 7.

(2) Dynamics Explorer (DE) 1.

(3) International Sun-Earth Explorer (ISEE) 1 and 2.

B. TDRSS Compatible Missions

Missions designed for support by the TDRSS that may require support prior to the TDRSS operational date.

(1) Landsat 4 and 5.

(2) Earth Radiation Budget Satellite (ERBS).

(3) Solar Maximum Mission (SMM).

(4) Solar Mesosphere Explorer (SME).

(5) Space Telescope (ST).

C. Special Support Missions

Missions that require a capability for emergency support or support of special spacecraft events.

(1) Space Transportation System (STS).

(2) Tracking and Data Relay Satellite System (TDRSS).

(3) International Cometary Explorer (ICE) (Encounter).

(4) Active Magnetospheric Particle Tracer Explorers (AMPTE).

(5) Geostationary Operational Environmental Satellite (GOES) G and H.

D. Reimbursable Missions

Missions that have contracted with NASA for support by the DSN on the basis of cost reimbursement.

(1) Broadcasting Satellite (BS) 2B.

(2) German Direct TV Broadcast Satellite (TV-SAT).

(3) French Direct TV Broadcast Satellite (TDF) 1.

III. Configuration

A representation of the 26-meter Subnet is included in Fig. 1, Mark IVA DSN 1985 configuration, where the 26-meter stations are identified as DSS 16, 46, and 66 at the Goldstone, Canberra, and Madrid complexes, respectively. As illustrated in Fig. 1, each station is essentially a stand alone element within a complex. Initially the 26-meter control room functions will not be integrated into the Mark IVA DSN Signal Processing Centers (SPCs). However, operational control will be exercised from the JPL Network Operations Control Center. The control rooms will have direct NASCOM communications with GSFC. Additionally, Fig. 1 indicates a cross-support capability within each complex that will allow command and telemetry signals at baseband to be routed between the 26-meter control room and the SPC.

The physical location and configuration of the 26-meter control room will be different at each complex, although the equipment will be essentially identical. At the Canberra Deep Space Communications Complex, the control room will be located in the pedestal building of the 64-meter antenna. At the Madrid Deep Space Communications Complex, the control room will be located in Room 900, adjacent to the Mark IVA SPC. The existing GSTDN 26-meter link at Goldstone, California will undergo minor modifications, but will not be physically relocated and will become the 26-meter Subnet station at the Goldstone Deep Space Communications Complex (GDSCC).

IV. System Characteristics

The basic equipment complement and interface design are illustrated in Fig. 2, 26-meter Subnet station design. The equipment consists of existing GSTDN hardware with the exception of two new devices, the Programmable Data Formatter (PDF) and the Shuttle Forward Link.

A. Telemetry System

Each 26-meter station will have the capability for processing up to three spacecraft telemetry channels simultaneously depending upon communication circuit limitations. The telemetry system includes a new throughput processor, the PDF. The PDF is a Digital Equipment Corporation LSI-11 based system that will accept the input data streams and block them asynchronously for transmission off-site.

The received spacecraft signal is routed to the various telemetry subsystem assemblies consisting of tunable filters,

discriminators, bit synchronizers, frame synchronizers, PSK demodulators, and project unique assemblies, as appropriate. The data is then routed to the PDF where it is block formatted for transmission. In addition to the telemetry blocking function the PDF also performs multiplexing of the blocked serial streams from the Spacecraft Command Encoder and Tracking Data Processor System. This data is then transferred to wideband NASCOM circuits. The PDF is capable of transferring data to a maximum of three 56-kb/s circuits or one 224-kb/s circuit.

A second PDF will be used by the station for monitoring data quality. This unit will sample an output channel of the first PDF, deblock the data stream, and route it to a frame synchronizer. The frame synchronizer will be used to verify the ability to lock-on the data stream.

(1) Key characteristics:

(a) Process data rates up to 800 kb/s.

(b) Process and/or record three data streams simultaneously (e.g. NIMBUS: 4-kb PCM and two 800-kb real time recordings; or, ISEE 1 and 2: two real-time PCMs and one real-time recording).

(c) Demodulation of NRZ-L or Bi-phase-L data on a subcarrier or carrier.

(d) Demodulate subcarriers up to 1.024 MHz.

(e) On-site wideband analog recording of detected and demodulated data.

(2) RF characteristics:

(a) Frequency range: 2200 to 2300 MHz.

(b) Bandwidth (baseband): 10 MHz.

(c) Antenna gain: 52 dbi.

(d) System temperature: 100K (33K Paramp, antenna at zenith).

(e) Polarization: RCP/LCP/linear.

(f) Cross-support interfaces include telemetry baseband signals from 26-meter control room to SPC, or SPC to 26-meter control room.

B. Command System

Two Spacecraft Command Encoders (SCEs) will be installed at each 26-meter station. The SCE is a Honeywell 316 based controller capable of processing NASCOM command blocks and controlling and testing the output functions for spacecraft commanding. New command software will be utilized to provide a throughput data format. The SCE input channel will be directly connected to an incoming NASCOM circuit. The SCE can perform its own polynomial error detection so that polynomial errors in an arriving block will cause that block to be rejected. At the option of a Project Operations Control Center (POCC), each valid block received can be returned for confirmation of valid SCE reception (command echo). The output (echo) channel will access the wideband link through the Telemetry System PDF, which will perform all necessary output multiplexing functions to the NASCOM circuit.

Uplink support for the Space Transportation System (STS) involves additional unique hardware consisting of a Comm Tech Console, the Delta Modulation System (DMS), and the Shuttle Command Voice Multiplexer (SCVM), as well as the SCE. The DMS receives analog voice from the Comm Tech Console and converts it to digital voice. The SCVM multiplexes digital voice from the DMS with command data from the SCE and applies it to the uplink carrier. The SCVM allows selection of uplink at either the high data rate (72 kb/s) or the low data rate (32 kb/s). The command rate of 128 bits/s is the same for either uplink data rate.

The Shuttle Forward Link will be installed prior to TDRSS becoming operational when the Shuttle requirement for support by the 26-meter Subnet is reduced to an emergency mode only. This device provides for transfer of uplink data (voice and command) in a throughput mode from the Johnson Spaceflight Center (JSC) to the Space Shuttle Orbiter (SSO).

A dedicated Command and Telemetry Processor (CTP), provided by the GOES Project, will be used in place of the SCE for launch and early orbit support of the GOES-series of spacecraft. As a back-up, the SCE will provide non-synchronous command capability.

(1) Key characteristics:

(a) PCM data rates 125 to 2000 b/s for NRZ-L, NRZ-L/Bi-phase-L, and NRZ-M.

(b) Subcarrier demodulation

PSK: 8 to 16 kHz

PSK Summed: 8 to 16 kHz

PSK Summed/FM: 2-kHz PSK/70-kHz FM

PSK AM/FM: 3- to 4-kHz PSK/70-kHz FM

FSK AM: 5 to 9 kHz.

(c) S-band uplink carrier with phase modulation index of 0.3 to 1.5 radians, peak.

(d) Local monitor and control, including system status and keyboard entry for emergency commanding

with voice communications to DSN operations and GSFC POCCs.

(e) Baseband cross-support interfaces include SCE commands to the 34-meter exciter or SPC commands to the 26-meter exciter.

(f) Redundant transmitters at each station.

(2) RF characteristics:

(a) Uplink frequency range: 2025 to 2120 MHz.

(b) Transmitter step tuning: 20 MHz bandwidth with 5-MHz overlap between steps.

(c) Transmitter power: continuously variable, 1 to 10 kw.

(d) Antenna gain: 51.5 dbi.

(e) Polarization: RCP/LCP.

C. Tracking System

One Tracking Data Processor System (TDPS) and STDN Ranging Equipment (SRE) will be installed at each 26-meter station.

The TDPS is a Modcomp minicomputer used to predict spacecraft position and provide pointing information to the antenna-servo system. It can process two sets of spacecraft state vectors simultaneously. The TDPS logs tracking data during a pass and simultaneously formats and transmits an edited version of these data off-site. Tracking data sampling rates are selectable at 0.1, 1, 10, and 60 second intervals.

The SRE measures the range to the spacecraft in nanoseconds, unambiguously, to 6.4 seconds round-trip light time, and S-band doppler to ±220 kHz.

Tracking and ranging data will be transmitted as low-speed data (LSD) via Teletype (TTY) lines, or as high-speed data (HSD) on wideband lines. Acquisition data will be received on TTY lines.

(1) Key characteristics:

(a) 26-meter X-Y antenna with rim mounted acquisition aid antenna.

(b) Antenna control (including autotrack) provided through tracking subsystem (TDPS/SRE/ACC).

(c) Carrier tracking via phase lock loop equipped receivers.

(d) Doppler and range measurements provided simultaneously.

(e) Data formatted and transmitted by tracking subsystem.

V. Network Operations Control Center (NOCC) Support

The JPL NOCC will have operations control responsibility for the 26-meter Subnet stations, as it does for other elements of the DSN, for the long term. However, until the TDRSS becomes operational and the GSTDN phase down occurs, GSFC will participate in the control of station operations. During this period of time the JPL NOCC will exercise operations control for support of the AMPTE mission and the Extended Missions by the 26-meter stations, but will turn over control of the stations to the GSFC NOCC for support of the TDRSS Compatible Missions (refer to Section II). Development of the sequence of events (SOEs) for these spacecraft passes will be consistent with this division of control responsibilities.

During this same period the GSFC will be responsible for scheduling the 26-meter stations. DSN requirements for support will be provided to the GSFC scheduling system via the JPL NOCC. The GSFC-generated schedules will then be supplied to the JPL NOCC for integration into the overall DSN schedules.

The GSFC will also participate in the development of antenna pointing predicts for the 26-meter stations when GSFC has orbit determination (OD) responsibility for the spacecraft requiring support. The Network Support Subsystem (NSS) in the JPL NOCC receives spacecraft position data either in the form of a P-file from the JPL Navigation Subsystem (NAV), or as an Inter Center Vector (ICV) from the GSFC Operations Support Computing Facility (OSCF), to use in the development of Improved Inter Range Vectors (I^2RVs). The I^2RVs are forwarded via TTY to the TDPS at the station for the generation of antenna pointing predicts. In some cases the OSCF at GSFC will generate the I^2RVs and transmit them directly to the TDPS at the station.

VI. Implementation Activities

Implementation of the 26-meter Subnet is a phased activity with the following planned completion dates:

DSS 16: December 1, 1984

DSS 46: October 1, 1984

DSS 66: February 1, 1985

Maintenance, operations and scheduling of the subnet stations will be the full responsibility of GSFC until they are transferred to JPL on February 1, 1985.

The 26-meter Subnet implementation has been divided into two tasks, the Facilities Development Task, for which JPL has management and implementation responsibility, and the Equipment Reconfiguration Task, for which GSFC has management and implementation responsibility.

(1) Facilities Development Task.

(a) Relocate the decommitted DSS 44 26-meter antenna in Australia and the GSTDN 26-meter antenna in Spain, including dismantling, moving, reassembling, and aligning the antenna structural and mechanical subsystems.

(b) Provide support facilities at each complex as required for the 26-meter antenna and its associated electronics (e.g. power, HVAC, roads, cableways, and buildings).

(2) Equipment Reconfiguration Task.

(a) Acquire, relocate, and install the electronics equipment required to implement an operational 26-meter station at each DSN complex.

(b) Provide documentation for the electronics equipment that is adequate for day-to-day maintenance and operations support.

(c) Develop, conduct, and document mission support tests and demonstrations for each station.

A. DSS 16, Goldstone, California

The 26-meter link at the existing Goldstone GSTDN station (GDS), with minimal modifications, will become the 26-meter Subnet station at the GDSCC. The modifications include installation of a PDF and an optical fiber cross-support link between DSS 16 and SPC 10.

(1) PDF installation:

(a) Equipment delivery, October 1, 1984.

(b) Installation and engineering tests complete, December 1, 1984.

(2) Cross-support link: completed.

B. DSS 46, Canberra, Australia

Implementation of DSS 46 is essentially completed with the remainder of the task to be accomplished by station maintenance personnel. DSS 46 is currently supporting the AMPTE Project using the 26-meter antenna and the RF front end in a cross-support mode with DSS 42. The station is scheduled to assume support responsibility for the Extended Missions, the TDRSS Compatible Missions and the STS (refer to section II, Mission Set) on October 1, 1984.

(1) Antenna relocation: completed.

(2) Facilities development: completed.

(3) Electronic reconfiguration and installation: completed.

(4) Maintenance and operations documentation: in process, completion due December 1, 1984.

(5) Mission support tests: in process, completion due December 1, 1984.

C. DSS 66, Madrid, Spain

(1) Antenna relocation: in process for November 1, 1984 completion.

(2) Facilities development: in process for November 15, 1984 completion.

(3) Electronic reconfiguration and installation: in process for December 20, 1984 completion.

(4) Maintenance and operations documentation: in process for February 1, 1985 completion.

(5) Mission support tests: completion due February 1, 1985.

References

1. Yeater, M. L., Herrman, D. T., and Sanner, G. E., "Networks Consolidation Program," *TDA Progress Report 42-59*, July and August 1980, pp. 107-120 Jet Propulsion Laboratory, Pasadena, California, Oct. 15, 1980.

2. Yeater, M. L. and Herrman, D. T., "Networks Consolidation Program," *TDA Progress Report 42-65*, July and August 1981, pp. 19-24, Jet Propulsion Laboratory, Pasadena, California, Oct. 15, 1981.

3. Yeater, M. L., Herrman, D. T., and Luers, E. B., "Networks Consolidation Program," *TDA Progress Report 42-69*, March and April 1982, pp. 10-13, Jet Propulsion Laboratory, Pasadena, California, June 15, 1982.

