# The Telecommunications and Data Acquisition Progress Report 42-123

## July–September 1995

Joseph H. Yuen
Editor



November 15, 1995

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

# The Telecommunications and Data Acquisition Progress Report 42-123

## July–September 1995

Joseph H. Yuen
Editor

November 15, 1995


National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

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

# Note From the Editor

Since issue 42-121, published on May 15, 1995, *The Telecommunications and Data Acquisition Progress Report* has been available electronically to all its readers on the World Wide Web at http://tda.jpl.nasa.gov/progress_report. Printed copies were also produced but, as the Editor's Note in that issue stated, the ultimate goal was to publish solely in electronic form. Now that goal is close to being realized. Beginning with issue 42-125, due to be published on May 15, 1996, *The TDA Progress Report* will be published entirely electronically at the above-mentioned URL. This issue and the next, 42-124, will be the last issues for which printed copies are produced; as a convenience for readers who are currently on our distribution list, however, we will distribute a hard copy of the table of contents for each issue. Readers with questions or concerns regarding this change are welcome to contact the editor.

# Preface

This quarterly publication provides archival reports on developments in programs managed by JPL's Telecommunications and Mission Operations Directorate (TMOD), which now includes the former Telecommunications and Data Acquisition (TDA) Office. In space communications, radio navigation, radio science, and ground-based radio and radar astronomy, it reports on activities of the Deep Space Network (DSN) in planning, supporting research and technology, implementation, and operations. Also included are standards activity at JPL for space data and information systems and reimbursable DSN work performed for other space agencies through NASA. The preceding work is all performed for NASA's Office of Space Communications (OSC).

TMOD also performs work funded by other NASA program offices through and with the cooperation of OSC. The first of these is the Orbital Debris Radar Program funded by the Office of Space Systems Development. It exists at Goldstone only and makes use of the planetary radar capability when the antennas are configured as science instruments making direct observations of the planets, their satellites, and asteroids of our solar system. The Office of Space Sciences funds the data reduction and science analyses of data obtained by the Goldstone Solar System Radar. The antennas at all three complexes are also configured for radio astronomy research and, as such, conduct experiments funded by the National Science Foundation in the U.S. and other agencies at the overseas complexes. These experiments are either in microwave spectroscopy or very long baseline interferometry.

Finally, tasks funded under the JPL Director's Discretionary Fund and the Caltech President's Fund that involve TMOD are included.

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

# Contents

## OSC TASKS
## DSN Advanced Systems
### TRACKING AND GROUND-BASED NAVIGATION

**Sensitivity of Planetary Cruise Navigation to Earth Orientation Calibration Errors** ..... 1 -1
J. A. Estefan and W. M. Folkner
NASA Code 312-30-11-60-02

**Wind Gust Models Derived From Field Data** ..... 30 -2
W. Gawronski
NASA Code 314-30-11-00-12

**A New Model for Yaw Attitude of Global Positioning System Satellites** ..... 37 -3
Y. E. Bar-Sever
NASA Code 314-30-11-50-02

**A Light-Induced Microwave Oscillator** ..... 47 -4
X. S. Yao and L. Maleki
NASA Code 314-30-11-40-04

### COMMUNICATIONS, SPACECRAFT–GROUND

**Optimum Detection of Tones Transmitted by a Spacecraft** ..... 69 -5
M. K. Simon, M. M. Shihabi, and T. Moon
NASA Code 314-30-11-30-02

**On the Design of Turbo Codes** ..... 99 -6
D. Divsalar and F. Pollara
NASA Code 315-91-20-20-53

**The Trellis Complexity of Convolutional Codes** ..... 122 -7
R. J. McEliece and W. Lin
NASA Code 315-91-20-20-53

**System Noise Temperature Investigation of the DSN S-Band Polarization Diverse Systems for the Galileo S-Band Contingency Mission** ..... 140 -8
J. E. Fernandez and D. L. Trowbridge
NASA Code 314-30-61-02-14

# Sensitivity of Planetary Cruise Navigation to Earth Orientation Calibration Errors

J. A. Estefan
Navigation and Flight Mechanics Section

W. M. Folkner
Tracking Systems and Applications Section

*A detailed analysis was conducted to determine the sensitivity of spacecraft navigation errors to the accuracy and timeliness of Earth orientation calibrations. Analyses based on simulated X-band (8.4-GHz) Doppler and ranging measurements acquired during the interplanetary cruise segment of the Mars Pathfinder heliocentric trajectory were completed for the nominal trajectory design and for an alternative trajectory with a longer transit time. Several error models were developed to characterize the effect of Earth orientation on navigational accuracy based on current and anticipated Deep Space Network calibration strategies. The navigational sensitivity of Mars Pathfinder to calibration errors in Earth orientation was computed for each candidate calibration strategy with the Earth orientation parameters included as estimated parameters in the navigation solution. In these cases, the calibration errors contributed 23 to 58 percent of the total navigation error budget, depending on the calibration strategy being assessed. Navigation sensitivity calculations were also performed for cases in which Earth orientation calibration errors were not adjusted in the navigation solution. In these cases, Earth orientation calibration errors contributed from 26 to as much as 227 percent of the total navigation error budget. The final analysis suggests that, not only is the method used to calibrate Earth orientation vitally important for precision navigation of Mars Pathfinder, but perhaps equally important is the method for inclusion of the calibration errors in the navigation solutions.*

## I. Introduction

Radio metric data, particularly two-way coherent Doppler and range, have been used to navigate robotic spacecraft since the inception of planetary exploration. For a spacecraft in interplanetary cruise or transit, much of the information content inherent in the data for position determination comes from the signature imposed on the station–spacecraft radio signal by the Earth's rotation [1–3]. The diurnal signature in the radio metric data yields information about the right ascension and declination of the spacecraft with respect to the direction of the Earth's spin axis at the time of observation. The orientation of the Earth, as a function of time, must be known with respect to inertial space in order to effectively utilize the radio metric data to deduce spacecraft position with respect to the target planet. Errors in Earth orientation thus lead to targeting errors for spacecraft approaching other planetary bodies.

Evidence of the need to adequately account for Earth orientation errors came as early as April 1965 when flight project navigation teams for the Rangers VII and VIII lunar probes observed a large difference in station longitude solutions for all deep-space stations using radio metric data [4]. This was later determined to be the result of improper Earth orientation calibration. As a result, Earth orientation calibration methods were later refined to support the Mariners IV and V planetary exploration missions.

To assess the effect of Earth orientation calibration errors on interplanetary cruise navigation for both current and future Deep Space Network (DSN) Earth orientation calibration techniques, a navigation error analysis of the Mars Pathfinder approach scenario was performed. Mars Pathfinder has the most stringent planetary cruise navigation requirements of any currently planned mission. Other Mars lander missions similar to Pathfinder are being studied. Navigation performance for these future missions may exhibit different sensitivity characteristics to Earth orientation calibration errors since the sensitivity is trajectory dependent. In this study, two Mars approach trajectories were evaluated, the nominal Pathfinder cruise trajectory with arrival at Mars on July 4, 1997, and a second trajectory with a longer transit time. Clearly, restricting the study to only two trajectories is far from encompassing the entire range of possible planetary approach scenarios. Moreover, actual navigation performance will vary depending on targeting point and targeting requirements, the data type and arc length, filtering strategy, and observation geometry, which could vary depending on each launch opportunity (especially spacecraft right ascension and declination at encounter). The study of two "representative" trajectories, while limited, at least provides some insight into the possible range of navigational uncertainties caused by Earth orientation calibration errors.

Other types of navigation problems have varying sensitivity to Earth orientation error. For spacecraft in close orbit about another planetary body, such as Magellan or Mars Global Surveyor (MGS), the primary signature on the spacecraft radio signal is imposed by the orbit about the planet. Doppler measurements can be used to determine all spacecraft orbital elements in most cases, and the resultant orbit determination is largely insensitive to Earth orientation errors. If, however, the orbit determination using Doppler data is not accurate enough to meet the mission requirements, two-station differenced-Doppler or narrow-band very long baseline interferometry (VLBI) observations can be used for improved orbit determination in some cases. In fact, the Magellan project utilized differenced-Doppler data for this purpose. A detailed sensitivity analysis of differenced-Doppler navigation to Earth orientation calibration errors is not presented in this article, but a cursory approximation is given in Appendix A. In contrast to spacecraft in close planetary orbit, the Galileo and Cassini spacecraft will be in long-period ($\sim$120-day) orbits about Jupiter and Saturn, respectively. These represent an intermediate case between planetary approach and low planetary orbit, so some sensitivity to Earth orientation errors might be expected. Onboard optical images of planetary satellites will be an important data type in determining the orbits for Galileo and Cassini. The added complexity of blending onboard optical data with radio metric data precluded this study from assessing the navigation sensitivity to Earth orientation errors for these outer planet orbiters.

In this article, a navigation error analysis is described that was used to assess the impact of various Earth orientation calibration strategies on predicted spacecraft orbit determination accuracies during interplanetary cruise. Section II provides the fundamental framework for defining the principal parameters that are used to characterize Earth orientation, while Section III focuses on the Earth orientation calibration process used by the DSN. These discussions are followed by a description in Section IV of the origin and format of the functional requirements levied on the DSN tracking system by the flight projects. In Section V, a simple information content analysis is presented to obtain a rough estimate of the influence of Earth orientation errors on Doppler cruise navigation performance. Section VI describes the assumptions used in a linear covariance analysis to evaluate the sensitivity of spacecraft navigational accuracies to Earth orientation calibration errors for two Mars Pathfinder approach scenarios. Various Earth orientation calibration strategies are described, together with tracking data simulation and error modeling assumptions. Results and key observations from the numerical assessment are summarized and discussed at the conclusion of the article.

## II. Earth Orientation Parameters

The Earth is an oblate, spinning body that undergoes precession and nutation due to the torques exerted upon it by the Sun, Moon, and other planets. The north pole of a body-fixed (crust-fixed) coordinate system varies unpredictably with respect to the spin direction, due to internal dynamics of the Earth and its atmosphere (a process called "polar motion"). Similar effects cause the Earth's rotation rate to vary unpredictably. (The variations in the rotation rate are several times larger than the polar motion variations.)

The orientation of a body in inertial space can be completely described by three Euler angles. Because the Earth rotates rapidly, the three angles describing the orientation of the surface with respect to inertial space vary rapidly with time. Conventionally, the orientation of the Earth is described by five angles that vary slowly with time, rather than by three rapidly varying angles. These five angles are described in greater detail below.

In the development of the 1980 International Astronomical Union (IAU) theory of nutation [5], the concept of the celestial ephemeris pole (CEP) was introduced. The CEP was defined such that there are no nearly diurnal motions of the CEP with respect to either space-fixed (inertial) or body-fixed coordinates. For a rigid body with no polar motion, the CEP corresponds to the body axis about which the body is spinning.

The motion of the CEP in space-fixed coordinates, due to precession and nutation, can be described by the two angles, $\psi$ and $\varepsilon$, where $\varepsilon$ is the obliquity (inclination of the equatorial plane to the Earth's orbital plane) and $\psi$ is the intersection of the equator and orbit with respect to a fixed equinox. The variation in the Earth's rotation about the CEP affects the time at which celestial objects cross the apparent meridian and is measured by a quantity called Universal Time (UT) (specifically, Universal Time 1, or "UT1"). Variations of the CEP in body-fixed coordinates are measured by the quantities polar motion "X" and polar motion "Y".

Because of the random variation of UT1 and polar motion (along with imperfect modeling of precession and nutation), an accurate description of Earth orientation requires continual monitoring. VLBI data can be used to determine all components of Earth orientation with 5 nrad or better accuracy (1-sigma).[1] Because VLBI measurements require correlation of large volumes of data from ground stations separated by large distances, there is usually a time delay between the acquisition of raw VLBI data and the processing of the data that determines the Earth orientation angles. This processing delay is currently 2 to 3 days for DSN VLBI measurements made for rapid determination of Earth orientation (i.e., TEMPO measurements, described in Section III); the delay is longer for VLBI data from external services. Satellite laser ranging (SLR) or Global Positioning System (GPS) data can also be used to determine polar motion and small changes in UT1 with shorter data-processing times but are not able to directly measure all five Earth orientation angles. Atmospheric angular momentum (AAM) data are highly correlated with variation in UT1 and the length of the day (LOD), a parameter proportional to the rate of change of UT1. Therefore, AAM data, both measurements and forecasts, have been used to improve predictions for both UT1 and LOD [6].