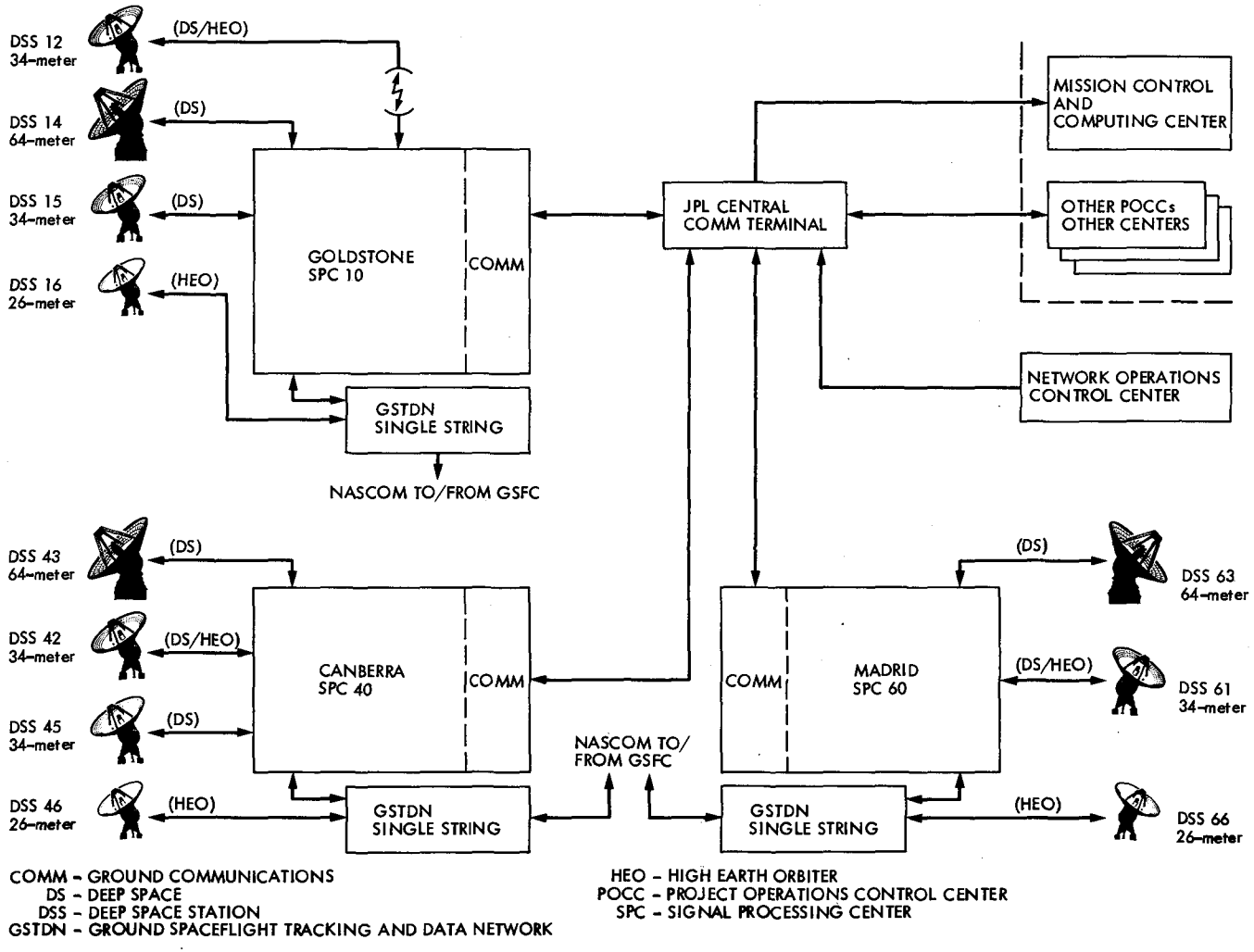


Fig. 1. Mark IVA DSN 1985 configuration

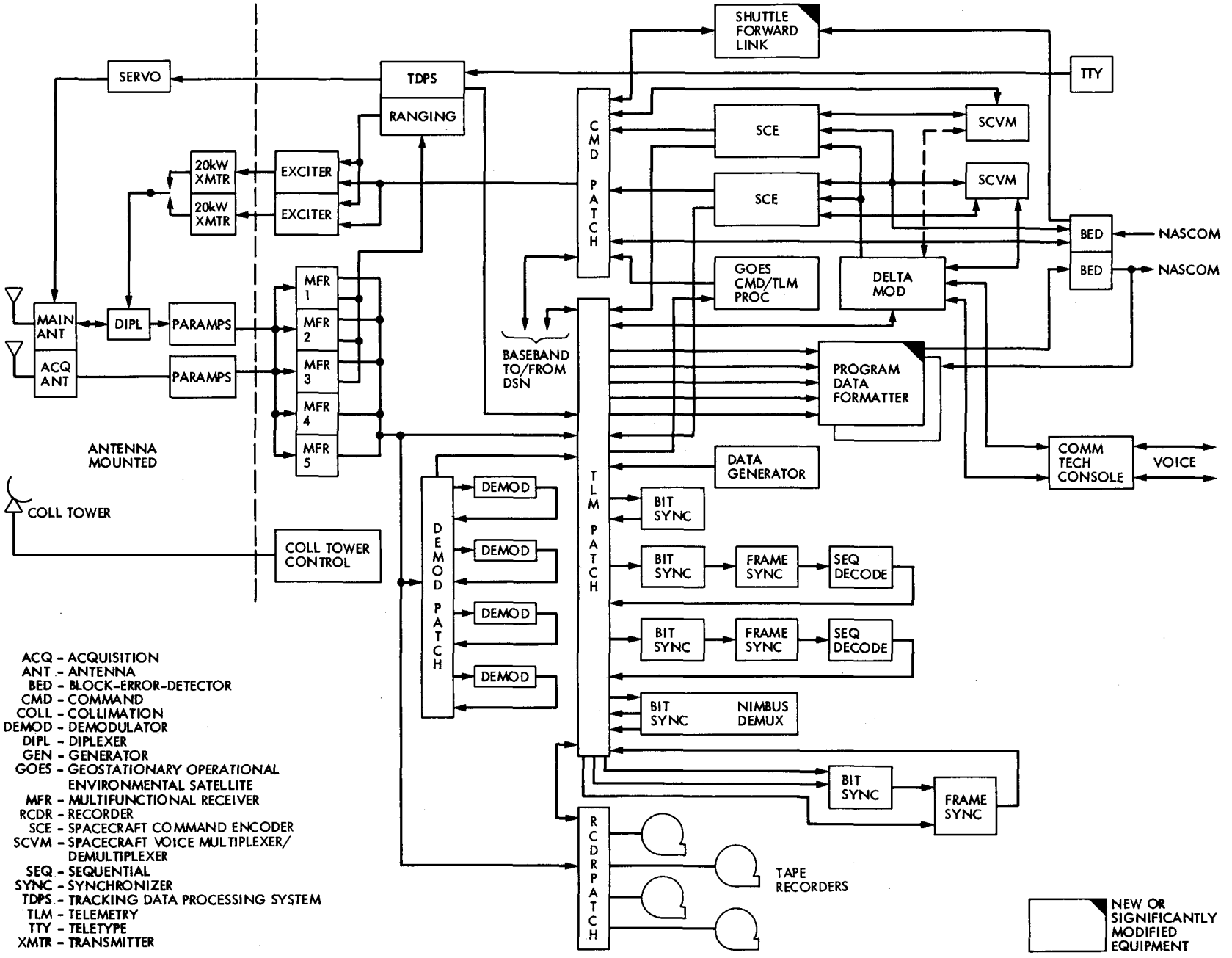


Fig. 2. 26-meter subnet station design

64-Meter to 70-Meter Antenna Extension

D. H. McClure and F. D. McLaughlin
Tracking and Data Acquisition Office

The DSN 64-m antennas are being upgraded to improve their performance by increasing the diameter to 70 m, installing precision panels, and using a shaped surface. Performance improvement is expected to be 1.9 dB at X-band and 1.4 dB at S-band.

This article is the first of a series of articles that will describe the engineering analyses, trade-off studies, and plans that will lead to the performance upgrading of the DSN 64-m antennas. This first article will review the overall concepts and describe some of the approaches used.

The first 64-m antenna was dedicated at Goldstone in April 1966. This was followed by the erection of two additional antennas, one in Australia in July 1972 and one in Spain in January 1973. Following the completion of the 64-m antennas, a DSN Advanced Systems program was started in which potential improvements in the microwave performance were explored. This program resulted in a menu of options considered technically feasible and cost effective. (This work was reported in JPL Document No. 890-47 by J. R. Hall, June 1, 1975.) This menu included:

(1) Install improved feed waveguides, dichroic filters, and travelling wave masers.

(2) Install an X-band dual hybrid horn.

(3) Extend the primary reflector from 64-m to 70-m.

(4) Install a highly efficient shaped primary reflector and subreflector.

Since the first two items were considered the most economical and cost effective, they were implemented to provide a needed performance improvement for the Voyager encounter at Jupiter in 1979.

Drawing on the results of the 1975 study, a 1982 study indicated that it would now be economical to improve the performance of the 64-m antenna by 1.9 dB at X-band and 1.4 dB at S-band by increasing the diameter to 70 m, installing a new precision-shaped surface, adding structural braces, and installing a Y-axis controller. The last two items are intended to compensate for gravity distortions. The structural braces were installed on the Goldstone antenna in 1973 with an improvement in the gain characteristics as a function of elevation angle. The Mark IVA Project, currently being implemented in the DSN, includes the installation of these braces at Australia and Spain, and the installation of the Y-axis controllers at all three antennas.

Figure 1 is an illustration of the antenna that shows the principle elements of the performance upgrade. These elements are:

(1) Extend the diameter from 64-m to 70-m. Enlarge the diameter of the main reflector to 70-m, the maximum practical limit developed in the 1975 study. This includes the following:

(a) Replace and extend radial ribs. The supporting structure beyond the 34-m diameter will be

removed and replaced. (The time available to install these changes is very limited and the plan is to replace, rather than rework, the existing ribs. The new ribs will be assembled on the ground before the antenna is removed from service.)

(b) Reinforce reflector backup structure. It is impractical to follow the same philosophy as for the ribs so the necessary reinforcements will be welded on during the downtime.

(c) Reinforce elevation wheel. The elevation wheel, which carries the counterweight, will be strengthened to support the additional counterweight.

(d) Additional counterweights. The added weight of the reflector, quadripod, and subreflector will require additional counterweight.

(2) New surface panels. New reflector surface panels, which will be smoother and fabricated to a shaped surface, will be installed.

(3) Newly shaped subreflector. The larger diameter main reflector will require a larger diameter subreflector. In addition, the subreflector will be of a shaped design and manufactured to a high precision.

(4) New quadripod. The larger subreflector will require the quadripod to be redesigned to carry the additional weight.

(5) Modify microwave feed. A 22-dB-gain X-band feed will replace the existing feed. The existing narrowband X-band feed was developed to provide a pseudo-shaped beam to be used with a parabolic main reflector surface. It will not perform satisfactorily with a shaped surface.

(6) Add 1-m S-band noise shield. Optimizing performance at X-band results in a significant loss of performance at S-band, caused by an increase in system temperature due to edge spillover. This is compensated for by adding a screen to the edge of the main reflector.

The source of the performance increase is illustrated in Fig. 2. The main difference in performance improvement between X- and S-band is due to the fact that improvement in surface accuracy has little effect at the lower frequency. The physical increase in aperture and the use of a shaped, rather than a parabolic, surface yields the same performance improvement at both frequencies.

There are several decisions and design optimizations required to get the largest performance improvement while keeping the cost within the established bounds. These include the following, each of which will be the topic of a future article.

1. Subreflector RF optical design. The feeds of the 64-m antenna are not on the antenna centerline. Three feedcones are mounted offset such that the subreflector must be rotated to "see" each of them as required. The existing configuration is a conventional hyperboloid-paraboloid reflector system, with the subreflector a section of a hyperboloid. When the reflectors are "shaped," to improve illumination efficiency, this subreflector becomes an asymmetric shape, which complicates the design and fabrication, and hence increases the cost. The question then becomes: What is the impact on performance if the subreflector is made symmetrical? Furthermore, is there a surface of revolution that approximates the desired shape close enough so that further economies might be made without compromising the RF performance?

There is also the question of the proper diameter. The larger main reflector requires the diameter of the subreflector to be increased. There is an optimum size based on RF efficiency, noise temperature, weight, and ease of fabrication. Weight affects quadripod design, and as will be pointed out later, this can have a significant effect on the overall performance of the antenna.

2. Noise Shield. Designing an antenna for more than one frequency is a compromise. Originally this antenna was designed for optimum performance at S-band. In the years that followed, X-band was added, and new antennas now being built for the DSN are optimized for X-band. Early performance analyses of the 70-m antenna showed that when the system was designed for the higher frequency, noise from edge spillover almost negated other improvements at S-band. One way to avoid this is to include a flange around the edge of the subreflector, and, in fact, this has been done on some of the DSN antennas in the past. The trouble with this approach is that this tends to be frequency sensitive; that is, while it may help at S-band and has no effect at X-band, it may have deleterious effects at K-band. Another solution is to install a "noise shield" around the edge of the main reflector, such as was done on the 100-m antenna in West Germany. It must also be remembered that the 1975 study showed that diameters much over 70-m required extensive drive modifications. Therefore, any added shield must be designed to not only have the desired RF performance but to minimize the impact on structural and mechanical performance.

3. Feed configuration options. Figure 3 illustrates the feed configurations under consideration. The shape of the reflectors is dependent on the characteristics of the horn used to illuminate the main reflector. Three X-band horns are under consideration: (1) the horn currently in use on the 64-m antennas,

(2) the new X/S horn being installed on the new 34-m antennas, or (3) the standard 22-dB-gain horn used on the 64-m antennas until 1979. Since the horn currently in use was designed to provide quasi-shaped performance, it is not practical to consider it in the design of a shaped surface. However, it is practical to consider it as a temporary feed for a reflector shape based on one of the other two (i.e., the 22-dB standard horn or the X-S common aperture feed). Therefore, there are the six possible configurations shown in Fig. 3.

4. Quadripod design. The quadripod performs a difficult function: it must support the subreflector in a fixed position and be relatively invisible to the RF beam. It also is used as a "crane" during the removal of the subreflector and feedcones. Thus, a design must be found that provides adequate stiffness and strength while minimizing the RF blockage.

5. Operating environment. The 1975 study pointed out that 70-m was about the maximum diameter that could be considered before significant changes would be required in the antenna drive system. The original 64-m antenna design provided for maximum antenna motion in either axis of 0.5 degrees per second. After the antenna was built, the maximum operational velocity was set to 0.25 degrees per second. Maximum operating wind velocity for the 64-m antenna is 88 km/h (55 mph). At that point, there is sufficient drive torque to move the antenna to the stow position. A 70-m antenna, with a 1-m noise shield, could encounter attitudes at which it could stall at wind velocities below 88 km/h (55 mph), though not for both axes simultaneously. The problem becomes one of identifying the controlling factors and either modifying the design or instituting acceptable operational restrictions.

6. Construction. In considering when to upgrade the 64-m antennas, careful consideration had to be given to the support of on-going and planned deep-space missions. Figure 4 is a plot of Deep Space Network loading for Priority 1 support. A characteristic of Priority 1 support requirements is that they are essentially inflexible and must be fully met on the committed and planned schedule. There is only one opportunity before the Voyager encounter at Neptune: mid-1986 to the end of 1987. This means that only eighteen months are available to modify in succession the three antennas. Each will require six months to remove the outer rib assemblies, remove the quadripod and subreflector, strengthen the backup structure, replace the outer ribs, quadripod and subreflector, install and set the new surface panels, and perform the necessary tests. The accomplishment of this task, in such a short time, will require a careful and well thought-out detailed construction plan.

1. EXTEND 64-m TO 70-m
   REPLACE AND EXTEND RADIAL RIBS
   REINFORCE REFLECTOR BACK UP STRUCTURE
   REINFORCE ELEVATION WHEEL
   ADDITIONAL COUNTERWEIGHTS
2. NEW SURFACE PANELS, SHAPED
3. NEW SUBREFLECTOR, SHAPED
4. NEW QUADRIPOD
5. MODIFY MICROWAVE FEED
6. ADD 1-m S-BAND NOISE SHIELD

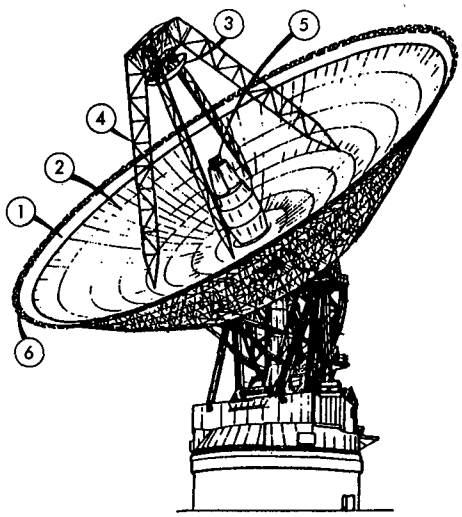


Fig. 1. Modifications to be made to the 64-m antenna

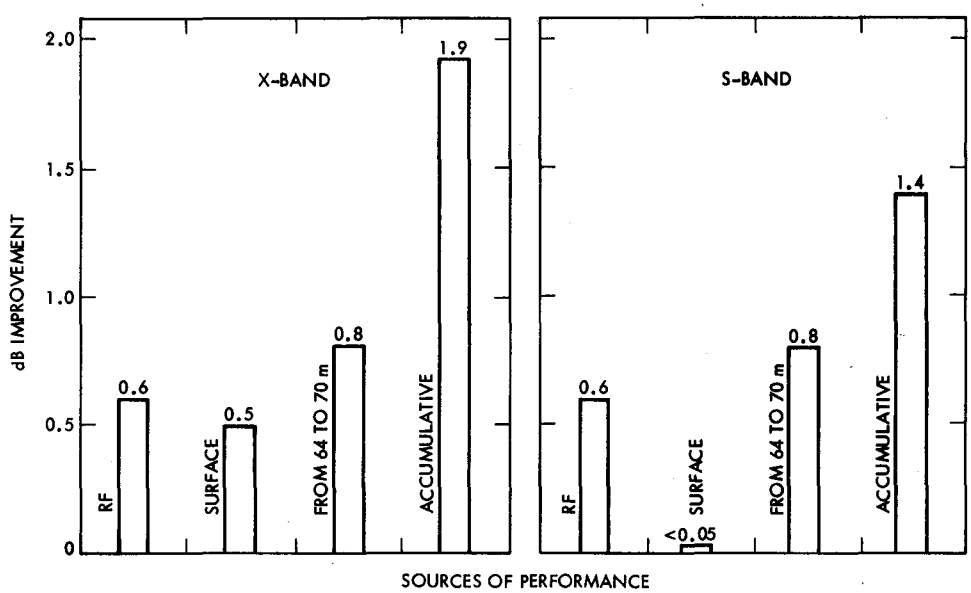


Fig. 2. Sources of performance increase for 70-m antenna

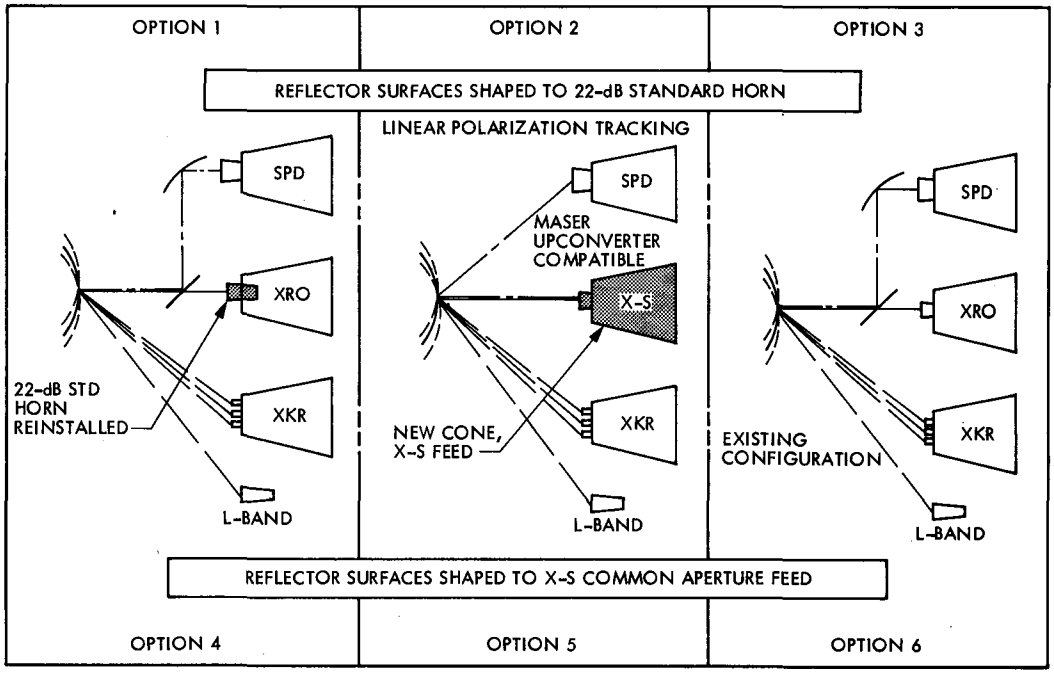


Fig. 3. Feed configurations for 70-m antenna

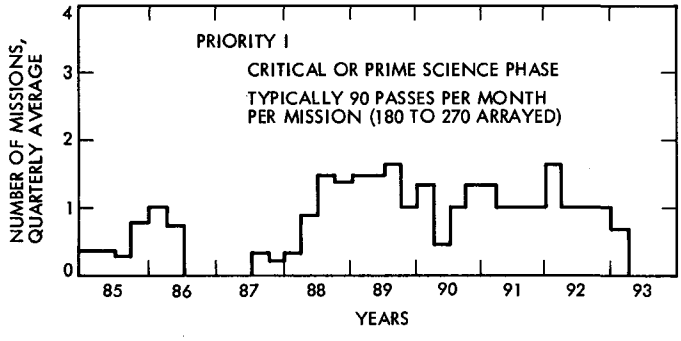


Fig. 4. Network loading

Antenna Microwave Subsystem

P. L. Parsons
TDA Engineering Office

The Deep Space Net consists of 64-m, 34-m, and 26-m antennas, each with its own characteristics. This article describes the changes that are being made to each under the Mark IV program and additions required for some special spacecraft projects.

I. Introduction

The antenna microwave subsystem includes the antenna feeds, the low-noise amplifiers (LNA), and the transmission paths between them. Various block diagrams later in the article provide a graphic definition of the subsystem. The feeds consist of the feed horn, a polarizer, perhaps an orthomode transducer for polarization diversity, and other specialized components as required by the particular feed. The LNA may be a traveling wave maser (TWM) or a field effect transistor (FET), which may be cooled or uncooled. The transmission path may include switches, a diplexer, couplers for various purposes, and waveguide or semirigid coaxial cable.

Work in the antenna microwave subsystem since the last report has been concentrated on completion of the MK IVA implementation. Specific tasks include: modification of 34-m track antennas to include HEO frequencies; acceptance testing of the feed cone for the new 34-m HEF antennas; modification of 64-m stations to accept International Comet Explorer (ICE) dual-frequency reception and ICE and AMPTE uplink frequencies.

Other tasks include an L-band receive capability for the Venus Balloon mission and design of the 70-m extension for the 64-m stations.

II. MK IVA Status

A. 34-m Standard Antennas

The 34-m standard stations are used for transmitting commands to, and receiving telemetry and metric data (range and doppler) from, deep-space and earth orbiting spacecraft. They will be used to supplement the 26-m stations to be acquired from GSTDN in tracking earth orbiters, although their original design did not allow operation at those frequencies. Modifications are presently underway to extend the S-band frequency range.

These stations have reflector antennas that are 34-m in diameter and Cassegrainian optics and feed horns for S- and X-band frequencies. The antenna mount is hour-angle declination (HA-DEC). Combining of the subreflector illumination paths of the two frequency bands is provided by a dichroic mirror assembly, permitting simultaneous operation at both bands.

In each band, the low-noise amplifier for deep-space frequencies is a maser. Because the earth orbiting spacecraft (S/C) do not require the low system temperature obtainable with a maser, a room temperature FET will be provided for those S/C frequencies at S-band. The microwave subsystem covers the

entire X-band range (8400 to 8500 MHz), but receiver components restrict the upper frequency to 8440 MHz.

A 20-kW transmitter capability is provided at S-band. Transmission at X-band would require redesign of portions of the antenna feed and dichroic assembly, and is not planned.

Figure 1 is a simplified block diagram. Only major elements are shown, assemblies being represented by rectangles and transfer switches by circles.

The dichroic assembly allows simultaneous transmission of S-band and reception of S- and X-band signals. Right circular polarization (RCP) or left circular polarization (LCP) can be selected for each of the feeds, but neither has an orthomode; therefore, only one polarization is available at a time. The second maser is, therefore, a backup to the first. Although the diagram shows two X-band masers installed, only Madrid (DSS 61) has this configuration so far.

Modification of these antennas for orbiter frequencies is proceeding in two phases, temporary and final. The temporary phase, now completed, extended the transmit range to 2090 to 2120 MHz and the receive range to 2270 to 2300 MHz. It has been accomplished by replacement of some waveguide components and tuning of the remainder. The final phase will provide a subsystem capable of operating over the frequency ranges of 2025 to 2120 MHz uplink and 2200 to 2300 MHz downlink.

A redesign of certain components, particularly the diplexer, has been necessary to obtain the required bandwidth. The design work has been completed and installation will be completed as scheduled during the MK IV downtime.

A distribution assembly, for each of the frequency bands, is being designed for DSS 42 and will likely be installed at all of the sites.

B. 34-m High-Efficiency (HEF) Antenna

The high-efficiency antennas, so called because the aperture efficiency at X-band is quite high (72%), will be used to supplement the standard 34-m net in tracking deep-space craft. They will also be arrayed with the 64-m antennas to increase the received SNR, thus allowing reception of a higher data rate (Ref. 1).

These antennas are being installed during Mark IV. They differ from the 34-track stations primarily in having an elevation over azimuth mount (AZ-EL), a single feed horn for both frequency bands, and no transmitter. The common feed horn eliminates the need for a dichroic mirror assembly and makes the implementation neater but does require compromises.

Figure 2 is a simplified block diagram. Either RCP or LCP may be selected from the S-band output of the S-X combiner. The X-band output is followed by an orthomode transducer that provides the possibility of simultaneous RCP and LCP. The initial implementation uses only one port of the orthomode because only one maser and receiver are installed. Future additions will include a diplexer, an X-band transmitter (Ref. 2), a second maser and a wide band FET for VLBI use.

The X-band LNA in these cones is a BLK II maser that has been transferred from the 64-m station. Two new X-band masers, BLK IIA, have been built for the 64-m stations to replace the BLK II units. These masers have been transferred to operations.

The S-band LNA is a room-temperature FET designed primarily for VLBI use. There are plans to replace it with a cooled FET to reduce the system noise temperature from a projected value of 110 K to about 50 K. This will make the station more suitable for spacecraft use.

The cone assemblies for the new 34-m antennas have been accepted from the contractor and have been transferred to operations. These cones provide for reception at S- and X-bands through a common feed horn. The design is optimized for X-band, therefore the antenna does not have as high a level of efficiency at S-band as is possible to achieve.

Performance measurements of the complete antenna are scheduled for this summer.

C. 64-m Antennas

The 64-m stations are used for transmitting commands to and receiving telemetry and metric data from spacecraft in deep space. They are also providing backup support to AMPTE.

The 64-m stations have three feed cones installed on a structure at the center of the reflector. The feeds are 120° apart on a circle. Selection of a feed is made by rotation of the (asymmetric) subreflector. A dichroic mirror assembly, half on the S-band and half on the X-band cone, permits simultaneous use of S- and X-band frequencies. The third cone is devoted to R&D or host country work. The planetary radar system also uses the third cone.

Figure 3 is a simplified diagram of the X-band portion of the subsystem. The orthomode provides the capability of simultaneous RCP and LCP. No changes are being made to these units.

Figure 4 is a simplified block diagram showing the S-band portion of the 64-m subsystem. Assemblies that were changed

during Mark IVA are identified with a solid triangle in the lower right corner.

The S- and X-band feeds each have orthomodes and two masers, thus permitting simultaneous reception of RCP and LCP. Differentially rotatable quarter-wave plates in the S-band feed allow any angle of linear polarization to be selected instead of circular for those spacecraft that use linear polarization. Because of changes for MK IV, autotracking of the polarization vector is no longer possible.

The S-band waveguide is cooled to allow transmission of CW at up to 500 kW. There is presently no X-band transmitter and major changes to the feed system will be required when one is added.

The microwave subsystem covers the earth orbiter X-band frequency range, but not the S-band transmit frequencies below 2090 MHz because of high VSWR. The nominal lower downlink frequency limit is 2270, but the second ICE frequency of 2217 has been provided for by the replacement, at DSSs 14 and 63 only, of the second maser by a new, BLK V, maser that was built especially for this purpose.

Special modifications are required to the 64-m stations to accommodate ICE, AMPTE and Venus Balloon/Halley Pathfinder missions. These changes include extension of the transmit bandwidth (from 2090 to 2120 MHz with limitations) (Ref. 3), extension of the maser bandwidth (2200 to 2320 MHz in the new, BLK V, maser) and the addition of L-band (Ref. 4).

The 64-m stations must be able to transmit in excess of 100 kW to ICE at 2090 MHz and 20 kW to AMPTE at three frequencies between 2090 and 2105 MHz. All of these frequencies are beyond the original design capability of the station, the VSWR being too high for the transmitter. Extensive measurements have preceded the design of a matching section that provides an excellent match at 2090 and 2110 to 2120 MHz. The sections have been completed and sent to the complexes. This permits the stations to transmit at high power (>20 kW) at the match frequencies, but not between 2091.5 and 2109 MHz. AMPTE requirements for transmission at 2093 and 2103 MHz are being met by installation of isolators in the low-power transmitter output waveguide. The S-band preamp filter (SPF) was retuned to accommodate the new transmitter frequency range (2090 to 2120 MHz).