Precession/nutation models with parameters adjusted to fit the observed space-fixed motion of the CEP, e.g.,[7,8], have an accuracy of 5 nrad or better over the time of the fit. These models can be used to predict precession and nutation for periods of about 1 year before discrepancies systematically exceed 5 nrad. Figure 1 shows a comparison of the daily correction calibrations (from 1991 to 1995) with a model by Steppe et al. fit to data through the end of 1993 [9]. The nutation corrections are with respect

[1] Earth orientation accuracies are often quoted in a variety of units. An angle of 5 nrad is approximately equal to 1 milliarcsecond (mas). An angular rotation of 5 nrad corresponds to a change in position on the surface of the Earth (equatorial displacement) of about 3 cm. A change in UT1 of 1 millisecond (ms) corresponds to an angle of about 15 mas, which is equivalent to an angle of 75 nrad, or to an equatorial displacement of roughly 50 cm.

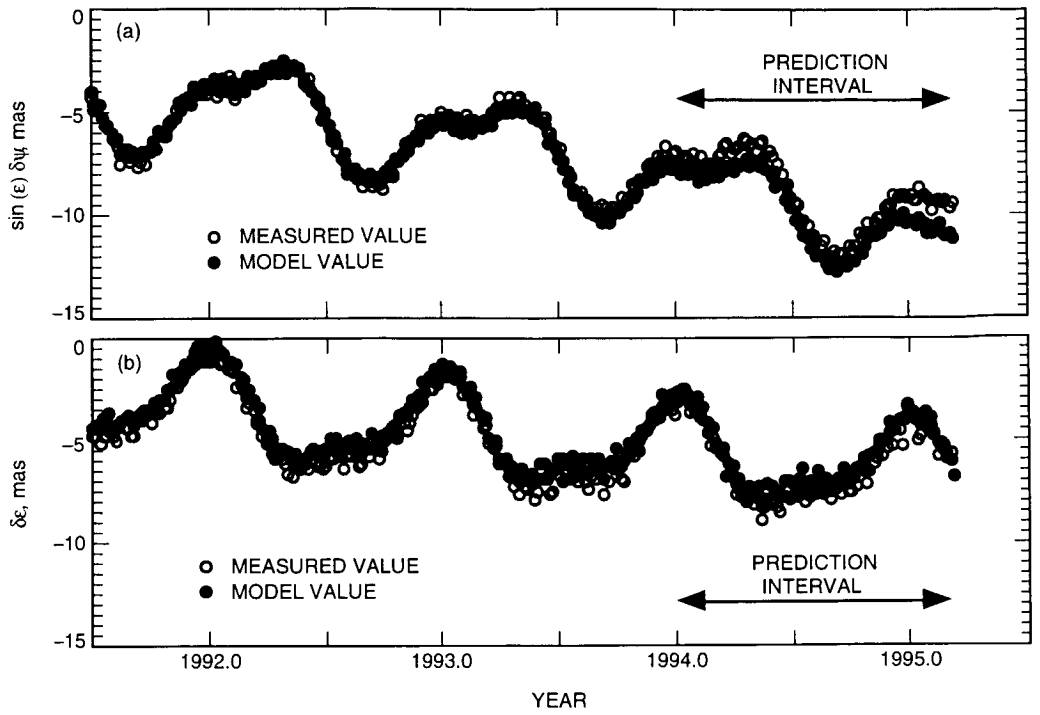


**Fig. 1. A comparison of a nutation correction model to observations. The correction angles (a) sin(ε) δψ and (b) δε are corrections to the IAU (1976) nutation model [5]. (A change of δε or sin(ε) δψ of 1 mas corresponds to a shift in the inertial position of a point on the Earth's surface of about 3 cm.)**

to the 1976 IAU precession model [10,11] and 1980 IAU nutation model [5]. The corrections are currently about 10 mas (~50 nrad) and are increasing with time. It can be seen that the predictions of the model fit to data through the end of 1993 agree with the later measurements to an accuracy of about 1 mas for about 1 year.

UT1 and polar motion (collectively referred to as UTPM throughout this article) vary randomly due to the interaction of the atmosphere and the crust. UT1 varies much more rapidly than polar motion. Random variation in UT1 can be characterized by an integrated random walk, while polar motion behaves approximately as an integrated Gauss–Markov process [12]. UT1 varies by an amount corresponding to an angle of 1 mas in about 1 day, so continual, rapid calibration is required to be able to completely describe Earth orientation to 1-mas accuracy.

## III. Earth Orientation Calibrations

At the Jet Propulsion Laboratory (JPL), Earth orientation calibrations are currently determined by the DSN's Time and Earth Motion Precision Observations (TEMPO) activity. TEMPO, which became operational in late 1983, was chartered to provide an operational Earth orientation service both to support JPL's spacecraft navigation efforts and to serve the worldwide community [13]. TEMPO supports Earth orientation calibration by performing VLBI measurements at regular intervals (currently twice per week) using the DSN's 70-m antenna subnetwork. (Prior to 1983, Earth orientation calibrations were provided by the DSN's Tracking System Analytic Calibration (TSAC) activity, which produced calibrations based on monthly estimates of UTPM disseminated by the Bureau Internationale de l'Heure (BIH) in Paris, France.)[2] The Kalman Earth Orientation Filter (KEOF) is used to combine the TEMPO measurements

[2] T. F. Runge, personal communication, Tracking Systems and Applications Section, Jet Propulsion Laboratory, Pasadena, California, February 1995.

with other sources of Earth orientation information. By performing regular VLBI measurements and including AAM measurements and forecasts, together with other data from Earth orientation services outside JPL, the DSN can deliver, at any time, an Earth orientation calibration accurate to 50 nrad (1-sigma) [6]. Earth orientation accuracy is better for times 15 days or more in the past, for which a greater amount of processed VLBI data from external services is available. Earth orientation predictions are also delivered, with accuracies that degrade with time due to the random behavior of UTPM. Efforts are ongoing to improve these predictions both by modeling improvements within the KEOF and by better utilization of geodetic and AAM data.[3]

The standard DSN Earth orientation calibration file (referred to as a UTPM STOIC file) is a text file of polynomial coefficients that provides UTPM calibrations for 37 specified times.[4] Precession and nutation calibrations are not included. The limitation to 37 calibration times has implications for the accuracy of Earth orientation available to the end-user (e.g., navigation teams) because of the integrated-random walk characteristic of UTPM. Several flight projects utilize calibration files that span a year or more, giving 10-day spacing (or more) between calibration times. Midway between respective calibrations at 10-day intervals, the expected (1-sigma) error in UT1 is about 0.4 ms (~20 cm) even if the calibration is perfect at the calibration times. This limitation, together with the lack of precession/nutation calibrations, has led to a new DSN calibration file—the Earth-Orientation Parameter (EOP) file—which includes precession and nutation corrections and has no limit on the number of calibration times.[5] It should be noted that all timing calibrations and their rates are given with respect to a reference time defined by atomic clocks, specifically, International Atomic Time (TAI).

## IV. Functional Requirements

Navigation-related requirements for current and future missions are defined primarily by flight projects and future mission study teams. These requirements serve as a starting point to establish DSN ground support requirements to satisfy mission navigation. In the past, navigation requirements for calibrations such as Earth orientation, station locations, and transmission media typically have been arrived at in an ad hoc manner without thorough analysis. This practice has at times resulted in confusion and later cancellation of implementation plans to develop calibrations for which there was an erroneously believed need.

Arguably, flight projects and future mission study teams find it more economical to simply adopt past calibration performance or to adopt anticipated improvements in the calibrations rather than conduct a parametric study in which all possible navigation calibrations are investigated. In order to meet mission navigation needs, the DSN has documented UTPM calibration capabilities and requirements for the tracking and navigation subsystems.[6,7] The current UTPM requirements are stated as: "(a) 30 cm (1-

[3] J. O. Dickey, personal communication, Tracking Systems and Applications Section, Jet Propulsion Laboratory, Pasadena, California, August 1995.

[4] In the early 1970s, all UTPM calibration data for mission operations were supplied in a single computer card deck called a PLATO deck (Platform Observables). The PLATO system replaced the former Timing and Polynomial (TPOLY) computer program for generating separate timing calibration data [14]. For contingency purposes, a smaller and simpler backup program was developed to generate PLATO-style decks that could be delivered rapidly in the event PLATO was not operable. This program was called STOIC (Standby Timing Operation In Contingencies)—hence, the frequently encountered convention "STOIC" file or, more appropriately, "UTPM STOIC" file. Sometimes, these files are referred to by their historical convention as "TPOLY" files or simply as "TP" arrays.

[5] *DSN Tracking System Interfaces, Earth Orientation Parameter Data Interface (TRK-2-21), DSN System Requirements Detailed Interface Design*, JPL 820-13, Rev. A (internal document), Jet Propulsion Laboratory, Pasadena, California, April 19, 1985.

[6] *DSN System Functional Requirements and Design: Tracking System (1988 Through 1993)*, JPL 821-19, Rev. C (internal document), Jet Propulsion Laboratory, Pasadena, California, pp. 3-20, April 15, 1993.

[7] *NOCC Subsystem Functional Requirements: Navigation Subsystem (1988 Through 1993)*, JPL 822-18, Rev. A (internal document), Jet Propulsion Laboratory, Pasadena, California, pp. 3-7–3-8, May 15, 1988.

sigma) in each component, predictive, for the days on which the calibrations are generated; (b) 5 cm (1-sigma) in each component, non-predictive, for periods through 14 days prior to the day on which the calibrations are generated; (c) 5 to 25 cm (1-sigma) in each component of polar motion, non-predictive, for periods from 1962 through 1984; and (d) 10 to 40 cm (1-sigma) in UT1, non-predictive, for periods from 1962 through 1984."[8]

The exact origin of the 30-cm real-time knowledge requirement is not widely known, although it is clear that it was arrived at via the common practice of synthesizing past flight project navigation team requirements and what the current calibration activity claimed could be delivered in terms of accuracy and timeliness. There is a common misconception that the 30-cm functional requirement for all three components of UTPM was driven by Magellan mission requirements. In actuality, the Magellan 30-cm requirement was inherited directly from the Galileo project for a 30-cm real-time UTPM knowledge requirement.[9] The UTPM requirements levied by future missions (e.g., Cassini, MGS) vary from flight project to flight project and are subject to change. Therefore, mission-specific requirements will not be presented here. It is fair to state that an effort is under way to update the overall Earth orientation calibration functional requirements for Mars Pathfinder (precession/nutation as well as UTPM) based on the analysis presented in this article.

## V. Information Content Analysis

Early analytic studies suggest that Earth orientation uncertainties result in equivalent uncertainties in the instantaneous location of tracking stations, which leads to a degradation in the apparent quality of the radio metric data used for navigation [15–17]. As noted in the introductory remarks, timing (UT1) errors in particular can lead to an erroneous prediction of the spacecraft coordinates near planetary encounter. Much of Doppler data's information content, when acquired during interplanetary cruise, comes from the diurnal signature of the Earth's rotation. This is evident in a simple analytic representation of the instantaneous range rate, $\dot{\rho}$, observed by an Earth-based tracking station [1–3]:

$$\dot{\rho} = v_r + r_s\omega_e \cos\delta \sin\omega_e t + (-\Delta\alpha + \Delta\lambda + \omega_e \Delta UT1) r_s \omega_e \cos\delta \cos\omega_e t \tag{1}$$

Here, $v_r$ denotes the spacecraft radial velocity with respect to the Earth; $r_s$ is the distance of the station from the Earth's spin axis, $\omega_e$ denotes the rotation rate of the Earth, and $t$ is measured from the nominal time the spacecraft crosses the tracking station's meridian. The $\delta$ is the instantaneous declination of the spacecraft, $\Delta\alpha$ the correction to the a priori value of spacecraft right ascension, $\Delta\lambda$ the correction to the station longitude, and $\Delta UT1$ the correction to rotation about the spin axis. There are, of course, other parameters to be estimated. Moreover, this simple model neglects the additional geometric strength that comes from the motion of the Earth about the Sun and the use of multiple tracking stations. Nevertheless, this model is useful to illustrate (to first order) the effect of Earth orientation errors on the Doppler data.

It is clear from Eq. (1) that an error in rotation about the Earth's spin axis would directly affect the right ascension estimate. For example, at Deep Space Station (DSS) latitudes (~35 deg), a 1-ms timing error is equivalent to a longitude error of about 0.4 m, or a right ascension error of about 0.07 $\mu$rad [13]. Polar motion affects the spacecraft position estimate by producing displacements in the station spin radius, longitude, and height above the equator. These displacements can be as large as 10 m if not properly calibrated. Equation (1) expresses the spacecraft right ascension and declination with respect to the Earth's equator of date. Errors in precession and nutation models can lead to errors in the transformation of the "of-date" right ascension and declination estimate into the inertial coordinate

[8] Ibid.

[9] S. N. Mohan and W. L. Sjogren, "Revised Navigation Requirement Specification for the VRM Mission Requirements Document 630-6 and Preliminary Spacecraft Instrumentation Requirements Document (SIRD)," JPL Interoffice Memorandum 314.10-348, Rev. 1 (internal document), Jet Propulsion Laboratory, Pasadena, California, September 22, 1983.

system of the planetary ephemeris. Precession/nutation modeling errors are rarely significant for Earth-orbiting spacecraft, where the observational data tie the spacecraft orbit much more tightly to a local coordinate system. For interplanetary spacecraft, the trajectory determined by Earth-based radio metric data must be related to the position of a distant planet.

## VI. Navigation Error Analysis

To investigate the effect of various levels of Earth orientation calibration accuracy on interplanetary cruise navigation, an error covariance analysis of the Mars Pathfinder approach segment was performed. The Mars Pathfinder approach scenario was selected because it has the most stringent planetary approach navigation requirements of any currently planned mission. Future Mars lander missions may utilize different trajectory designs and potentially could exhibit a lesser or greater level of sensitivity to Earth orientation calibration errors than those presented herein. This analysis is intended to serve as a representative model.

### A. Calibration Strategies (Test Cases)

In order to study the effect of Earth orientation calibrations on Mars Pathfinder cruise navigation, six test cases were developed to cover a wide range of possible Earth orientation calibration strategies. The level of calibration errors, which are a function of time, depends on the amount of data included in creation of the calibration files and on the timeliness of their deliveries. Precession/nutation calibrations were not included in this study since it is possible to predict the corrections for about a year with an accuracy approaching ~1 mas. All cases of Earth orientation studied here assume a basic set of VLBI measurements that can provide this level of precession/nutation accuracy (cf., Section II).

The Earth orientation calibration cases are characterized by the uncertainty in UT1 and polar motion as a function of time and by the correlations between the errors at different times. The reference day for the Earth orientation calibration cases is the day on which the navigation solution is performed. Figure 2 shows the assumed uncertainty in UT1 for the six cases, while Fig. 3 illustrates the assumed polar motion uncertainties. To simplify the analysis, the indicated level of polar motion uncertainty was assumed for both the X and Y components independently, even though actual measurements show that uncertainty in predictions for polar motion Y increase about 30 percent slower than for polar motion X [12]. Table 1 gives additional information about the statistics of each Earth orientation case. In all cases, the errors due to the potentially sparse array of calibration times imposed by the STOIC file have been neglected since this effect can be removed either by use of the EOP file or by use of a STOIC file that spans the shortest time possible.

The baseline Earth orientation case is the current nominal DSN capability and is indicated as TEMPO in the figures and in Table 1 [6].[10] This is based on two DSN VLBI measurements per week combined with data available from other sources. It is assumed that the last processed TEMPO VLBI measurement of UT1 was acquired 5 days before the KEOF filter run and that the KEOF filter run is performed 1 day prior to the navigation solution. For UT1 prediction, the rate of change in UT1 is important. The UT1 rate is dependent on the last two processed TEMPO measurements. For this particular case, UT1 was characterized as a first-order Gauss–Markov random process with a 1-sigma steady-state uncertainty of 0.11 ms and a 5-day correlation time until 7 days prior to the navigation solution; from this time forward, UT1 uncertainty was characterized by an integrated random walk (through the time of the navigation solution). The current KEOF filter solutions include the TEMPO VLBI measurements and AAM measurements and forecasts to give the stated capability for UT1 accuracy on any given day. In addition, daily VLBI measurements from external services are included in the KEOF filter solutions to provide the steady-state uncertainty of 0.11 ms for times in the past. (The 0.11 ms is larger than the

[10] A. P. Freedman, "Polar Motion Prediction With KEOF," JPL Interoffice Memorandum 335.2-92.01 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 5, 1992.

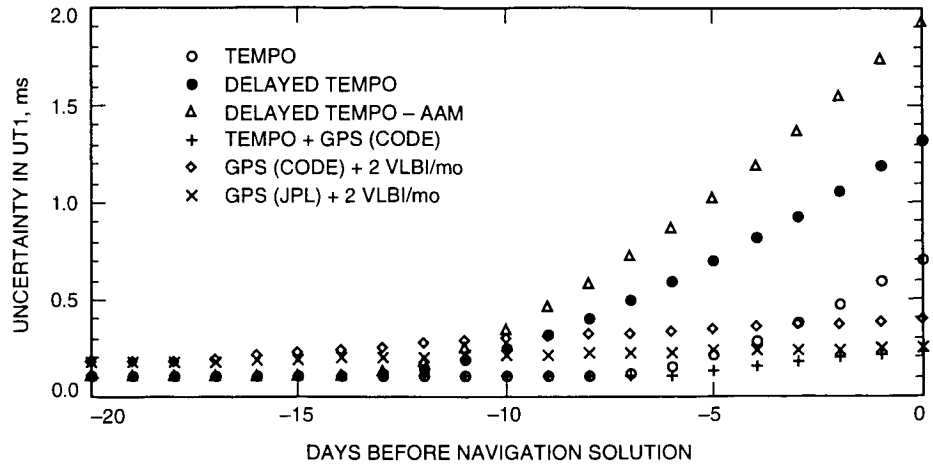


**Fig. 2. Modeled UT1 errors versus time for six different calibration strategies.**

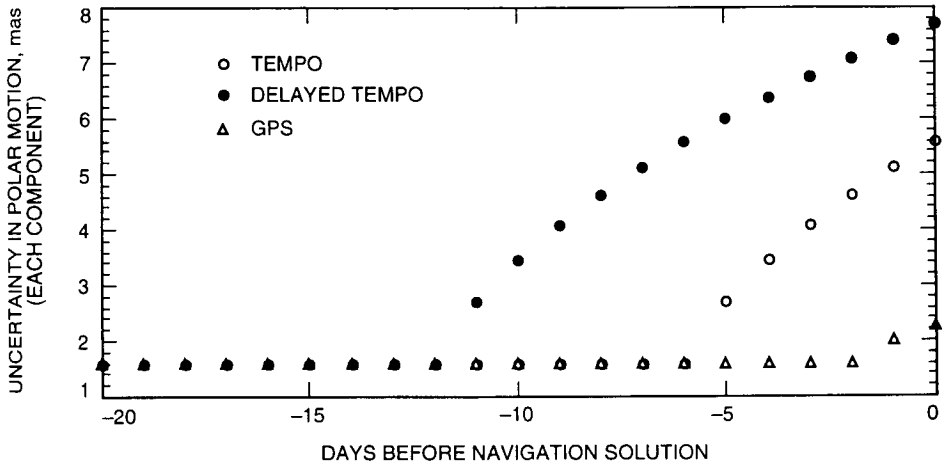


**Fig. 3. Modeled polar motion errors versus time for three different calibration strategies.**

**Table 1. Earth orientation calibration accuracies for various strategies.**

| Calibration strategy | TAI – UT1 1-sigma, ms | | Polar motion 1-sigma, mas | |
|---|---|---|---|---|
| | −21 days | 0 days | −21 days | 0 days |
| TEMPO (current) | 0.11 | 0.71 | 1.6 | 5.6 |
| Delayed TEMPO | 0.11 | 1.32 | 1.6 | 7.7 |
| Delayed TEMPO – AAM | 0.11 | 1.93 | 1.6 | 7.7 |
| TEMPO + GPS (CODE) | 0.11 | 0.23 | 1.6 | 2.3 |
| GPS (CODE) + 2 VLBI/mo | 0.18 | 0.39 | 1.6 | 2.3 |
| GPS (JPL) + 2 VLBI/mo | 0.18 | 0.25 | 1.6 | 2.3 |

quoted measurement uncertainties in order to accommodate possible offsets and drifts between various Earth orientation services.) In addition, the KEOF includes polar motion determinations from SLR measurements of Earth-orbiting satellites, which are available up to 5 days before the filter run (6 days before the navigation solution). The uncertainty in each component of polar motion was modeled as first-order Markov with a 5-day correlation time and a 1-sigma steady-state uncertainty of 1.6 mas, up to 6 days before the navigation solution, at which time the uncertainty was modeled as a random walk increasing to the final time. For times later than 6 days before the navigation solution, polar motion was modeled as a random walk that approximated the observed polar motion statistics [12]. (An integrated Gauss–Markov process was not used due to the difficulty in implementing it in the covariance analysis software.)

To investigate the importance of timely Earth orientation calibration delivery, two cases were included with a 6-day delay between the KEOF filter run and the navigation solution. The case labeled "delayed TEMPO" in Figs. 2 and 3 and in Table 1 is identical to the baseline case except for an additional 6-day delay. For the delayed TEMPO case, UT1 uncertainty is assumed to grow as an integrated random walk 13 days before the navigation solution, and polar motion begins to grow as a random walk 11 days before the navigation solution. A third case, labeled "delayed TEMPO – AAM," is identical to the delayed TEMPO case except that AAM data in the KEOF solution are not included. Without the AAM data, the UT1 uncertainty begins growing as an integrated random walk 13 days prior to the navigation solution, but at a faster rate. The polar motion uncertainty is identical for those two cases, i.e., delayed TEMPO and delayed TEMPO – AAM.

Measurements of GPS satellites have been used extensively for geodetic purposes, and GPS data have a demonstrated capability to measure polar motion and LOD. (Recall LOD is directly related to UT1 rate.) For the past 2 years, the Center for Orbit Determination, Europe (CODE) has been producing daily measurements of polar motion and LOD with a week or more delay between data acquisition and Earth orientation delivery. There are plans for the DSN to begin rapid processing of GPS data to supplement and partially replace TEMPO VLBI measurements in an effort to reduce loading on the DSN's 70-m subnetwork. If the measurements of LOD are uncorrelated (i.e., "white"), then including LOD measurements implies a random walk noise on UT1. Three cases of VLBI and GPS data combinations are included here assuming that the GPS LOD measurements are uncorrelated. The actual noise characteristics are under investigation. If the LOD measurements turn out to be correlated,[11] then the effect of GPS Earth orientation calibrations on navigation error may be different than the results presented in this article.

The fourth Earth orientation case, labeled "TEMPO + GPS(CODE)," assumes the current level of external VLBI measurements, the current two TEMPO VLBI passes each week, plus GPS polar motion and LOD measurements with a 1-day processing time. For this case, UT1 was assumed to behave as a Gauss–Markov process with a 5-day correlation time and a 1-sigma steady-state uncertainty of 0.11 ms until 7 days before the navigation solution, at which time the UT1 uncertainty was assumed to grow as a random walk at a level characteristic of the CODE GPS LOD deliveries. Each component of polar motion is described by a Gauss–Markov process with a 5-day correlation time and a 1.6-mas steady-state uncertainty (1-sigma) until 2 days before the navigation solution, at which time the polar motion uncertainty increases as a random walk. (This polar motion uncertainty model is assumed for all three cases that include GPS data.)