A new maser has been designed and built to receive the second ICE frequency of 2217 MHz. These masers will be installed in the tricone position at DSSs 63 and 14. Reception at this frequency requires another filter in the transmitter waveguide to reject klystron beam noise at this new receive frequency. These filters, and a matching section, have been built and will be installed on schedule.

The DSN is committed to providing tracking and telemetry support to the Venus Balloon/Halley Pathfinder mission. Spacecraft transmission will be at 1668 MHz, a frequency well outside the normal DSN band. To accommodate this frequency, a new approach has been taken. The design includes a new feed horn mounted to the outside of the XRO cone, dual cooled FET amplifiers, and an upconverter to convert the signal to S-band. The converted signal is put into an unused port in the maser select assembly and follows the normal S-band signal path from that point.

Figure 5 is a simplified block diagram. The second FET is a backup to the first and only one polarization (RCP) is provided. LCP can be provided by unbolting one of the polarizer segments, rotating it and rebolting it in place.

Although each refrigerator (CCR) is capable of holding two S-band and two X-band FETs, it can hold only one of the larger L-band FETs. Therefore, separate CCRs are required for the main and the backup FET. Since each CCR requires a compressor, two new compressors have been obtained, one each for DSS 43 and 63. The other CCR at those sites will use the spare DSN compressor that is located, with the operational compressors, on the antenna structure. At DSS 14, the DSN spare and the R&D compressor will both be used. Switching of helium lines will be required to bring R&D or radar equipment back on line.

III. Studies

A. 70-m Extension (Ref. 5)

Selection of a feed for the X-band cones has been made. Studies of material for the noise shield have been completed. Other RF studies are continuing within the limits of available resources.

References

1. Stevens, R., "Applications of Telemetry Arraying in the DSN," *TDA Progress Report, 42-72*, pp. 18-82. Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1983.

2. Meeker, J. G., and Timpe, C. T., "X band Uplink Technology Demonstration at DSS 13," *TDA Progress Report 42-77*, pp. 24-32. Jet Propulsion Laboratory, Pasadena, Calif., May 15, 1984.

3. Galvez, J. L., Marlin, H., Stanton, P., "ISEE-3 Microwave Filter Requirements," *TDA Progress Report 42-76*, pp. 114-119. Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1984.

4. Withington, J. R., Reilly, H. F., Bathker, D. A., "RF Performance of a Proposed L Band Antenna System," *TDA Progress Report 42-75*, pp. 91-97. Jet Propulsion Laboratory, Pasadena, Calif., Nov. 15, 1983.

5. Stevens, R., "Implementation of Large Antennas for Deep Space Mission Support," *TDA Progress Report 42-76*, pp. 161-169. Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1984.

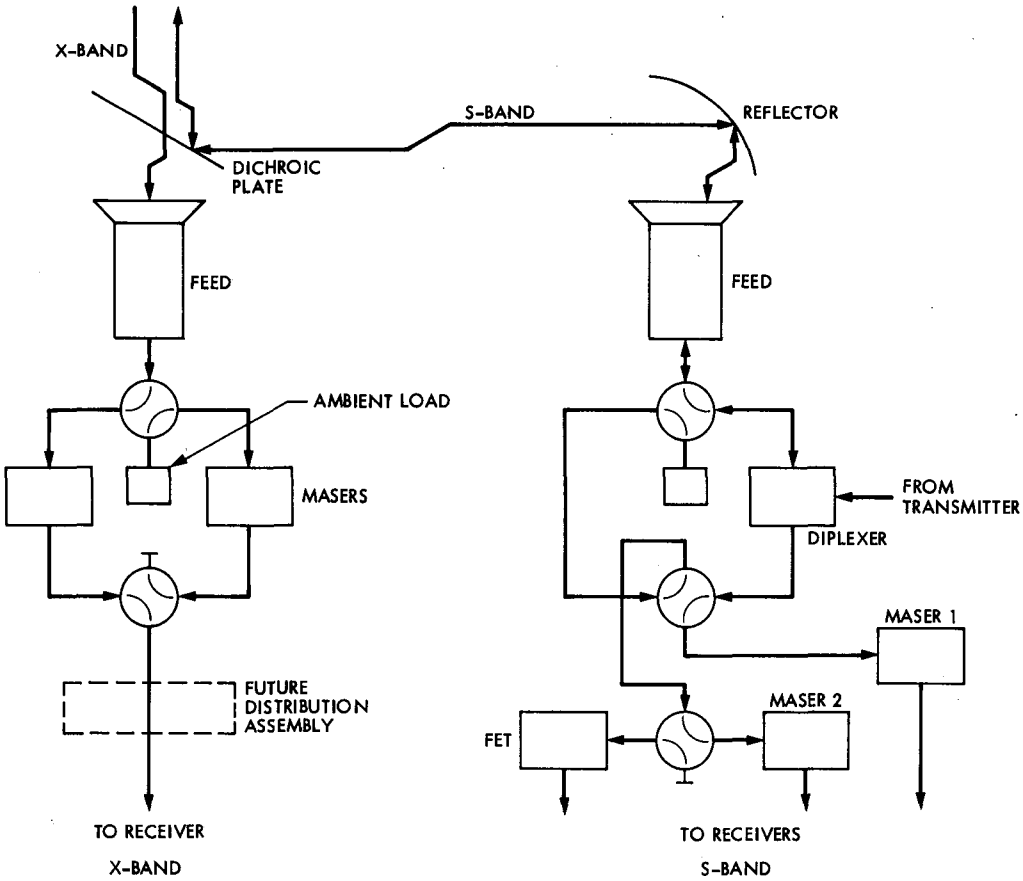


Fig. 1. Simplified block diagram, 34-m standard stations

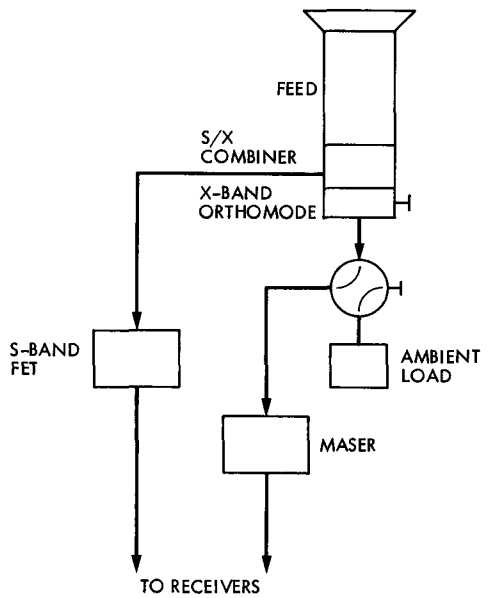


Fig. 2. Simplified block diagram, 34-m HEF stations

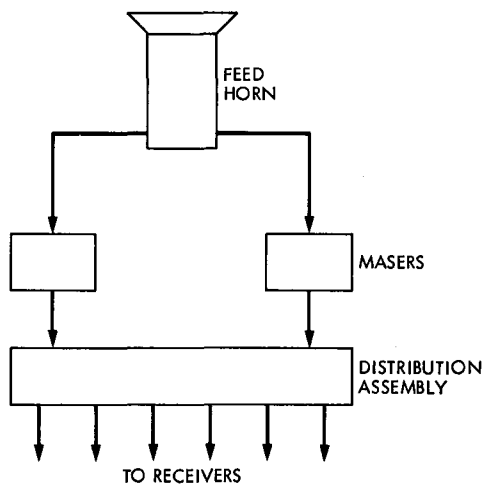


Fig. 3. Simplified block diagram, 64-m station X-band

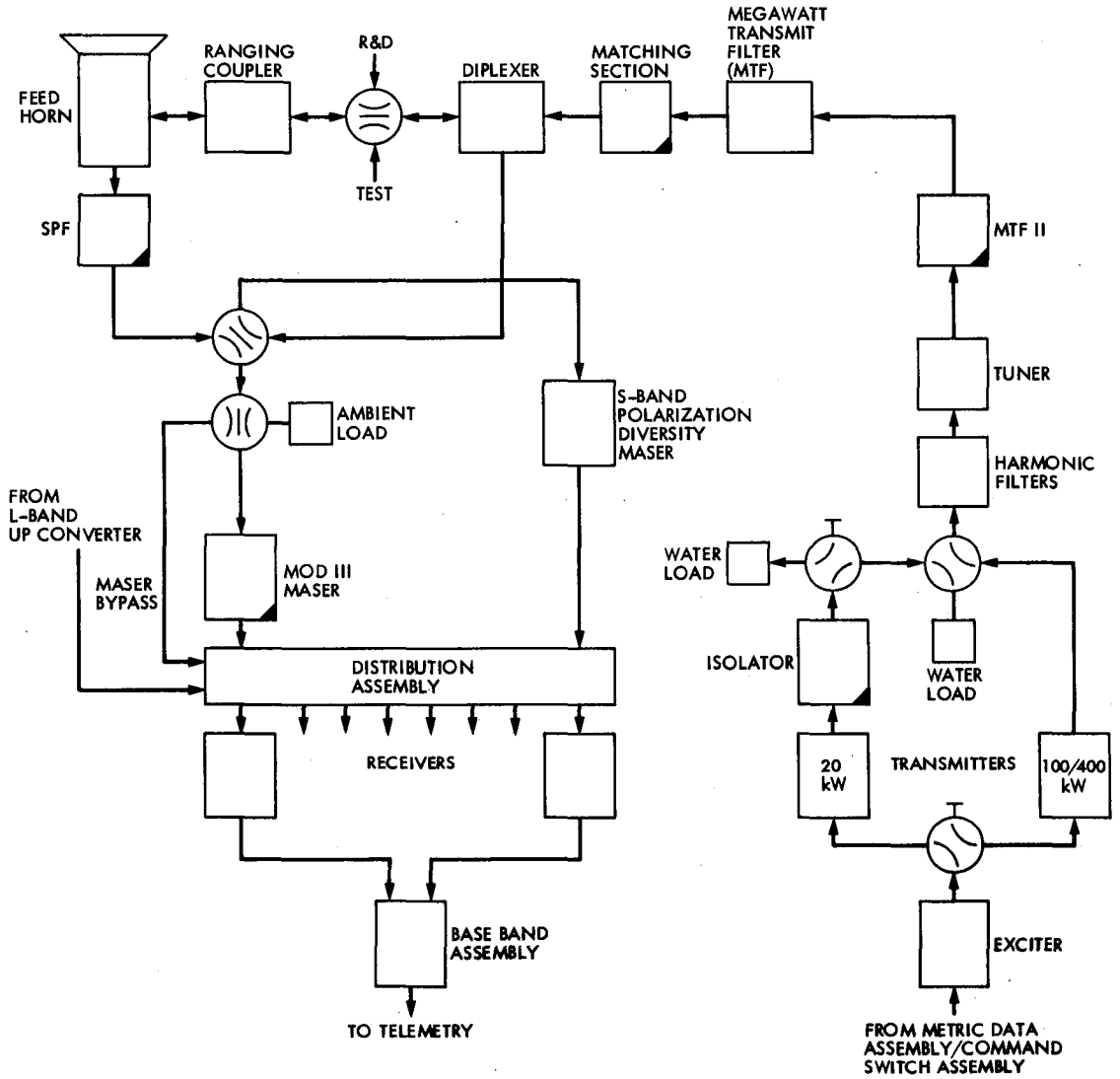


Fig. 4. Simplified block diagram, 64-m station S-band

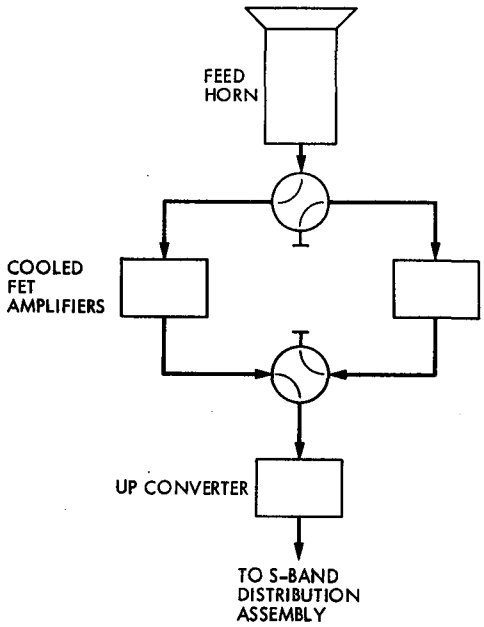


Fig. 5. Simplified block diagram, 64-m station L-band

VLBI System (BLK I) IF-Video Down Conversion Design

N. C. Ham
Telecommunications Systems Section

The prototype IF-video down converter of the DSN BLK I VLBI system, a hybrid of analog and digital circuits, has been completed and tested. The initial results have met all design specifications, establishing a milestone in new design technology for such devices. This device has uses in other similar applications and can form the basis for expanded design and applications.

I. Introduction

The IF-video down converter is an assembly of the Deep Space Communication Complex (DSCC) BLK I Very Long Baseline Interferometry (VLBI) system which down converts the receiver intermediate frequency (IF) spectrum to video-band frequencies. The video-band spectrum, in digitized form, is then stored on high-capacity magnetic discs and simultaneously transmitted via wideband data lines (WBDL) from the DSCC to the VLBI Processing Center at the Jet Propulsion Laboratory (JPL) in near-time for further processing with similar data from other DSN complexes. The VLBI signal can be either an extragalactic radio source (ERS) or a spacecraft line-spectra signal (Ref. 1).

A fundamental problem in down converting a radio frequency (RF) and IF spectrum to the lower video-band frequencies is that of suppressing the image (or adjacent band) noise which would cause signal-to-noise ratio (SNR) degradation to the received signal spectrum.

Similarly, in the conversion, the signal phase information in the video output data must not have been subjected to phase instabilities and fluctuations. This is an essential requirement when the system is used to accurately determine the group delay of a limited span bandwidth spacecraft radiated signal during a delta differenced one-way range (ΔDOR) observation. The alternate observation of a spacecraft and ERS signal during ΔDOR aids in the navigation of the spacecraft into deep space.

A prototype down-converter assembly, comprised of hybrid analog and digital circuits, has been designed, constructed and tested and meets the specified system requirements for image-band rejection and limited SNR degradation.

Other important specifications met were the capability for down converting narrow segment bandwidths of the wide bandwidth RF S- and X-band spectra through separate wide-band RF channels, providing a dynamic signal amplitude reception range, selectable channel output bandwidths, and low imparted phase instabilities and fluctuations.