Because the current DSN plan is to utilize GPS LOD measurements so as to acquire fewer VLBI measurements, and because the number of external VLBI services has been steadily declining, the fifth Earth orientation case assumes that only two VLBI measurements are acquired per month and combined

[11] Preliminary studies of JPL GPS-derived LOD measurements exhibit "nonwhite" behavior on time scales longer than 3–5 days, A. P. Freedman, personal communication, Tracking Systems and Applications Section, Jet Propulsion Laboratory, Pasadena, California, August 18, 1995.

with GPS measurements. The case, labeled "GPS(CODE) + 2VLBI/mo" in the figures and in Table 1 assumes a 10-day delay in processing the VLBI measurements. This delay may occur as a result of reducing the load on the 70-m subnetwork, whereby VLBI measurements are acquired using the 34-m subnetwork. This strategy would require tapes to be used to record the VLBI data and shipped back to JPL for processing. With a 10-day delay between VLBI data acquisition and final processing, the latest VLBI measurement the KEOF could possibly include would be 11 days before the navigation solution with a worst-case delivery of 25 days before the navigation solution. For this case, it was assumed that the UT1 uncertainty behaved as a Gauss–Markov process with a 5-day correlation time and a 1-sigma steady-state uncertainty of 0.18 ms until 18 days prior to the navigation solution. This higher steady-state uncertainty is due to the lack of daily VLBI measurements from external services and reflects the uncertainty from using daily GPS LOD measurements to interpolate between VLBI UT1 measurements. At 18 days before the navigation solution, the UT1 uncertainty is assumed to grow as a random walk at a level characteristic of the CODE LOD measurements.

The current DSN plan is to have in place a JPL rapid GPS processing system for Earth orientation. The 3-year implementation cycle will begin in fiscal year 1996 with provisional operations beginning as early as fiscal year 1998. This will give way to a fully operational system by fiscal year 1999.[12] The processing implementation plan is under development but, as a test, there have been daily GPS solutions for LOD performed since late 1994. These solutions do not span a long enough time period to provide a good statistical measure of performance, but preliminary results indicate the JPL LOD measurements may be twice as accurate as the CODE deliveries. The sixth Earth orientation case, labeled "GPS(JPL) + 2VLBI/mo," is identical to the previous test case, GPS(CODE) + 2VLBI/mo, except that at 18 days before the navigation solution, the uncertainty in UT1 is assumed to begin a random walk behavior with a slower growth rate.

### B. Mars Pathfinder Tracking and Error Modeling Assumptions

The Mars Pathfinder spacecraft will directly enter the Martian atmosphere from Earth transfer orbit for landing on the Martian surface. Other missions (e.g., Cassini, MGS) will either fly by the target planet or enter orbit through a series of orbital correction maneuvers. The primary atmospheric entry constraint for Mars Pathfinder is the flight path angle, the angle between the incoming velocity vector of the spacecraft and the vector normal to the Martian atmosphere. If this angle is too large (shallow), the spacecraft may overheat before parachute deployment, and if the angle is too small (steep), excess pressure may develop that could potentially damage the spacecraft's aeroshell from ablation. This entry angle constraint is expected to place the most stringent requirements on calibration of Earth orientation. A secondary requirement is to target the spacecraft to land within a predetermined landing footprint on the Martian surface. The size of the landing footprint is 100 km × 300 km.

The Mars Pathfinder spacecraft will be spin stabilized throughout its interplanetary cruise to Mars and will communicate through its high-gain antenna. The onboard telecommunications system has an X-band (7.2-GHz) uplink/X-band (8.4-GHz) downlink radio system, which will be used to acquire Doppler and ranging measurements and to transmit science and engineering telemetry data. The nominal launch window is a 30-day launch period beginning on December 5, 1996.[13] Arrival at Mars is scheduled to occur on July 4, 1997. The launch vehicle will be targeted so that it will not impact the Martian surface. In the first 60 days after launch, two trajectory correction maneuvers (TCMs) will be performed to remove the effects of launch vehicle injection errors and to remove the targeting bias. A third TCM (TCM-3) is scheduled to be executed 60 days prior to Mars atmospheric entry. The critical navigation event time is just before the final maneuver (TCM-4). Five days prior to TCM-4, a navigation solution will be generated to design the final maneuver. The maneuver design command parameters will

[12] S. M. Lichten, personal communication, Tracking Systems and Applications Section, Jet Propulsion Laboratory, Pasadena, California, January 1995.

[13] At the time this article went to print, the actual launch window was not yet fixed since the mission profile and spacecraft launch mass were still being refined.

be uplinked to the spacecraft for execution from 10 to 15 days before atmospheric entry. Expected trajectory uncertainties for this critical navigation delivery have been carefully studied by Thurman and Kallemeyn[14,15,16] via linear covariance analysis and Monte Carlo simulation. The covariance analysis assumptions adopted herein to assess the sensitivity of the critical Mars Pathfinder navigation solution to various Earth orientation calibration strategies were derived in large part from these earlier navigation performance assessments.

The nominal Mars Pathfinder trajectory is a so-called "Type I" trajectory, where the heliocentric longitude of the spacecraft changes by less than 180 deg between launch and arrival. An alternative "Type II" trajectory, where the heliocentric longitude of the spacecraft changes by more than 180 deg and less than 360 deg between launch and arrival, was originally considered for Mars Pathfinder. Analysts who first studied the Type II trajectory option suggest that the principal reasons the Type I trajectory option was preferred were (1) to attempt to minimize 70-m antenna conflicts between Mars Pathfinder at arrival and the Galileo mission at Jupiter, (2) to shorten the cruise time from ~11 months to ~7 months, which would yield less consumables in terms of propellant, and (3) to attain a more favorable geometry for the spacecraft to remain at Earth-point during cruise while maximizing the Sun's exposure to the solar arrays.[17] (The Sun–probe–Earth angle is small for this mission.) A navigation error analysis for the Type II option was included in this assessment because some future missions to Mars (including MGS) will utilize Type II trajectories.

**Table 2. Assumed data arc lengths for Mars Pathfinder navigation analysis.**

| Trajectory | Launch/arrival date | Data arc specification | | |
|---|---|---|---|---|
| | | Begin, days[a] | End, days[b] | Length, days |
| Type I | January 3, 1997/ July 4, 1997 (fixed) | $L + 60$ | $M - 15$ | 107 |
| Type II | December 2, 1996/ November 10, 1997 (fixed) | $L + 236$ | $M - 15$ | 107 |

[a] $L$ = launch.

[b] $M$ = Mars arrival.

The tracking data arcs assumed for the covariance analysis are shown in Table 2 for both the Type I and Type II trajectories. X-band two-way coherent Doppler and ranging data were simulated over these intervals. DSN coverage varied according to the nominal DSN data acquisition schedule specified in the Mars Pathfinder *Navigation Plan.*[18] For the Type I transfer phase ($L+60$ days to $M-45$ days), the DSN coverage was taken to be one 4-h pass/week per complex; during the Mars approach phase ($M-45$ days to Mars arrival), continuous coverage was assumed; and for the TCM-3 phase, one 8-h pass/day (continuous for 12 h before and after TCM-3) was assumed over the interval TCM ± 3 days. The same data arc

[14] S. W. Thurman, "Orbit Determination Filter and Modeling Assumptions for MESUR Pathfinder Guidance and Navigation Analysis," JPL Interoffice Memorandum 314.3-1075 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 15, 1993.

[15] *Navigation Plan: Preliminary Version*, Pathfinder Flight Project, JPL D-11349 (internal document), Jet Propulsion Laboratory, Pasadena, California, December 1993.

[16] *Navigation Plan: Critical Design Review Version*, Mars Pathfinder Project, JPL D-11349 (internal document), Jet Propulsion Laboratory, Pasadena, California, July 1994.

[17] V. M. Pollmeier, personal communication, Navigation and Flight Mechanics Section, Jet Propulsion Laboratory, Pasadena, California, March 1995.

[18] *Navigation Plan: Critical Design Review Version*, op. cit.

length was used for the Type II trajectory; thus, simulated data points began at $L + 236$ days. In an effort to minimize the effects of potential station or complex outages while maximizing the angle-finding capability of the ranging data, tracking passes were scheduled to alternate between DSN complexes.

The Doppler and ranging data were assumed to have measurement uncertainties of 0.09 mm/s (60 s average) and 2 m, respectively (1-sigma). Although recent X-band Doppler data residuals are typically smaller than 0.09 mm/s, a higher Doppler uncertainty was assumed in order to reflect the low-frequency power of the solar plasma noise spectrum that is not properly characterized by the root-mean-square of the residuals.[19] A 20-min integration time was assumed for each data point (for both data types).

The Mars Pathfinder trajectories were integrated from initial position and velocity conditions (epoch state) using models for the dynamic forces on the spacecraft. The modeled gravitational forces were due to the masses of the Sun and the planets; relative locations of these bodies were based on the JPL DE200 ephemeris. Other forces modeled were nongravitational accelerations due to solar radiation pressure (SRP), gas leaks from valves and pressurized tanks, and attitude maintenance activity. In addition, TCM-3 maneuver execution errors were modeled.

Parameters estimated by the data reduction algorithm (a variant of the sequential Kalman filter [18]) included a wide array of dynamic and observational error sources categorized as (1) spacecraft epoch state, (2) spacecraft nongravitational force modeling errors, (3) maneuver execution errors, (4) errors in the orbital elements of the Earth and Mars, (5) systematic Doppler and ranging error biases, (6) transmission-media zenith delay calibration errors for the ionosphere and troposphere, (7) crust-fixed station location errors, and (8) Earth orientation calibration errors for UTPM. All of these error sources and their assumed a priori and steady-state values are summarized in Table 3. A priori uncertainties for the spacecraft initial state were large enough to leave it essentially unconstrained, while nongravitational forces were modeled as first-order Gauss–Markov random processes. (Note that all nongravitational forces except the slowly varying SRP accelerations were modeled using a stochastic gas leak model and are lumped under the category "NGA" in the table, where NGA denotes nongravitational accelerations.) TCM-3 execution errors (for the TCM-4 delivery) were modeled as random biases in all three body-fixed components. The uncertainty in the Earth–Mars ephemeris was taken from the JPL DE234 ephemeris error covariance by Standish,[20] but constrained with the knowledge that the orientation of the Earth's orbit is now known to 15 nrad [19]. For processing the two-way ranging data, the filter model included a bias parameter associated with each ranging pass from each station in order to approximate the slowly varying nongeometric delays in the ranging measurements that are caused principally by station delay calibration errors and uncalibrated solar plasma effects. The spacecraft spin rate, detectable in the Doppler signature, was estimated as a Gauss–Markov process with a 5-day correlation time. Uncertainty in knowledge of the station locations was assumed to be 10 cm for each component. This station location uncertainty is expected to be characteristic of the new DSN beam-waveguide (BWG) antennas once surveys are complete. More accurate station locations exist for antennas for which VLBI data are available, including the 70-m antennas and the 34-m high-efficiency (HEF) antennas.[21]

Although this study is restricted to the orbit determination problem and does not address the influence of guidance errors on navigational accuracy, it is important to note that, upon completion of TCM-4, the contribution of maneuver execution errors to the overall guidance dispersions are expected to be negligible. This was demonstrated in preflight error analyses and is discussed in greater detail in the Mars Pathfinder *Navigation Plan*.[22]

---

[19] W. M. Folkner, "Effect of Uncalibrated Charged Particles on Doppler Tracking," JPL Interoffice Memorandum 335.1-94-005 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 1, 1994.

[20] E. M. Standish, "The JPL Planetary Ephemerides, DE234/LE234," JPL Interoffice Memorandum 314.6-1348 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 8, 1991.

[21] In actuality, the Mars Pathfinder spacecraft will be "uplink-limited" and will, therefore, require use of the 34-m HEF antennas for telecommunication. A more conservative assessment is made herein by assuming the 34-m BWG antennas.

[22] *Navigation Plan: Critical Design Review Version*, op. cit.