The design specifications, including error budgets and element contribution limits, that form part of the overall VLBI system design are listed in Table 1.

II. Design Configuration and Rationale

The basic design of the down-converter is an extension of an existing all-analog design which has been used for this application and achieves the image spectrum rejection by utilizing a phase quadrature cancellation technique (Ref. 1).

Figure 1 is the block diagram of the converter and labels the analog and digital portions. Major components are the analog IF mixer, local oscillators and selector, video amplifiers and anti-aliasing filters, and digital input signal multiplexer, analog-to-digital converter (ADC), 90° phase shifter, combiner, low-pass filters, decimator and formatter, and computer interface for control and operation.

The hybrid design exploits the advantages of both circuit domains for meeting the design requirements. For example, the analog circuits are best applicable at the IF range from 265 to 400 MHz for achieving the IF-LO conversion to video band and in maintaining the quadrature phase for less than an octave range.

The implementation of the second quadrature phase shift over a 5-octave range while maintaining low-phase variation (ripple) was achieved in the digital domain. Similarly, the low-pass filters, decimator, and formatter are better achieved in this domain and in ensuring the success for overall repeatable characteristics and reproduction.

Two IF-video converters (analog portion) are used for the S- and X-band spectra to break the frequency ratio into lesser ranges of 265 to 305 MHz and 300 to 400 MHz, respectively, to aid in maintaining the quadrature phase, and video signal amplitude match. Lumped circuit hybrids were used to obtain the quadrature phase shift at these local oscillator frequencies (Ref. 2).

Eight frequency synthesizers are available for the X-band spectrum and four for the S-band spectrum down conversion. Figure 2 illustrates a typical spacecraft spectrum such as will be transmitted by the Galileo spacecraft during a ΔDOR observation. The two extreme sideband tones (X-band) are used to determine the signal group delay, while the close-in side tones and carrier signals are used to resolve the delay ambiguity in determining the outermost sideband tone phases. This is necessary when using several narrow bandwidth channels to determine the broader bandwidth signal group delay and is known as bandwidth synthesis (BWS) (Ref. 3).

The individual frequency synthesizers are tuned to align with the appropriate sidebands (S- and X-band spectra) which are sequentially time multiplexed in a cyclic manner to obtain their phase. Each oscillator is left operating at its tuned frequency during the multiplexing procedure to maintain continuity of the signal phase. A typical multiplexed sequence is shown in Fig. 3 for both the spacecraft signal and the alternate ERS spectrum (at approximately the same channel frequency) during a ΔDOR observation.

Only the ERS spectrum is observed when the system is used for obtaining VLBI data for time synchronization and Earth motion precision observations, and the system will utilize the full bandwidth of the S- and X-band IF spectra and the frequency synthesizers tuning capability.

For the digital portion, the two quadrature video-band signals from the mixers (Fig. 1) are input to the analog-to-digital converter for digitization. The two digitized signals are then input to the phase shifter for shifting the video spectrum of one signal an additional 90 degrees relative to the other. The phase shifter design is based on a 90° phase-shifting network using the Jacobian elliptic function together with the bilinear transform.

A pair of all-pass networks was derived whose phase difference was the closest approximation to 90° over the range of video-band frequencies. A total of 6 poles and zeros was used to achieve the desired response.

The digital low-pass filter design was adopted from an existing computer program (Ref. 4) for elliptic filters.

The hybrid converter design differs in another major way from previous all-analog designs in that the signal amplitude linearity must be maintained throughout the complete signal path (including digitization and digital processing) until the low-pass filtering is accomplished. Digitally this prevents register overflow and preserves the computation integrity during the video-band phase shift, image-band rejection and low-pass filtering process. Following this processing, the signal is reduced in data volume by sequential data selection (decimated, the selection rate is reducible to 1/16 in 1/2 increments). The reduced data rate coincides with the selected low-pass filter bandwidth while preserving the Nyquist criterion. Finally, the data are formatted into a 1-bit quantization as the data form. The amplitude differences in the carrier and sideband tones (shown in Fig. 2) together with the necessity for linearity define one requirement for the dynamic range of the ADC. Another requirement is to accommodate the amplitude difference between the spacecraft and ERS signals as the observation alternates between these two signals sources.

Although the dynamic range of the ADC was adequate to accommodate the amplitude differences of the above signals during a given observation period, it is not adequate (nor was it considered necessary) to handle the extremes of the received spacecraft signal level throughout the entire mission-time profile. Figure 4 illustrates the anticipated signal power-time profile for the Galileo mission.

Since the power level extremes occur over many days and the ADC dynamic range can accommodate the expected range during segments of this profile, a computer-controlled attenuator at the IF amplifier output prior to the input of the frequency down-converter can be preset during any observation day to optimize the expected signal level range to the ADC characteristics. This is applicable to either the line-spectra spacecraft or ERS continuous-spectra signals during either a ΔDOR or time and Earth-motion observations (Ref. 2).

III. Design Detail and Error Budget

There were subtle design details that were confronted in meeting the overall requirements, such as the technical management to allow subassemblies to be designed and developed with individual specification using independent and separate schedules, personnel, and laboratories.

The requirements were critically specified, top down from the system requirements, using engineering judgment towards the error budget allocation with the apportionment balanced on the state-of-the-art design and components technology, complexity, and ease of duplication. Budget allocation was reassessed and trade-offs made during the development with the greatest tolerance value rebudgeted to the most difficult circuit domain.

A typical example is the specification of parameters that affected the image rejection and phase variation. The parameter error sensitivity was formulated and provided the basis for distributing the budget allocation.

For example, from the diagram of Fig. 1 the equation for the top mixer and video amplifier string signal can be written as,

$$V_{TO} = V_S \left[\cos(\omega_S \rightarrow \omega_{US})t + \cos(\omega_S \rightarrow \omega_{LS})t\right] \quad (1)$$

(Note that $\omega_S \rightarrow \omega_{US}$ indicates the range of frequencies from ω_S to ω_{US} [Ref. 1].) The bottom string can be written as,

$$V_{BO} = (V_S + \delta)\left[\sin(\omega_S \rightarrow \omega_{US})t + \phi_A - \sin(\omega_S \rightarrow \omega_{LS})t + \phi_A\right] \quad (2)$$

where $(\omega_S \rightarrow \omega_{LS})$ is the lower spectrum band from the signal frequency ω_S and $(\omega_S \rightarrow \omega_{US})$ is the upper spectrum. The frequency ω_S is also the edge of the desired down-converted spectrum and is equal to the local oscillator signal ω_{LO} (Ref. 1). δ_A and ϕ_A are the amplitude (from equal value) and phase (from 90 degrees difference) errors of the two analog output signals. Small error values or near-perfect amplitude and quadrature phase tracking implies near-infinite image-band suppression.

Following digitization, the upper signal is shifted an additional 90° relative to the lower signal,

$$V''_{TO} = V''_S \left[\sin(0 \rightarrow \omega_{US})t + \phi_D + \sin(0 \rightarrow \omega_{LS})t + \phi_D\right] \quad (3)$$

where V_S is the new amplitude from the ADC. Sin 0 represents the zero frequency or dc component since $\omega_S = \omega_{LO}$, and ϕ_D is the error of the digital 90° phase shifter in producing a true quadrature phase to all frequencies across $0 \rightarrow \omega_{US}$ and $0 \rightarrow \omega_{LS}$.

Similarly, the bottom signal input to the combiner is,

$$V''_{BO} = (V_S + \delta)''\left[\sin(0 \rightarrow \omega_{US})t + \phi_A\right] - \left[\sin(0 \rightarrow \omega_{LS})t + \phi_A\right] \quad (4)$$

The combiner sums the two input signals, expressed by Eqs. (3) and (4) and cancels the lower sideband or image spectrum and obtains the desired passband spectrum as follows:

$$V_{\text{CANCEL}} = V''_S \left[\sin(0 \rightarrow \omega_{LS})t + \phi_D\right] - (V_S + \delta)\left[\sin(0 \rightarrow \omega_{LS})t + \phi_A\right] \quad (5)$$

$$V_{\text{OUT}} = V''_S \left[\sin(0 \rightarrow \omega_{US})t + \phi_D + \sin(0 \rightarrow \omega_{US})t + \phi_A\right] \quad (6)$$

The vector diagram of Fig. 5 illustrates the upper and lower spectra signals relative to Eqs. (1) through (6). From the combiner summed output diagrams of Fig. 5(c) and (d) showing the desired passband and reject band signals, an equation for the image rejection magnitude can be derived.

The passband signal amplitude is

$$V_{\text{PASS}} = \left\{\left[(V_U + \delta) + (V_U \cos\sigma)\right]^2 + (V_U \sin\sigma)^2\right\}^{1/2} \quad (7)$$

and the reject band signal amplitude is

$$V_{\text{REJECT}} = \left\{\left[(V_L + \delta) - (V_L \cos\sigma)\right]^2 + (V_L \sin\sigma)^2\right\}^{1/2} \quad (8)$$

The image spectrum rejection magnitude is the ratio of the passband spectrum, amplitude relative to the image spectrum amplitude, and expressed in dB becomes,

$$\text{I.R.} = 10 \log \frac{2(1+\delta)(1+\cos\sigma)+\delta^2}{2(1+\delta)(1-\cos\sigma)+\delta^2} \qquad (9)$$

where δ is the amplitude difference between the two video-band signals prior to the combining; $\delta = V'_{BO} - V'_{TO}$; and σ is the net phase error from 90°.

The design error budget for the analog portion of the converter was set to 90° ±4° and ±0.4 dB in quadrature phase and amplitude imbalance, respectively, for the two video-output signals over any 600 kHz segment bandwidth within the S- and X-band IF range.

These error contributions would insure an image rejection to a value

$$\text{I.R. (analog)} = 10 \log \frac{2(1.05)(1+0.99756)+0.05^2}{2(1.05)(1-0.99756)+0.05^2}$$

or

$$\text{I. R.} = 27.4 \text{ dB}$$

For the digital portion, a value of 90° ±0.75° was specified. The ±0.75° error includes the phase ripple parameter across the passband range which is largely contributed to by the all-pass 90° phase-shift network.

The error budget of $\leqslant 1°$ peak-phase variation (ripple) is important when the spacecraft line-spectra signal is to be received during a ΔDOR observation. With a maximum sideband tone separation of 38 MHz, the contribution of 1° peak-phase error could result in an error of 1 cm in the signal group delay determination (Ref. 1). The phase error is manifest in the determination of the line-spectra sideband signal phase when it lies in between calibration tones during the reception-data-acquisition process.

The calibration tones are injected as a comb (equally spaced in frequency) of phase-stable tones into the VLBI system microwave input to aid in calibrating out the system-incurred phase variations during the final data processing. The phase of a fixed-frequency spacecraft signal falling in between the calibration tones would be difficult to determine should a large phase ripple exist. The error specification placed a limit on the resultant error contributed by the ripple between any two phase calibration tones. As will be discussed later in the test data, the measured ripple was better than specified and resulted in a linear interpolation factor.

Other contributors to the amplitude SNR degradation are the aliasing, quantization, and noise spectral density. The ADC largely contributed the noise spectral density and quantization noise, which, with the need for $\geqslant 48$ dB signal amplitude dynamic range, imposed the requirement for a quality ADC.

A 15-bit, state-of-the-art, ADC manufactured by Preston Scientific (Model GMAD-1A) is used and contains an equivalent noise spectral density (noise floor) of 0.75 millivolts rms.

The ADC has a ±5-V peak-volts signal input capability. Relative to this the 2^8-bit level minimum resolvable signal voltage of 19.5 mV results in a 48 dB dynamic capability.

The noise floor value, relative to 19.5 mV, results in a value 28 dB below this minimum signal and meets this error specification.

For quantization noise the equation is

$$Q_{\text{Noise}} = \frac{Q}{2\sqrt{3}} = \frac{\text{Full Scale Range}}{2^n\, 2/\sqrt{3}}$$

where Q is the quantum (or quantization size) and $2/\sqrt{3}$ is the conversion of the sampling residual triangular waveform to rms volts. Thus, the quantization voltage relative to the 5 V full-scale range is

$$V_{\text{Quant}} = \frac{5}{2^{15} \times 1.15}$$

$$= 0.13 \text{ mV rms}$$

and relative to the minimum resolvable signal level of 19.5 mV, the quantization noise is,

$$Q_{\text{Noise}} = 20 \log \frac{19.5}{0.1327}$$

$$= 43 \text{ dB}$$

and meets this error budget.

Aliasing noise is reduced by the shape characteristics of the anti-aliasing filter following the analog amplifiers relative to the sampling frequency of the ADC. Figure 6 illustrates the filter response, sampling frequency, and resultant aliasing noise response, and shows that the lower sampled sideband and data passband overlapping regions are > 73 dB below the passband spectrum. Thus aliasing noise meets the specified budget value of $\geqslant$ 25 dB even when the signal is 48 dB below the 5-V maximum ADC input.

IV. Test Results

The initial test results from the prototype unit met all design specifications. Individual tests on the analog portion were performed first because of its development schedule and were later combined with the digital portion as the latter was designed and constructed for the overall tests.

Table 2 tabulates the amplitude and quadrature phase tracking of the analog IF-VF mixer over the frequency range from 265 to 400 MHz. Measurements were at frequencies of 265, 300, 325, 350, 375, and 400 MHz where the input signal was stepped across a 10 to 300 kHz band at each of these frequencies.

Figures 7 and 8, for the 250 kHz and 125 kHz LPF bandwidth, are the test results of the complete IF-video converter amplitude response over a ±300 kHz spectrum about a local oscillator frequency of 300 MHz.

These two responses are typical for all the low-pass filter (LPF) bandwidths for any segment within the total IF spectrum. The results indicate that the analog and digital circuits tend to compensate the imbalance across the bandwidth and result in an image rejection greater than 34 dB.

The input signal for these results was set to the maximum value of the ADC at ±5 V. When the signal levels were lowered to input values of ±0.5 V and ±0.05 V into the ADC, the amplitude response had similar characteristics, indicating that the image rejection and passband characteristics would be perserved with a minimum signal 48 dB below the maximum signal capability. This range meets the dynamic range specification.

Figure 9 compares the expected phase shift versus frequency (calculated from the filter design) against the measured phase of the prototype unit and indicates close agreement.

Figure 10 is the design phase-ripple response versus frequency of the all-pass 90° digital phase shifter.

To determine the amount of phase ripple present in the total IF-video converter response shown in Fig. 9, a smoothing program was used to obtain an average phase versus frequency response for the measured data. The difference between the actual data and the average response was derived and plotted as shown in Fig. 11 for the 250 kHz LPF. This figure illustrates the nature of the ripple and indicates that the difference is less than a ±1° peak between 40 kHz to 205 kHz and within these frequency limits meets the specification relative to an average response. The nearly symmetrical variation of the phase ripple about the average response would imply that the phase response could be modeled between any 50 kHz segment band. This would aid in the operational utilization since some interpolation factor between any 50 kHz calibration tone could be adopted.

Figures 12, 13, 14, and 15 are such comparisons between the measured value and a linear regression analysis over a 50 kHz segment beginning from 10 kHz to 60 kHz and so on. For example, Fig. 13 is for the frequency segments from 60 to 110 kHz and 80 to 130 kHz and shows the linear curve comparison versus the data points. The correlation coefficient is 1.000, and resultant residuals are averaging below ±1°.

The remaining plots for other segment frequency ranges indicate low residual values except for the two end segments, i.e., 10 kHz to 80 kHz and 160 kHz to 230 kHz.

These results are acceptable since the tuning resolution of the digitally tuned frequency synthesizers (IF-video converter local oscillators) can place the spacecraft signal near the center of the LPF passband. Similarly, since the residuals are well within the 1° peak limit near the passband center region, the phase value between the 50 kHz calibration tones can be linearly interpolated to limit expected errors to well within the total overall system error allocation. There is no reason that the linear phase range could not be extended should such a requirement be necessary for other applications.

V. Summary

The requirements for good image-band rejection and low-phase variation IF-video converters was dictated by spacecraft navigation application. These were necessary because of the low-flux densities of available extragalactic radio sources that will be viewed near the spacecraft trajectory and the nature of accurately determining the phase information of the spacecraft line-spectra type signal having low-span bandwidth.

Assemblies of the DSCC BLK I VLBI system have been developed and tested which met these stringent requirements using a hybrid of analog and digital circuits. The performance has exceeded the characteristics of existing analog converters presently being used for similar VLBI applications.

The use of this converter design is not restricted to VLBI uses. It has application wherever a radio frequency spectrum must be down-converted to video-band frequencies for detection, data recording, or transmission over moderate bandwidth transmission circuits with tuning flexibility covering a broad spectrum.

Typical of such application is the observation and analysis of spacecraft carrier signals during planet occultation, orbiting satellite signals, and signal detection over an extremely broad sprectrum such as the Search for Extra Terrestrial Intelligence (SETI).

Good amplitude and quadrature phase matching over a wide range of IF were achieved with analog circuits while low phase variation all-pass phase shifters and realizable low-pass filters were accomplished using digital circuits.

The digital circuits following ADC will always cover a common range of video-band frequencies, regardless of the RF and IF range. This simplifies the circuit design to few functional types and lends itself to Very Large Scale Integration (VLSI) modules.

Acknowledgment

I wish to acknowledge the contributions of V. M. Chavez, V. S. Chen, and T. Sato for the design, development, and testing of the IF video converter.

References

1. Ham, N. C., "Frequency Down-Converter as Applied to VLBI," *TDA Progress Report 42-53*, pp. 74-82, Jet Propulsion Laboratory, Pasadena, California, July 1979.

2. SOSA, E. N., et. al., "Narrow Channel Bandwidth Receiver for VLBI," *TDA Progress Report 42-64*, pp. 18-29, Jet Propulsion Laboratory, Pasadena, California, May 1981.

3. Molinder, J. I., "A Tutorial Introduction to Very Long Baseline Interferometry (VLBI) Using Bandwidth Synthesis," *TDA Progress Report 42-46*, pp. 16-28, Jet Propulsion Laboratory, Pasadena, California, May 1978.

4. Gray, A. H., Jr., Markel, J. D., "A Computer Program for Designing Digital Elliptic Filters," *IEEE Transaction on Acoustics, Speech and Signal Processing*, Vol. ASSP-24, No. 6, pp. 529-538, December 6, 1976.

Table 1. Design specifications

RF-IF Spectra	
X-band	8400–8500 MHz
S-band	2265–2305 MHz
IF_x	300–400 MHz
IF_s	265–305 MHz
Output Channels	
Number of channels (Time and Frequency Multiplexed)	8 for X-band 4 for S-band
Bandwidth Selection, kHz	250 125 62.5 31.25
Data Format	1-bit quantization
Signal Phase Continuity (During the time-frequency multiplexing)	Phase continuity of each channel signal must be maintained over the length of the observation period.
Signal Reception Amplitude Dynamic Range	$\geq$ 48 dB, ratio of maximum to minimum signal resolved
Operational Control	Automatic control of all functions during operation
Functions:	Frequency tuning Channel multiplexing RF spectra selection Channel bandwidth
Signal Amplitude SNR Degradation	$\leq$ 1%
Error Budget Allocation	
Image band rejection	$\geq$ 23 dB below passband level
Aliasing noise	$\geq$ 25 dB below minimum signal resolution
Quantization noise	$\geq$ 25 dB below minimum signal resolution
Signal Phase SNR Degradation	$<$ 1%
Error Budget Allocation:	
Dispersive phase error (random)	$\sigma_\phi < 1^\circ$ rms
Phase variation (nonrandom)	σ_ϕ 1° peak

Table 2. Test results of typical analog portion of IF video converter

Oscillation Frequency	265 MHz		300 MHz		325 MHz		350 MHz		375 MHz		400 MHz	
Output Frequency kHz	ΔA, dB	Δφ, deg	ΔA, dB	Δφ, deg	ΔA, dB	Δφ, deg	ΔA, dB	Δφ, deg	ΔA, dB	Δφ, deg	ΔA, dB	Δφ, deg
10	0.5	0.1	0.5	−0.7	0.5	−0.8	0.6	0	0.6	−0.2	0.6	0
25	0.4	0.1	0.5	−0.6	0.5	−0.8	0.6	0	0.6	−0.1	0.7	0.1
50	0.4	0.2	0.5	−0.6	0.5	−0.8	0.5	0	0.6	−0.1	0.7	0.1
75	0.5	0.3	0.5	−0.6	0.5	−0.7	0.5	0.1	0.6	−0.1	0.7	0.1
100	0.5	0.3	0.5	−0.5	0.5	−0.7	0.5	0.1	0.6	−0.1	0.7	0.1
125	0.4	0.3	0.5	−0.5	0.5	−0.7	0.5	0.1	0.6	0	0.6	0.1
150	0.4	0.3	0.5	−0.5	0.5	−0.7	0.5	0.1	0.6	0	0.8	0.2
175	0.4	0.3	0.5	−0.5	0.4	−0.7	0.4	0.1	0.6	0	0.8	0.2
200	0.3	0.3	0.5	−0.5	0.4	−0.7	0.4	0.1	0.6	0	0.8	0.1
225	0.3	0.3	0.5	−0.6	0.4	−0.7	0.4	0.1	0.6	−0.1	0.8	0.1
250	0.3	0.2	0.4	−0.6	0.5	−0.8	0.5	0	0.5	−0.1	0.7	0.1
275	0.3	0.1	0.4	−0.7	0.5	−0.9	0.5	−0.1	0.6	−0.2	0.6	0
300	0.3	0.1	0.5	−0.7	0.4	−0.9	0.5	−0.1	0.6	−0.2	0.7	−0.1

Output frequency = $F_{INPUT} - F_{OSCILLATOR}$
ΔA = Difference between two output video frequencies
Δφ = Difference from quadrature between output frequencies