**Table 3. A priori and steady-state uncertainties for orbit determination error model parameters.**

| Estimated parameter set | Uncertainty, 1 $\sigma$ | Remarks |
|---|---|---|
| Spacecraft epoch state | A priori | Constant parameters |
| Position components | 100 km | |
| Velocity components | 1 m/s | |
| Nongravitational force model | | |
| Solar radiation pressure (SRP) | Steady-state | First-order Markov |
| Radial ($G_r$) | 5% of nominal | 60-day correlation time |
| Transverse ($G_x/G_y$) | 5% of nominal | |
| Gas leaks (NGA) | Steady-state | First-order Markov |
| Radial ($a_r$) | $2 \times 10^{-12}$ km/s$^2$ | 5-day correlation time |
| Transverse ($a_x/a_y$) | $2 \times 10^{-12}$ km/s$^2$ | 5-day correlation time |
| Maneuver execution error model | | |
| TCM-3 ($\Delta V_x, \Delta V_y, \Delta V_z$) | A priori | Constant parameters |
| (for TCM-4 delivery) | $10^{-2}$ m/s | |
| Planetary ephemerides error model | | |
| Earth-Mars ephemeris | A priori | Constant parameters |
| Orbit orientation (3 Euler angles) | 15 nrad | |
| Longitude with respect to periapsis | 10 nrad | |
| Semimajor axis ($\Delta a/a$) | 5 parts in $10^{11}$ | |
| Eccentricity ($\Delta e$) | 3 parts in $10^{10}$ | |
| Ground system error model | | |
| Range biases | A priori | Constant parameters |
| (one per station per pass) | 1 m | |
| Transponder bias | Steady-state | First-order Markov |
| (ranging data only) | 1 m | 0.5-day correlation time |
| Doppler spin bias | Steady-state | First-order Markov |
| (Doppler data only) | $10^{-2}$ mm/s | 5-day correlation time |
| Transmission media | Steady-state | First-order Markov |
| Zenith troposphere | 5 cm | 0.1-day correlation time |
| Zenith ionosphere | $5 \times 10^{16}$ e/m$^2$ | 0.2-day correlation time |
| DSN station coordinates | A priori | Constant parameters |
| (crust-fixed $r_s, z_h, \lambda$) | 10 cm | (uncorrelated) |
| Earth orientation | (cf., Section VI.A) | (cf., Section VI.A) |
| Timing (UT1) | | |
| Polar motion (X,Y) | | |

## C. Encounter Geometry

Because much of the strength of the Doppler and ranging data comes from the signature imposed by the rotation of the Earth, interpretation of the covariance analysis results is aided by understanding the encounter geometry.

The spacecraft position in the Earth–spacecraft direction is directly measured by ranging data. Doppler data help determine the other two components of the spacecraft position, which lie in the plane of the sky. There is a well-known weakness in determining spacecraft declination from Doppler data for spacecraft

near zero declination [1-3]. Spacecraft declination can be inferred from ranging data using tracking stations located in both northern and southern latitudes [20]. Figure 4 shows the Pathfinder Type I trajectory on the plane of the sky as viewed from Earth. As seen in the figure, encounter occurs near zero declination. Because of this encounter geometry, the spacecraft declination will probably depend upon ranging data, and the declination uncertainty should exhibit sensitivity to station delay calibration errors. In contrast, the Type II trajectory has a relatively large, negative encounter declination, as shown in Fig. 5. For the Type II encounter, the Doppler data will have a larger role in determining spacecraft declination, which should thus be less sensitive to station delay calibration errors.

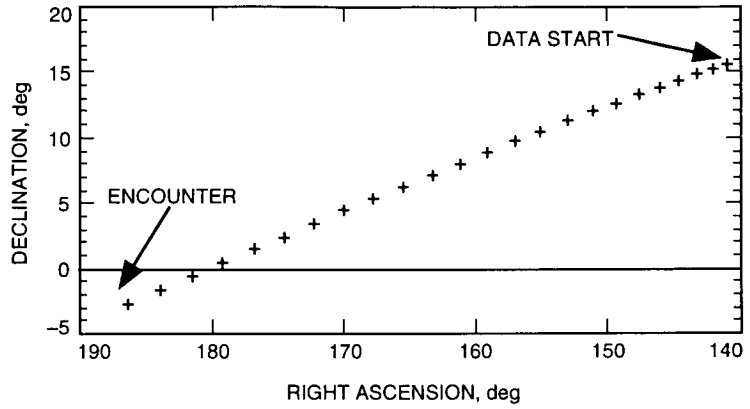


**Fig. 4. Mars Pathfinder Type I trajectory as viewed from Earth; shown at 5-day intervals from the beginning of the data arc to encounter.**

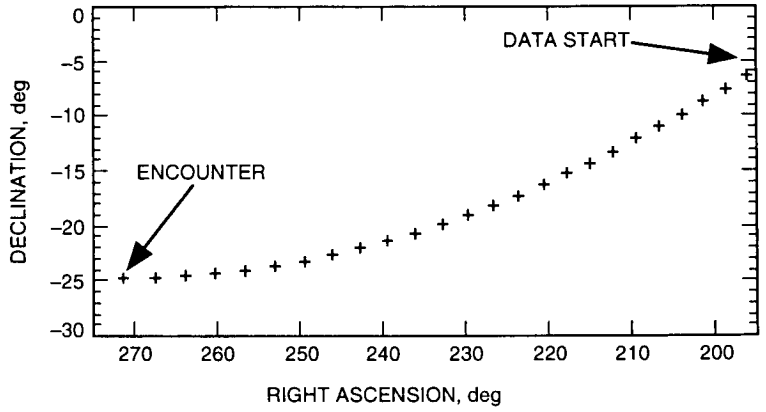


**Fig. 5. Mars Pathfinder Type II trajectory as viewed from Earth; shown at 5-day intervals from the beginning of the data arc to encounter.**

A typical uncertainty ellipsoid for the spacecraft position on approach would have principal axes approximately aligned with the plane-of-sky axes, with a much smaller uncertainty in the Earth–Mars direction than in the other two components (assuming ranging data are included). Planetary approach trajectories are typically described in aiming plane ($B$-plane) coordinates.[23] Figure 6(a) shows the relationship of the $B$-plane components to the plane-of-sky components for the Pathfinder Type I encounter. The approach direction, $\widehat{\mathbf{S}}$, is nearly parallel to the radial (Earth–Mars) direction, $\widehat{\mathbf{r}}$. The $-\widehat{\mathbf{R}}$ direction is in the plane normal to the approach direction, $\widehat{\mathbf{S}}$, and approximately parallel to the direction of increasing declination, $\widehat{\delta}$, while the $-\widehat{\mathbf{T}}$ direction is in the plane normal to the approach direction and approximately

[23] For a complete description of the $B$-plane coordinate system, please refer to Appendix B.

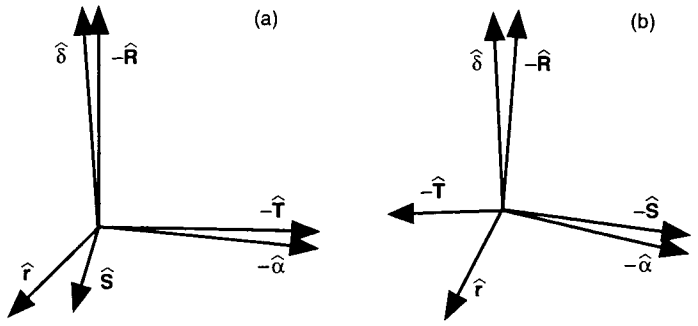


**Fig. 6. The *B*-plane components for Mars Pathfinder approach trajectories with respect to plane-of-sky coordinates: (a) Type I and (b) Type II.**

parallel to the direction of decreasing right ascension, $-\hat{\alpha}$. For the Type I trajectory, the well-determined component is approximately in the direction of approach. A small position uncertainty in this direction is expressed in the $B$-plane system as a small uncertainty in the time from encounter, i.e., linearized time of flight (LTOF). The position uncertainty is approximately related to the LTOF uncertainty by the approach velocity. For the Type I trajectory, the approach velocity is about 5.5 km/s (a 1-s uncertainty in LTOF corresponds to a position error of about 5.5 km). An error in right ascension, such as might be caused by a UT1 calibration error, will appear in the $\mathbf{B} \cdot \hat{\mathbf{T}}$ component.

Figure 6(b) shows the relationship of the $B$-plane components to the plane-of-sky components for the Type II trajectory. The direction of the spacecraft approach to Mars, $-\hat{\mathbf{S}}$, is about 11 deg from the direction of decreasing right ascension, $-\hat{\alpha}$. The $-\hat{\mathbf{R}}$ direction is about 23 deg from the declination axis, $\hat{\delta}$. The $-\hat{\mathbf{T}}$ direction is about 23 deg from the Earth–Mars direction, $\hat{\mathbf{r}}$. For this trajectory, an error in right ascension will be reflected mainly in LTOF. For the Type II trajectory, the approach velocity is approximately 3.9 km/s.

Mars Pathfinder navigation is required to deliver, prior to the final maneuver (TCM-4), a trajectory estimate with less than a 1-percent probability of exceeding the entry angle requirement. The latest assessment of the Type I flight path entry angle requirement is ±1 deg (99 percent), which implies a requirement on the navigation delivery corresponding to a 3-sigma uncertainty of 21 km in the magnitude of the impact parameter.[24] Stated another way, the entry corridor is 42-km wide, as depicted in Fig. 7.

## D. Results

In the covariance studies performed, a careful model was constructed for the time-dependent Earth orientation errors shown in Figs. 2 and 3. This model would be somewhat difficult to implement into the operational Orbit Determination Program (ODP),[25] which currently does not have a statistical reset capability or an integrated random walk model such as the one used in this analysis. Because of this limitation, the effect of each Earth orientation calibration strategy on the total orbit determination error was calculated in two ways. For the first estimation method, the contribution to orbit determination error from Earth orientation was determined with UT1 and polar motion included (i.e., estimated) in the navigation solution, with correctly modeled time-dependent a priori uncertainties. In the second estimation method, the contribution to orbit determination error was assessed under the assumption that the Earth orientation calibration errors were ignored (i.e., not estimated) in the navigation solution.

[24] P. K. Kallemeyn, personal communication, Navigation and Flight Mechanics Section, Jet Propulsion Laboratory, Pasadena, California, March 1995.

[25] The ODP is a large institutional software system used for research and navigation support of flight operations, N. D. Panagiotacopulos, J. W. Zielenbach, and R. W. Duesing, *An Introduction to JPLs Orbit Determination Program*, JPL 1846-37 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 21, 1974.

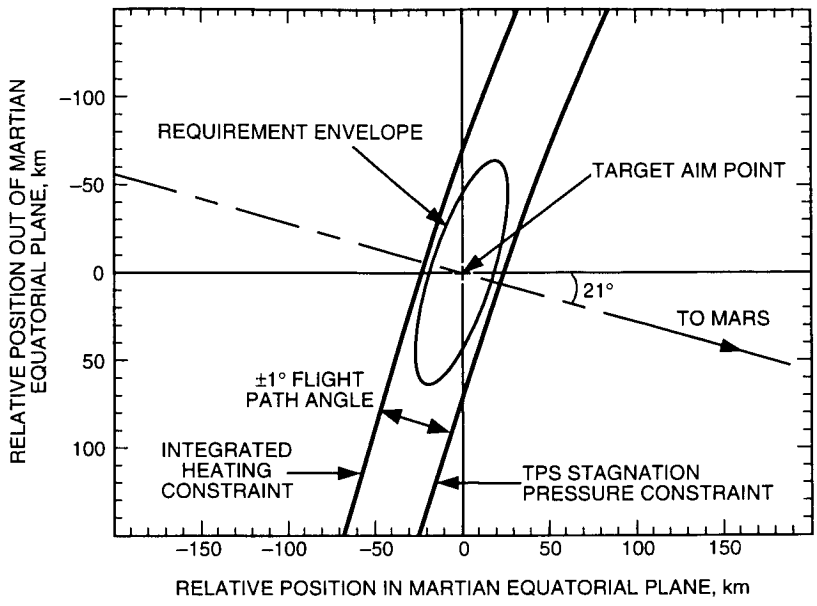


**Fig. 7. Mars Pathfinder entry corridor and landing accuracy requirements (99 percent). (Note: the Earth equatorial plane is nearly parallel to the direction to Mars.)**

Figure 8 shows the contribution to the total navigational uncertainty for the nominal (Type I) Mars Pathfinder trajectory from all error sources described in Section VI.B except Earth orientation. The covariance analysis results given below are expressed in $B$-plane components referred to the Earth mean equator of J2000 (EME2000), as described in Appendix B. Contributions were computed in a manner such that the sum of each error source, when added in quadrature, gives the total navigation uncertainty in a root-sum-square sense [21]. The critical navigation parameter for Pathfinder approach is the magnitude of the impact parameter, denoted $|\mathbf{B}|$. Recall from the previous discussion that $|\mathbf{B}|$ is related to the flight path entry angle. For the nominal Pathfinder Type I approach trajectory, the $\mathbf{B} \cdot \hat{\mathbf{T}}$ uncertainty, denoted $\sigma_{\mathbf{B}\cdot\hat{\mathbf{T}}}$, is nearly the same as the uncertainty in $|\mathbf{B}|$. In general, the relationship of the component uncertainties $\sigma_{\mathbf{B}\cdot\hat{\mathbf{R}}}$, $\sigma_{\mathbf{B}\cdot\hat{\mathbf{T}}}$, and $\sigma_{|\mathbf{B}|}$ depends upon the choice of the targeted entry point.

The uncertainties in arrival time (LTOF) are very small because of the approach direction nearly coinciding with the Earth–Mars direction, which is well determined by ranging data. The scale for LTOF in Fig. 8(c) is 3 s and corresponds to a position uncertainty of about 15 km. The major error source (other than Earth orientation) for $\mathbf{B} \cdot \hat{\mathbf{T}}$ (~right ascension) is the anomalous nongravitational accelerations (NGAs). The $\mathbf{B} \cdot \hat{\mathbf{R}}$ (~declination) uncertainty has roughly equal contributions from data noise, nongravitational forces, and station delay calibrations for ranging data. The 1-m accuracy of the range bias calibrations assumed for the covariance analysis has been inferred from observations of the day-to-day consistency of Mars Observer ranging data residuals [22]. This assumption should be interpreted cautiously since the systematic effects in the Mars Observer range biases could have been absorbed by other spacecraft trajectory parameters, such as nongravitational accelerations. Fortunately, this is not an issue for Mars Pathfinder since the critical navigation component, $|\mathbf{B}|$, is almost entirely in the right ascension direction. Further, this navigation error analysis was not intended to be the "official" Mars Pathfinder analysis. The principal purpose here was to provide a quantitative measure of the relative importance of potential error sources, specifically, Earth orientation calibration errors.

Figure 9 illustrates the contribution to the total orbit determination uncertainty from each case of Earth orientation calibration error described in Section VI.A. Here, it is seen that Earth orientation calibration errors are a significant source of error for Mars Pathfinder in the critical $\mathbf{B} \cdot \hat{\mathbf{T}}$ and $|\mathbf{B}|$

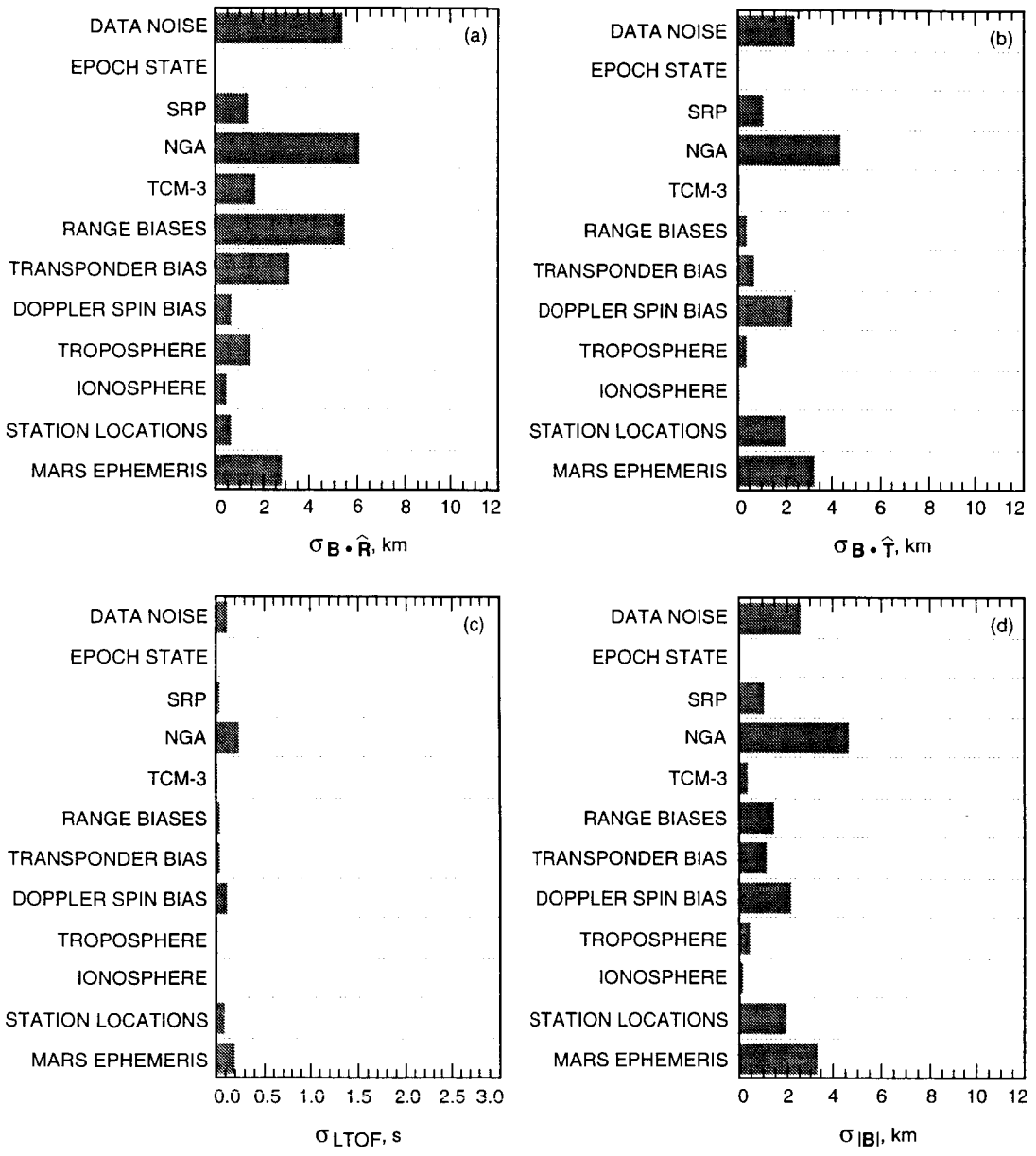


**Fig. 8. Relative contributions of the principal error sources (other than Earth orientation) to the total Mars Pathfinder orbit determination uncertainty. Uncertainties are shown in *B*-plane coordinates with respect to the mean Earth equator of 2000: (a) B • R̂ (~declination) uncertainty, (b) B • T̂ (~right ascension) uncertainty, (c) LTOF (time of encounter) uncertainty, and (d) uncertainty in the magnitude of the impact parameter, |B|.**

components.[26] Earth orientation calibration error is less significant in the $\mathbf{B} \cdot \hat{\mathbf{R}}$ and LTOF components. The lack of sensitivity to Earth orientation in the LTOF direction is due to the fact that the approach direction is nearly aligned with the Earth–Mars direction; therefore, LTOF is well determined by the ranging data. The spacecraft declination (nearly aligned with the $\mathbf{B} \cdot \hat{\mathbf{R}}$ direction) is determined largely by ranging data at northern and southern latitude stations since, at the low encounter declination, the Doppler data do not contribute much to the determination of declination. Because the declination is

[26] Recall that errors due to precession and nutation were neglected from this analysis; thus, the formal Earth orientation calibration errors are strictly due to UTPM calibration errors.

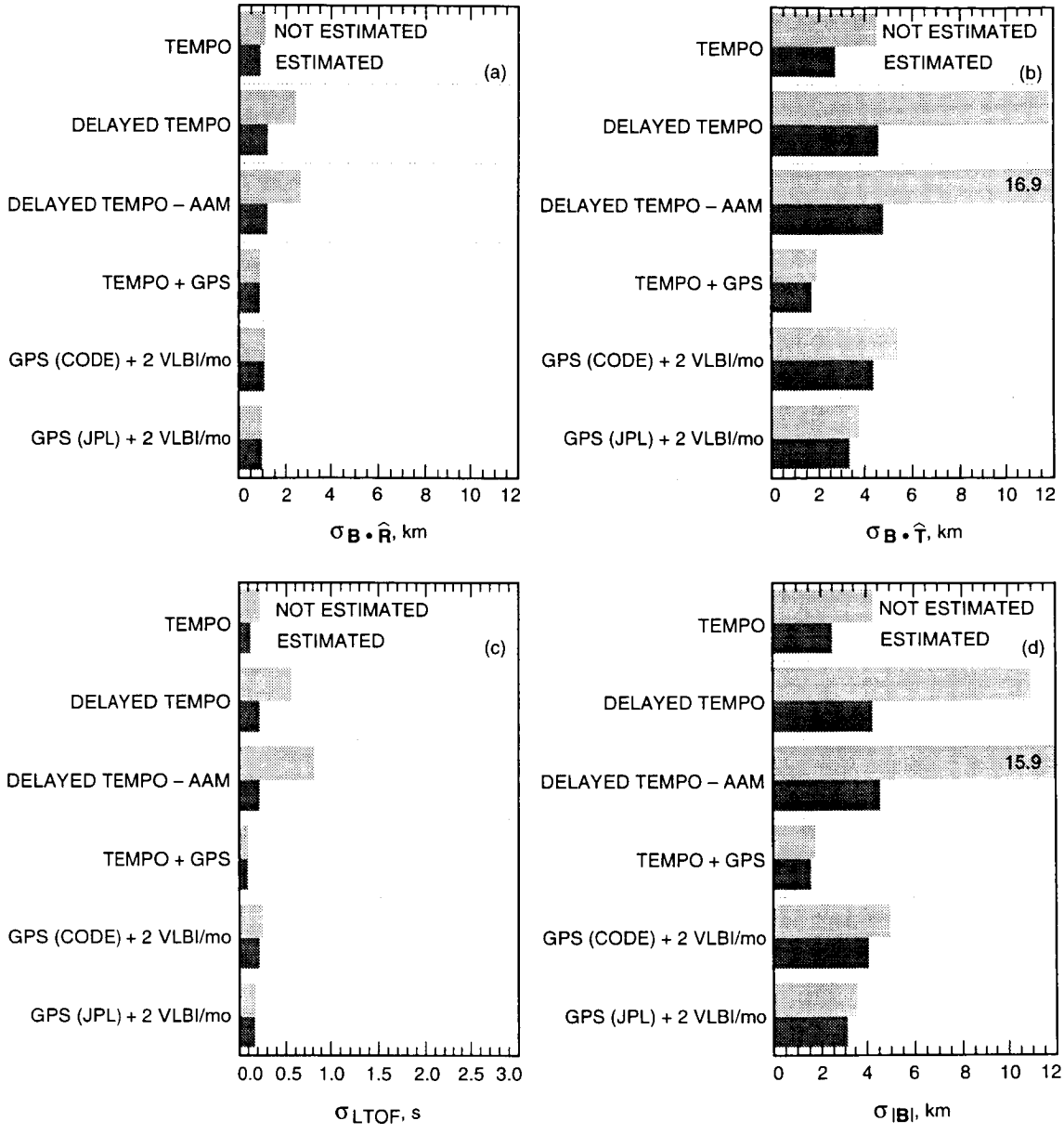


**Fig. 9. Relative contributions of Earth orientation to the total Mars Pathfinder orbit determination uncertainty. Uncertainties are shown in *B*-plane coordinates with respect to the mean Earth equator of 2000: (a) B • R̂ (~declination) uncertainty, (b) B • T̂ (~right ascension) uncertainty, (c) LTOF (time of encounter) uncertainty, and (d) uncertainty in the magnitude of the impact parameter, |B|. Uncertainties are given for both the case where Earth orientation parameters were estimated in the navigation solutions and for cases where Earth orientation was not adjusted in the navigation solution.**

determined principally by ranging data with an assumed accuracy of 1 m, there is not much sensitivity to Earth orientation errors for the calibration strategies studied here, which all give Earth orientation errors smaller than 1 m at the Earth's surface.