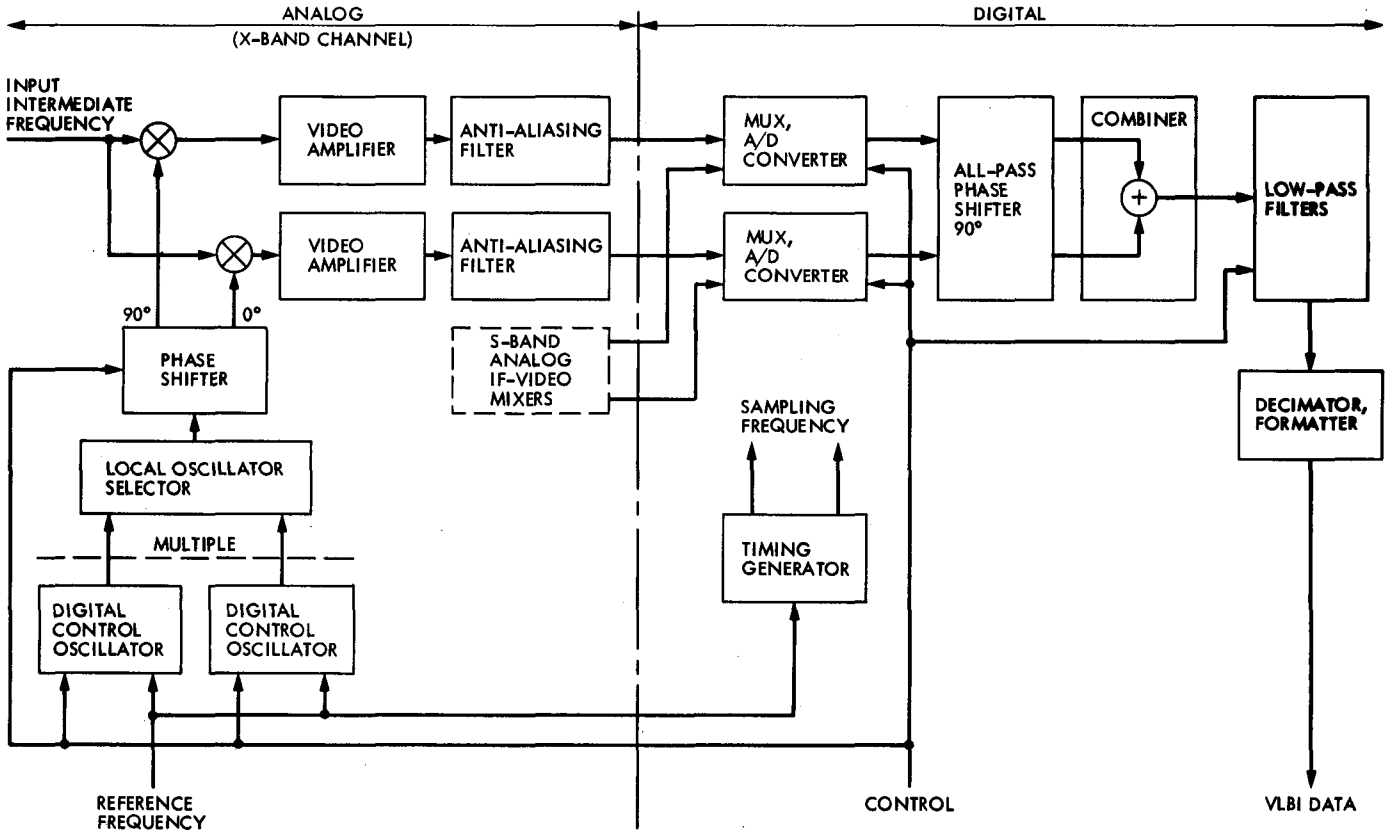


Fig. 1. IF-video converter

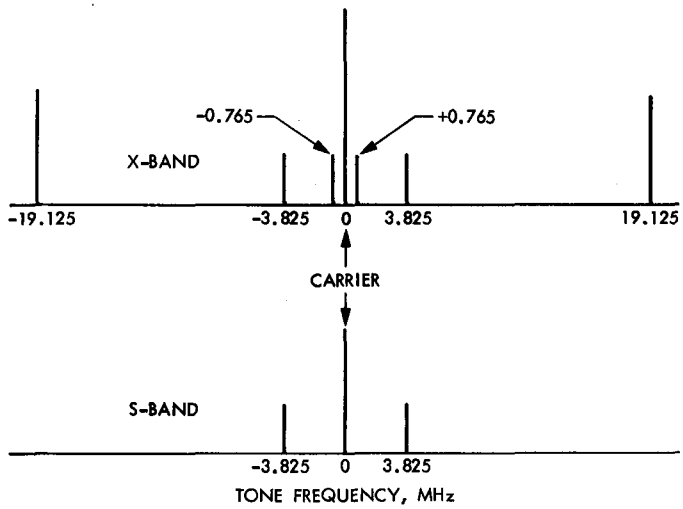


Fig. 2. Frequencies of the Galileo tones

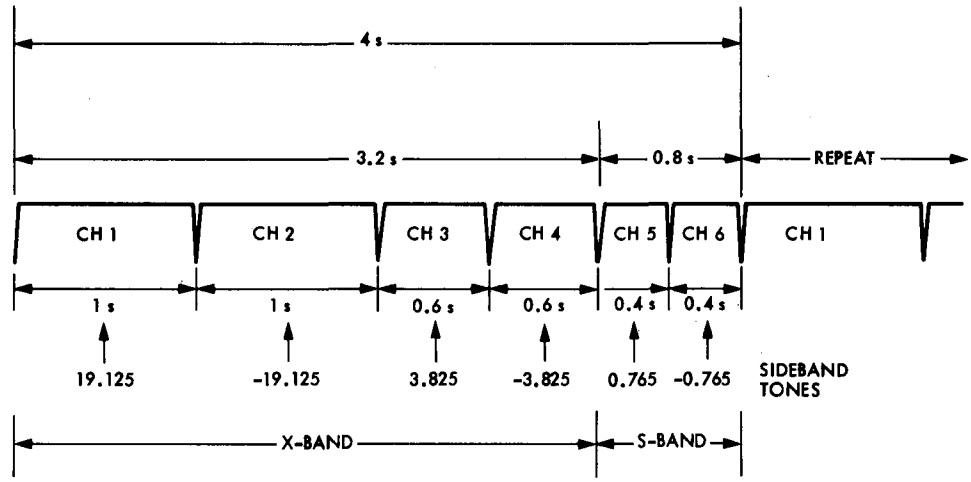


Fig. 3. Typical multiplexed channel sequence of spacecraft sideband tones

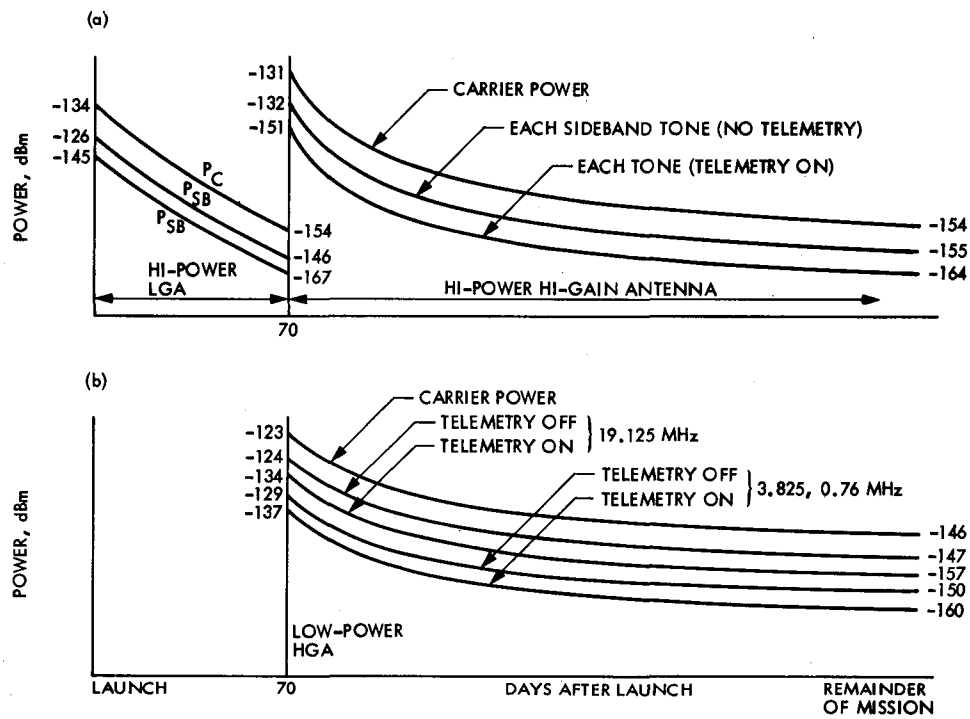


Fig. 4. Received spacecraft power level vs time profile: (a) S-band; (b) X-band

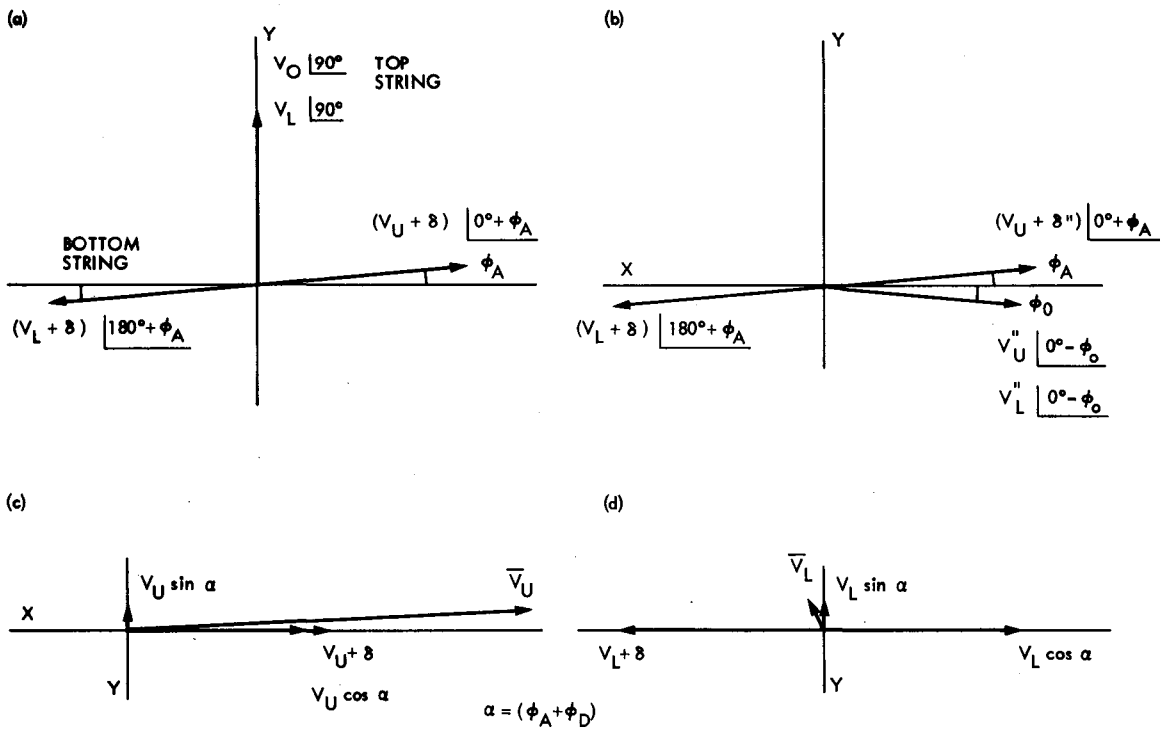


Fig. 5. Vector diagrams for upper- and lower-band spectrum: (a) video amplifiers output; (b) digital phase shifter output; (c) combiner summed output for desired passband upper-band spectrum; (d) combiner summed output for rejected image lower-band spectrum

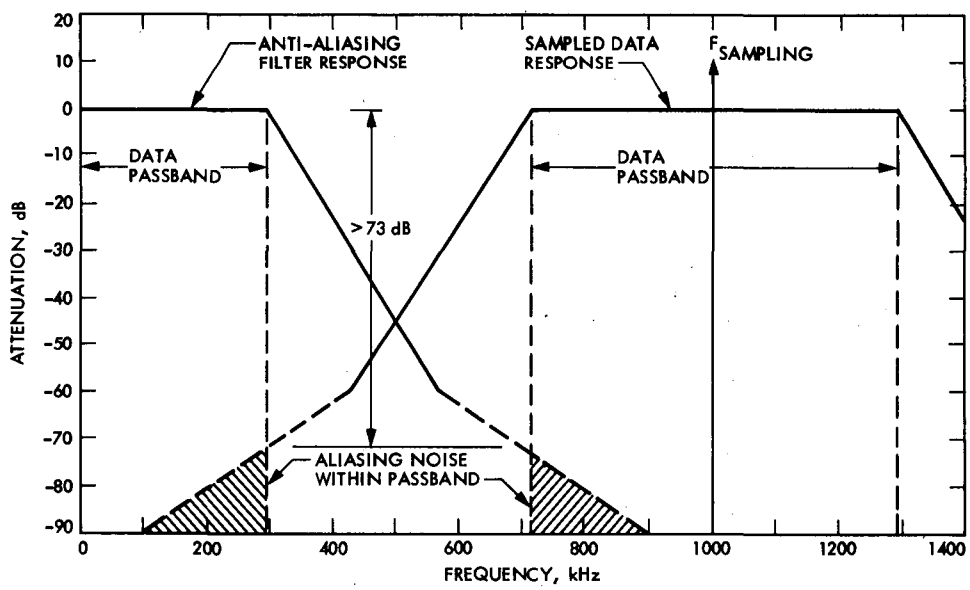


Fig. 6. Down converted video and sampled data spectra

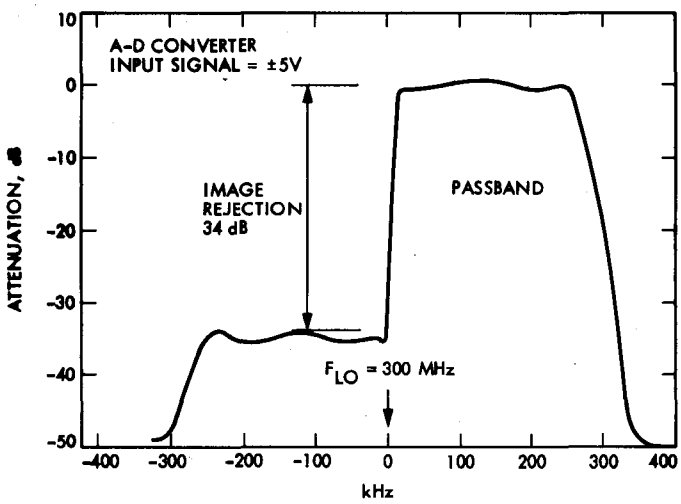


Fig. 7. Total response of IF-video converter for 250 kHz bandwidth

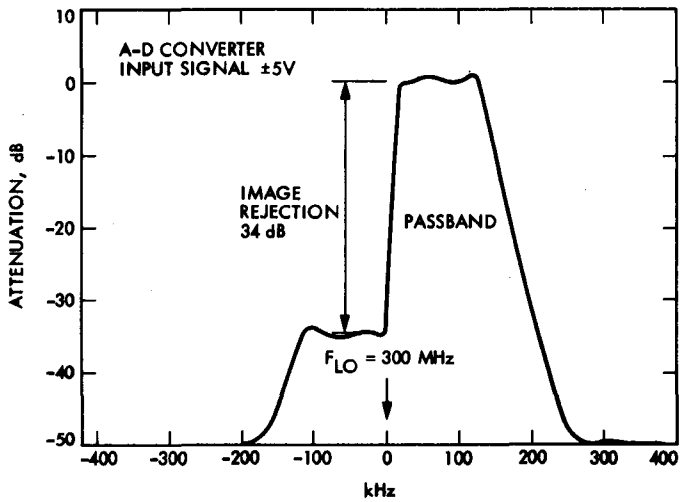


Fig. 8. Total response of IF-video converter for 125-kHz bandwidth

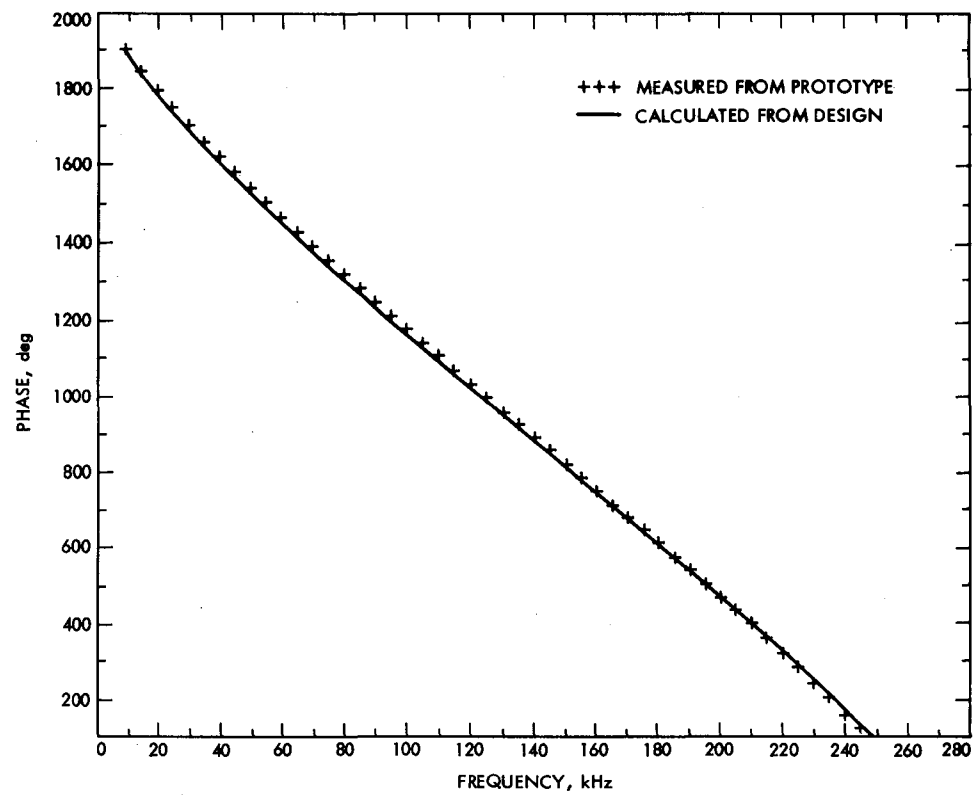


Fig. 9. Data channel filter, phase response (250 kHz BW)

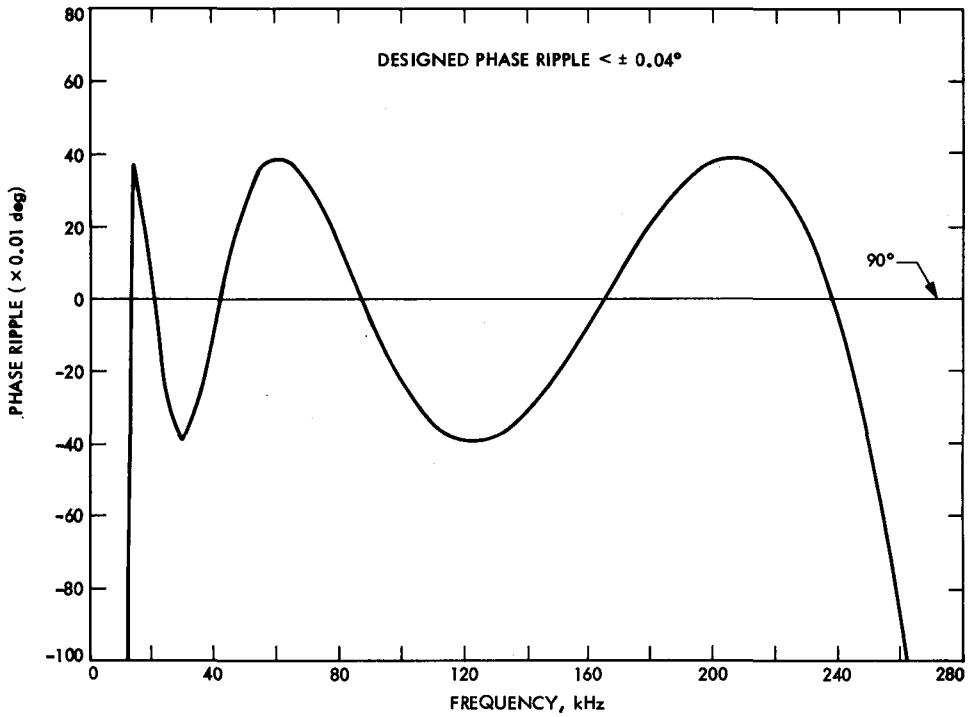


Fig. 10. Data channel filter, 90° phase shifter

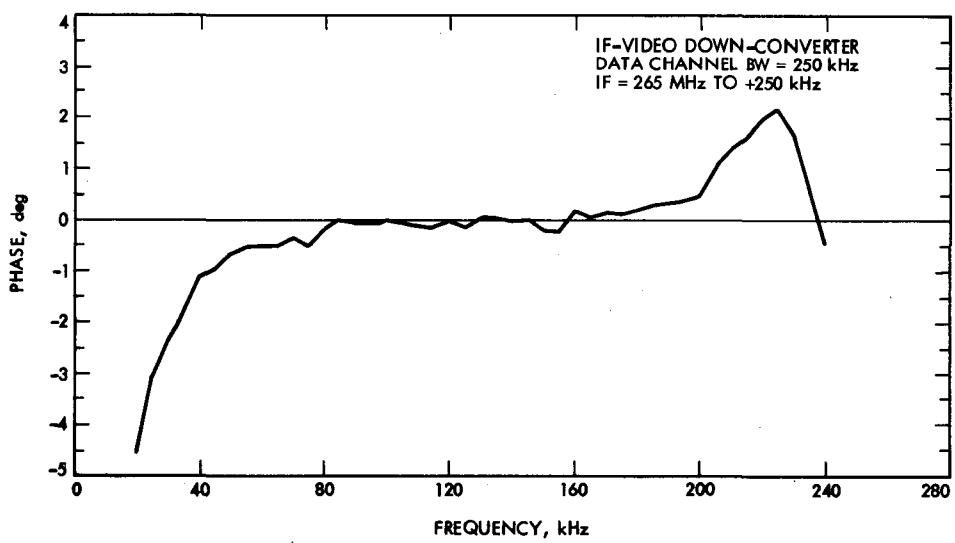


Fig. 11. Phase difference between actual and smoothed average phase response

R SQUARE = 0.997

YHAT = 2348.528 + -10.284 X

X(I)	Y(I)	YHAT	RESIDUALS
10.00	2265.70	2245.69	20.01
15.00	2198.70	2194.27	4.43
20.00	2138.40	2142.84	-4.44
25.00	2082.50	2091.42	-8.92
30.00	2029.50	2040.00	-10.50
35.00	1978.80	1988.58	-9.78
40.00	1930.10	1937.16	-7.06
45.00	1882.20	1885.74	-3.54
50.00	1835.50	1834.32	1.18
55.00	1789.30	1782.90	6.40
60.00	1743.70	1731.48	12.22