In the case of current DSN Earth orientation calibration performance, assuming a delivery of the calibration files on the day of the critical navigation solution from TEMPO VLBI data, Fig. 9(d) shows that Earth orientation errors contribute approximately 39 percent of the 1-sigma $|\mathbf{B}|$ (flight path entry angle) requirement of 7 km for the case where UTPM parameters were included in the navigation solution,

and 64 percent of the allowable error if UTPM were ignored in the navigation solution. The two Earth orientation calibration cases with delayed delivery show contributions to navigation uncertainty that are significantly larger. The delayed calibration cases are most likely unacceptable for the Mars Pathfinder mission. The optimistic Earth orientation case, in which the current twice-weekly TEMPO VLBI measurements are augmented with daily GPS data, shows a much smaller contribution to the navigation uncertainty than the nominal TEMPO case. The two calibration strategies with daily GPS data combined with reduced VLBI observations (2 VLBI/month) are comparable to the nominal TEMPO case. In contrast to the nominal TEMPO case, the GPS-based calibrations exhibit smaller differences between the strategy of including UTPM parameters and statistics in the navigation solution and the strategy of ignoring the UTPM parameters in the navigation solution.

Figure 9 shows a reduced sensitivity to Earth orientation errors when the UTPM parameters are estimated, along with the trajectory parameters, in the navigation filter. This improvement is large for the cases with poorest UTPM accuracy. The improvement is coupled to the assumptions about the level of nongravitational forces affecting the spacecraft. If there were no nongravitational forces acting on the spacecraft, or if the nongravitational forces were perfectly known, then the spacecraft would provide a reference against which Earth orientation changes could be measured using Doppler data. If there were large nongravitational forces affecting the spacecraft that are not well known, then the spacecraft could not be used as a reference against which Earth orientation changes could be measured. Because the Pathfinder spacecraft will be a simple, spinning platform, the nongravitational forces affecting it are assumed here to be well modeled. Because of this assumption, when the Earth orientation uncertainties increase beyond a certain level, the navigation filter begins to rely on the assumed level of nongravitational force uncertainties and can improve upon the a priori knowledge assumed for Earth orientation parameters. This would not be true for a spacecraft with larger uncertainties in the nongravitational force model.

Because of the different encounter geometry, the covariance analysis results for the Type II trajectory are quite different from the Type I trajectory. No attempt was made to quantify the critical navigation component, $|\mathbf{B}|$, since the Type II trajectory will not be used for Mars Pathfinder and the choice of the targeted point for this study was arbitrary. The $B$-plane component uncertainties should be interpreted in such a manner that the critical component could be more like $\mathbf{B} \cdot \widehat{\mathbf{R}}$ or $\mathbf{B} \cdot \widehat{\mathbf{T}}$, depending on the choice of the targeted point.

Figure 10 shows the navigation uncertainty from all error sources for the Type II trajectory with the exception of Earth orientation calibration error. The LTOF uncertainty is about a factor of six larger for the Type II case than for the Type I case because the approach direction is not aligned with the Earth–Mars direction. The scale in Fig. 10(c), 3 s, corresponds to a position uncertainty of about 12 km due to the approach velocity of 3.9 km/s. Nongravitational forces, Mars ephemeris uncertainty, and data noise are seen to be the dominant sources of error (other than Earth orientation) for the other components. The $\mathbf{B} \cdot \widehat{\mathbf{T}}$ component is most closely aligned with the Earth–Mars direction at encounter and, hence, is the best determined component. The uncertainty in $\mathbf{B} \cdot \widehat{\mathbf{R}}$ ($\sim$declination) shown in Fig. 10(a) for the Type II trajectory is less sensitive to ranging calibration errors and more sensitive to station location errors than is the Type I case. This is a reflection of the large, negative encounter declination enabling Doppler data to influence the determination of declination.

Figure 11 shows the contribution of Earth orientation calibration errors to the orbit determination uncertainty for the Type II trajectory. The $\mathbf{B} \cdot \widehat{\mathbf{R}}$ uncertainty is much more dependent on Earth orientation than is the Type I case. This sensitivity is related to the determination of declination by the Doppler data, which are sensitive to Earth platform errors. The sensitivity of declination to Earth orientation can be seen to be principally due to polar motion errors since cases with identical polar motion uncertainties, but different UT1 uncertainties, have the same effect on the $\mathbf{B} \cdot \widehat{\mathbf{R}}$ uncertainty. The $\mathbf{B} \cdot \widehat{\mathbf{T}}$ component shows some sensitivity to UT1 errors since the $\widehat{\mathbf{T}}$ direction is mostly in the Earth–Mars direction but partly in the direction of increasing right ascension. UT1 errors have a larger effect on LTOF since the

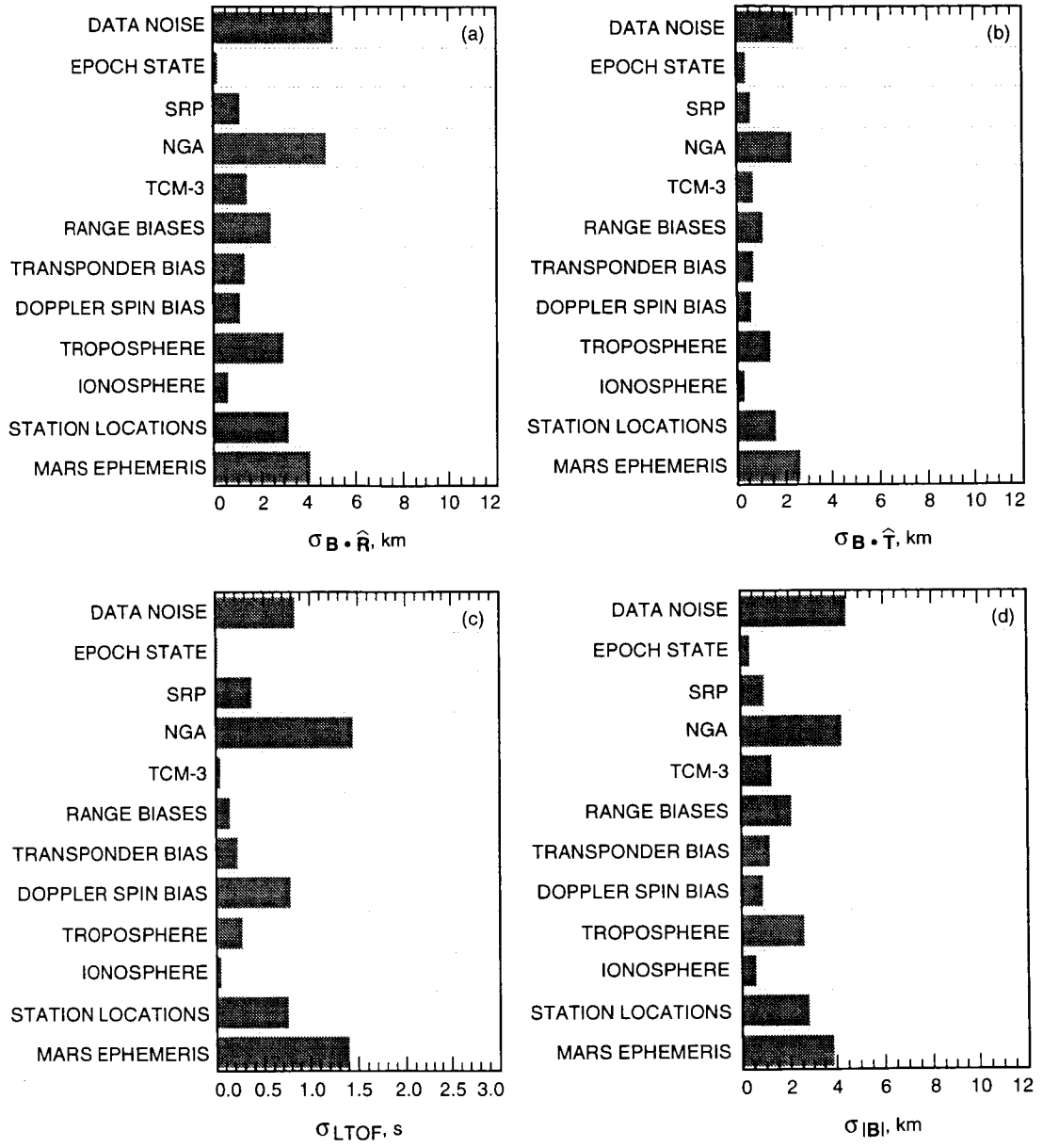


**Fig. 10. Relative contributions of the principal error sources (other than Earth orientation) to the total orbit determination uncertainty for a Mars Pathfinder Type II approach scenario. Uncertainties are shown in *B*-plane coordinates with respect to the mean Earth equator of 2000: (a) B • R̂ (~declination) uncertainty, (b) B • T̂ (~right ascension) uncertainty, (c) LTOF (time of encounter) uncertainty, and (d) uncertainty in the magnitude of the impact parameter, |B|.**

approach direction is more closely aligned with the right ascension direction. The nominal TEMPO Earth orientation errors would be one of the larger sources for error in $\mathbf{B} \cdot \hat{\mathbf{T}}$ and a moderate source of error in $\mathbf{B} \cdot \hat{\mathbf{R}}$ for this trajectory. The GPS-based cases contribute less to the navigation uncertainty in $\mathbf{B} \cdot \hat{\mathbf{R}}$ and $\mathbf{B} \cdot \hat{\mathbf{T}}$ than the nominal TEMPO case, but result in large errors in LTOF.

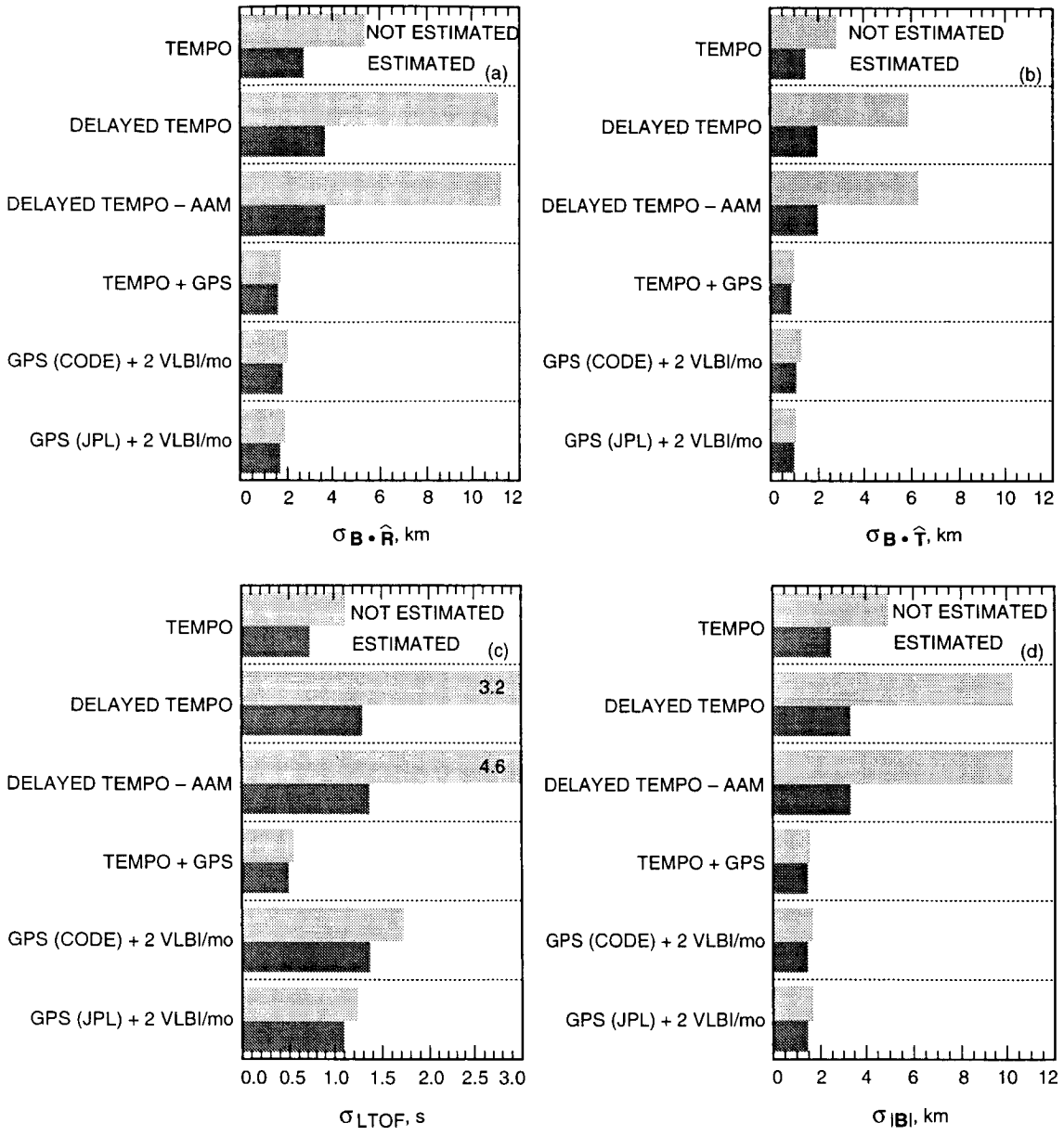


**Fig. 11. Relative contributions of Earth orientation for a Mars Type II approach trajectory. Uncertainties are shown in *B*-plane coordinates with respect to the mean Earth equator of 2000: (a) B • R̂ (~declination) uncertainty, (b) B • T̂ (~right ascension) uncertainty, (c) LTOF (time of encounter) uncertainty, and (d) uncertainty in the magnitude of the impact parameter, IBI. Uncertainties are given for both the case where Earth orientation parameters were estimated in the navigation solutions and for cases where Earth orientation was not adjusted in the navigation solution.**