R SQUARE = 1.000

YHAT = 2300.653 + -9.240 X

X(I)	Y(I)	YHAT	RESIDUALS
30.00	2029.50	2023.46	6.04
35.00	1978.80	1977.27	1.53
40.00	1930.10	1931.07	-0.97
45.00	1882.20	1884.87	-2.67
50.00	1835.50	1838.67	-3.17
55.00	1789.30	1792.47	-3.17
60.00	1743.70	1746.27	-2.57
65.00	1698.50	1700.08	-1.58
70.00	1654.00	1653.88	0.12
75.00	1609.60	1607.68	1.92
80.00	1566.00	1561.48	4.52

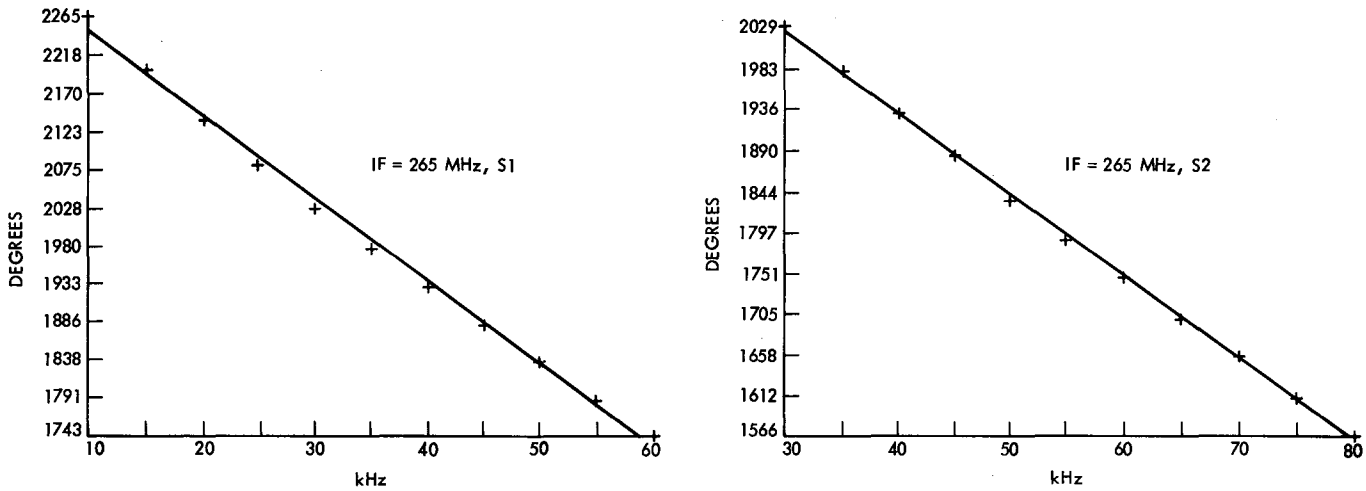


Fig. 12. Linear regression analysis results from 10 to 80 kHz

R SQUARE = 1.000

YHAT = 2266.979 + -8.751 X

X(I)	Y(I)	YHAT	RESIDUALS
60.00	1743.70	1741.91	1.79
65.00	1698.50	1698.16	0.34
70.00	1654.00	1654.40	-0.40
75.00	1609.60	1610.65	-1.05
80.00	1566.00	1566.89	-0.89
85.00	1522.50	1523.14	-0.64
90.00	1478.90	1479.38	-0.48
95.00	1435.40	1435.63	-0.23
100.00	1392.00	1391.87	0.13
105.00	1348.60	1348.11	0.49
110.00	1305.30	1304.36	0.94

R SQUARE = 1.000

YHAT = 2257.964 + -8.656 X

X(I)	Y(I)	YHAT	RESIDUALS
80.00	1566.00	1565.45	0.55
85.00	1522.50	1522.17	0.33
90.00	1478.90	1478.89	0.01
95.00	1435.40	1435.61	-0.21
100.00	1392.00	1392.33	-0.33
105.00	1348.60	1349.05	-0.45
110.00	1305.30	1305.76	-0.46
115.00	1262.10	1262.48	-0.38
120.00	1219.20	1219.20	-0.00
125.00	1176.10	1175.92	0.18
130.00	1133.40	1132.64	0.76

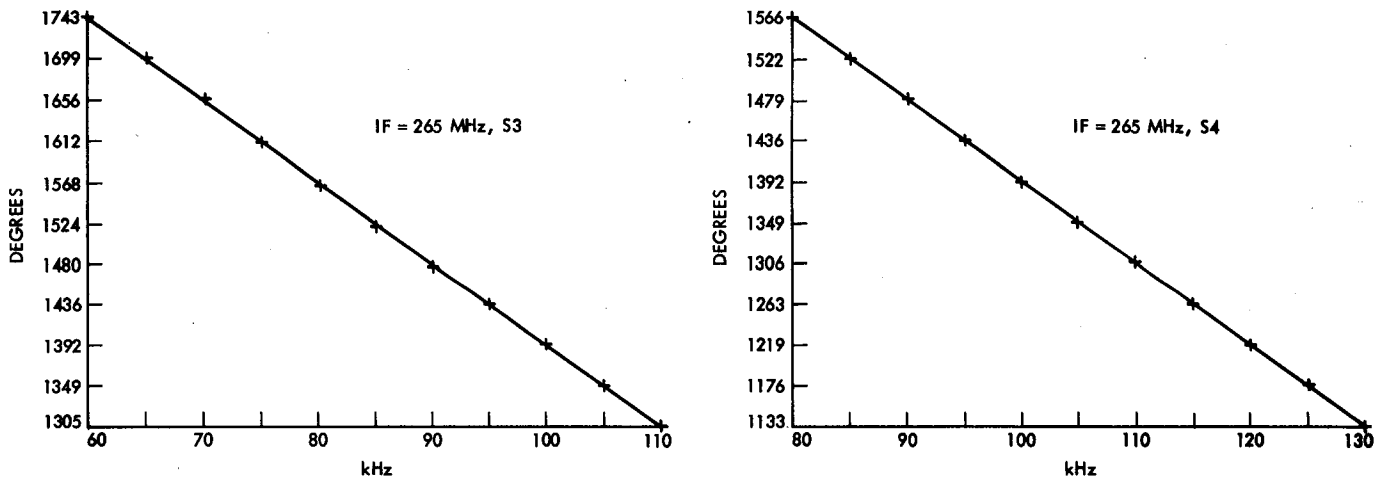


Fig. 13. Linear regression analysis results from 60 to 130 kHz

R SQUARE = 1.000

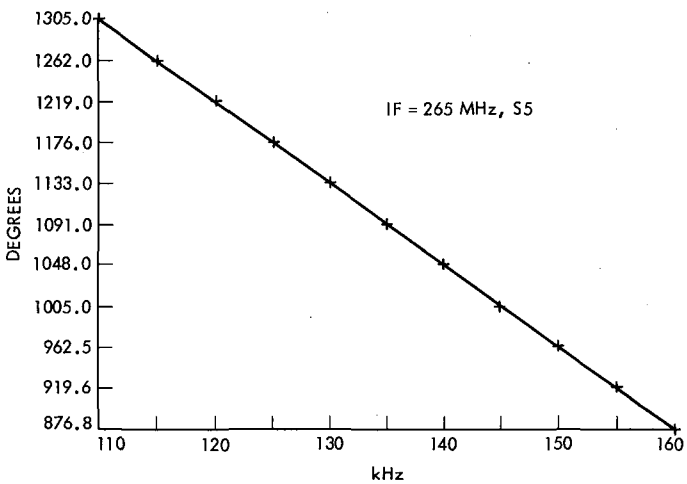
YHAT = 2248.161 + -8.575 X

X(I)	Y(I)	YHAT	RESIDUALS
110.00	1305.30	1304.94	0.36
115.00	1262.10	1262.07	0.03
120.00	1219.20	1219.19	0.01
125.00	1176.10	1176.32	-0.22
130.00	1133.40	1133.45	-0.05
135.00	1099.50	1090.57	-0.07
140.00	1047.50	1047.70	-0.20
145.00	1004.70	1004.83	-0.13
150.00	961.70	961.95	-0.25
155.00	919.00	919.08	-0.08
160.00	876.80	876.20	0.60

R SQUARE = 1.000

YHAT = 2243.333 + -8.541 X

X(I)	Y(I)	YHAT	RESIDUALS
130.00	1133.40	1132.99	0.41
135.00	1090.50	1090.29	0.21
140.00	1047.50	1047.58	-0.08
145.00	1004.70	1004.87	-0.17
150.00	961.70	962.17	-0.47
155.00	919.00	919.46	-0.46
160.00	876.80	876.76	0.04
165.00	834.10	834.05	0.05
170.00	791.60	791.35	0.25
175.00	748.80	748.64	0.16
180.00	706.00	705.94	0.06



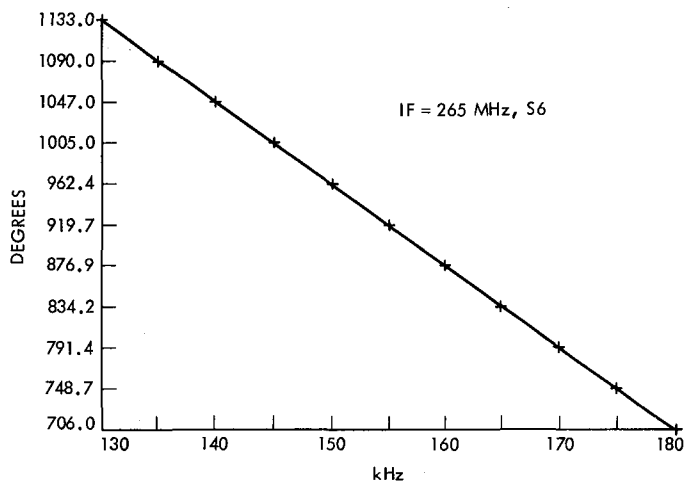


Fig. 14. Linear regression analysis results from 110 to 180 kHz

R SQUARE = 1.000

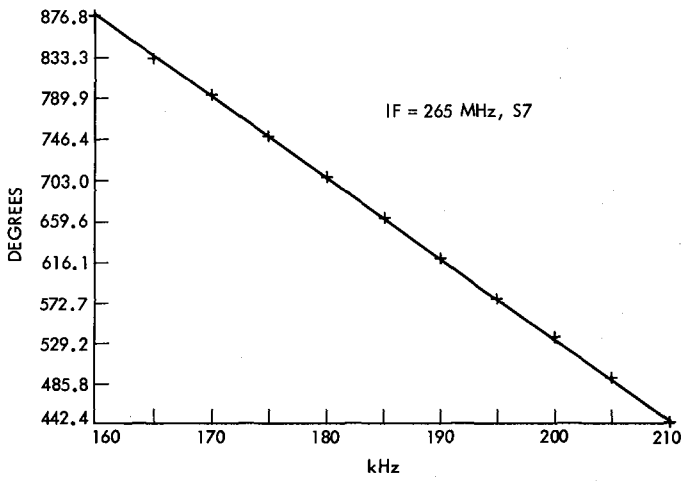
YHAT = 2265.253 + -8.668 X

X(I)	Y(I)	YHAT	RESIDUALS
160.00	876.80	878.37	-1.57
165.00	834.10	835.03	-0.93
170.00	791.60	791.69	-0.09
175.00	748.80	748.35	0.45
180.00	706.00	705.01	0.99
185.00	663.00	661.67	1.33
190.00	619.70	618.33	1.37
195.00	576.10	574.99	1.11
200.00	532.10	531.65	0.45
205.00	487.80	488.31	-0.51
210.00	442.40	444.97	-2.57

R SQUARE = 0.999

YHAT = 2361.363 + -9.163 X

X(I)	Y(I)	YHAT	RESIDUALS
180.00	706.00	711.94	-5.94
185.00	663.00	666.12	-3.12
190.00	619.70	620.31	-0.61
195.00	576.10	574.49	1.61
200.00	532.10	528.67	3.43
205.00	487.80	482.85	4.95
210.00	442.40	437.04	5.36
215.00	395.60	391.22	4.38
220.00	347.30	345.40	1.90
225.00	297.00	299.59	-2.59
230.00	244.40	253.77	-9.37



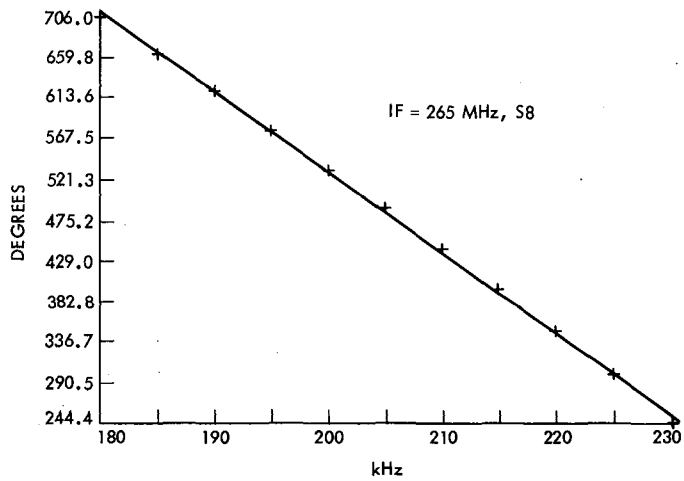


Fig. 15. Linear regression analysis results from 160 to 230 kHz

NASA—JPL—Coml., L.A., Calif.