## VII. Summary and Conclusions

A numerical assessment measuring the sensitivity of spacecraft delivery errors to the accuracy and timeliness of Earth orientation calibrations was completed for two interplanetary cruise scenarios derived from the Mars Pathfinder mission set. This study was motivated by the fact that, to date, errors in Earth orientation (i.e., precession/nutation, polar motion, and variation in Earth rotation rate) are still capable

of contributing significantly to the composition of the noise signature on radio metric data acquired by the DSN. These errors can thus lead to degraded spacecraft navigational accuracies if not adequately calibrated.

Results from the navigation sensitivity analysis concurred with the expected outcome that not only is Earth orientation calibration performance important in determining spacecraft navigational accuracy, but so is the timeliness of the calibration file deliveries. Based on the analyses presented in this article, the current best DSN Earth orientation calibration performance provided by the TEMPO activity yielded a contribution of about 39 to 64 percent of the total navigation error budget for the critical component of the nominal Mars Pathfinder Type I trajectory, depending on the navigation filtering strategy being used. These results assumed line-of-sight data types (i.e., two-way Doppler and range) were used in the navigation process. Use of differential data types could reduce the sensitivity to Earth orientation calibration errors.

Variations on the current DSN calibration method representing delayed TEMPO deliveries of the calibration files as well as delayed deliveries without use of AAM data were also assessed. Results for these cases showed that Earth orientation calibration errors dominated the total navigation error budget, irrespective of the trajectory type. Furthermore, a very large penalty was paid when the Earth orientation parameters were not adjusted in the navigation solution.

With the advent of GPS-based ground observations as a viable Earth orientation calibration system and the ongoing effort to reduce the loading on the DSN 70-m subnetwork, new Earth orientation calibration techniques are being devised. Statistical models representing examples of these calibration strategies were constructed and their effect on the Mars Pathfinder navigation delivery error assessed. In the (optimistic) case where the current level of TEMPO calibrations (2 per week) was used in concert with daily GPS-based calibrations, the influence of UTPM calibration errors on overall navigation performance was, as expected, minimal. Under the current environment where there is continual pressure to reduce the number of DSN-based VLBI observations (again, addressing the 70-m antenna loading issue), this calibration strategy will probably not be attainable operationally.

A sensitivity analysis was also performed for an operationally more realistic Earth orientation calibration strategy in which GPS-based calibrations were used as the principal means of generating frequent (daily) Earth orientation calibration information, augmented with periodic VLBI-based measurements ($\sim$2 per month). (The GPS system alone cannot determine all components of Earth orientation and, thus, requires an external calibration source such as VLBI.) In this assessment, analysis results suggest that the contribution of UTPM errors to the total navigation error budget for the critical component of the nominal Mars Pathfinder trajectory lies somewhere between 43 and 55 percent, depending on the accuracy of the GPS deliveries. These results assumed that the UTPM parameters were adjusted in the navigation solution. The true level of accuracy will depend, of course, on the actual system implemented.

Since the GPS calibration system is in the early stages of development, the statistical characteristics of the calibrations are not yet well determined. With the noise levels assumed for this analysis, the GPS-based Earth orientation calibrations appear to offer an advantage over the current TEMPO-based calibrations in that they relax the need for the navigation process to properly model the time-varying behavior of UTPM calibration errors. In addition, the proposed system is designed to provide rapid processing and timely deliveries of the calibration files to the flight projects. The overall performance (accuracy) levels, as evidenced in this study, were at or near the same level as the current DSN capability, perhaps only slightly better in some cases.

## Acknowledgments

The authors would like to thank the following individuals, who greatly assisted in preparation of this article: Alan Steppe, who graciously agreed to referee this article and assisted with several excellent technical discussions regarding correct statistical modeling of the Earth orientation calibration errors; Adam Freedman, who helped determine the level of uncertainty to assign to each Earth orientation calibration strategy; Carl Christensen, Jean Dickey, Tom Runge, Jim Border, and George Resch for providing a number of useful technical comments based on early drafts of this article; and Pieter Kallemeyn and Vince Pollmeier for the frequent tutorial sessions on the Mars Pathfinder mission. Without the help of all of these individuals, this research task could not have been accomplished.

It is hoped that this article will assist future mission study teams and end users in the understanding of errors associated with the DSN Earth orientation calibration process and the potential limiting effect these errors can have on spacecraft radio navigation performance.

## References

[1] J. O. Light, "An Investigation of the Orbit Redetermination Process Following the First Mid-Course Maneuver," *Supporting Research and Advanced Development, Space Programs Summary 37-33*, vol. IV, Jet Propulsion Laboratory, Pasadena, California, pp. 8–17, June 30, 1965.

[2] T. W. Hamilton and W. G. Melbourne, "Information Content of a Single Pass of Doppler Data From a Distant Spacecraft," *The Deep Space Network, Space Programs Summary 37-39*, vol. III, Jet Propulsion Laboratory, Pasadena, California, pp. 18–23, May 31, 1966.

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

[4] N. A. Renzetti, editor, *A History of the Deep Space Network: From Inception to January 1, 1969*, JPL Technical Report 32-1533, vol. I, Jet Propulsion Laboratory, Pasadena, California, September 1, 1971.

[5] P. K. Seidelmann, "1980 IAU Theory of Nutation: The Final Report of the IAU Working Group on Nutation," *Celestial Mechanics*, vol. 27, pp. 19–106, 1982.

[6] A. P. Freedman, J. A. Steppe, J. O. Dickey, T. M. Eubanks, and L.-Y. Sung, "The Short-Term Prediction of Universal Time and Length of Day Using Atmospheric Angular Momentum," *Journal of Geophysical Research*, vol. 99, no. B4, pp. 6981–6996, April 10, 1994.

[7] T. A. Herring, B. A. Buffett, P. M. Mathews, and I. I. Shapiro, "Forced Nutations of the Earth: Influence of Inner Core Dynamics 3. Very Long Interferometry Data Analysis," *Journal of Geophysical Research*, vol. 96, pp. 8259–8273, 1991.

[8] P. Charlot, O. J. Sovers, J. G. Williams, and X X Newhall, "Precession and Nutation From Joint Analysis of Radio Interferometric and Lunar Laser Ranging Observations," *Astronomical Journal*, vol. 109, pp. 418–427, 1995.

[9] J. A. Steppe, S. H. Oliveau, and O. J. Sovers, "Earth Rotation Parameters From DSN VLBI: 1994," *IERS Technical Note 17*, Observatoire de Paris, Paris, France, pp. R19–R32, 1994.

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

[11] J. H. Lieske, "Precession Matrix Based on IAU (1976) System of Astronomical Constants," *Astronomy and Astrophysics*, vol. 73, pp. 282–284, 1979.

[12] D. D. Morabito, T. M. Eubanks, and J. A. Steppe, "Kalman Filtering of Earth Orientation Changes," *Earth's Rotation and Reference Frames for Geodesy and Geodynamics*, edited by A. K. Babcock and G. A. Wilkins, Dordrecht, Holland: D. Riedel, pp. 257–267, 1988.

[13] N. A. Renzetti, J. F. Jordan, A. L. Berman, J. A. Wackley, and T. P. Yunck, *The Deep Space Network—An Instrument for Radio Navigation of Deep Space Probes*, JPL Publication 82-102, Jet Propulsion Laboratory, Pasadena, California, pp. 77–78 and 88–91, December 15, 1982.

[14] H. F. Fliegel and R. N. Wimberly, "Time and Polar Motion," *Tracking System Analytic Calibration Activities for the Mariner Mars 1971 Mission*, JPL Technical Report 32-1587, Jet Propulsion Laboratory, Pasadena, California, pp. 77–81, March 1, 1974.

[15] D. W. Trask and P. M. Muller, "Timing: DSIF Two-Way Doppler Inherent Accuracy Limitations," *The Deep Space Network, Space Programs Summary 37-39*, vol. III, Jet Propulsion Laboratory, Pasadena, California, pp. 7–16, May 31, 1966.

[16] P. M. Muller and C. C. Chao, "Timing Errors and Polar Motion," *Tracking System Analytic Calibration Activities for the Mariner Mars 1969 Mission*, JPL Technical Report 32-1499, Jet Propulsion Laboratory, Pasadena, California, pp. 35–43, November 15, 1970.

[17] J. F. Jordan, G. A. Madrid, and G. E. Pease, "Effects of Major Errors Sources on Planetary Spacecraft Navigation Accuracies," *Journal of Spacecraft and Rockets*, vol. 9, no. 3, pp. 196–204, March 1972.

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

[19] W. M. Folkner, P. Charlot, M. H. Finger, J. G. Williams, O. J. Sovers, X X Newhall, and E. M. Standish, Jr., "Determination of the Extragalactic-Planetary Frame Tie From Joint Analysis of Radio Interferometric and Lunar Laser Ranging Measurements," *Astronomy and Astrophysics*, vol. 287, pp. 279–289, 1994.

[20] S. W. Thurman, T. P. McElrath, and V. M. Pollmeier, "Short-Arc Orbit Determination Using Coherent X-Band Ranging Data," *Advances in the Astronautical Sciences*, vol. 79, part I, pp. 23–44, 1992.

[21] J. A. Estefan and P. D. Burkhart, "Enhanced Orbit Determination Filter Sensitivity Analysis: Error Budget Development," *The Telecommunications and Data Acquisition Progress Report 42-116, October–December 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 24–36, February 15, 1994.

[22] L. A. Cangahuala, E. J. Graat, D. C. Roth, S. W. Demcak, P. B. Esposito, and R. A. Mase, "Mars Observer Interplanetary Cruise Orbit Determination," *Advances in the Astronautical Sciences*, vol. 87, part II, pp. 1049–1068, 1994.

[23] L. J. Wood, "Orbit Determination Singularities in the Doppler Tracking of a Planetary Orbiter," *Journal of Guidance, Control, and Dynamics*, vol. 9, no. 4, pp. 485–494, July–August 1986.

[24] D. B. Engelhardt, J. B. McNamee, S. K. Wong, F. G. Bonneau, E. J. Graat, R. J. Haw, G. R. Kronschnabl, and M. S. Ryne, "Determination and Prediction of Magellan's Orbit," *Advances in the Astronautical Sciences*, vol. 75, part II, pp. 1143–1160, 1991.

[25] S. W. Thurman, "Information Content of Interferometric Delay-Rate Measurements for Planetary Orbiter Navigation," paper AIAA-90-2909, AIAA/AAS Astrodynamics Conference, Portland, Oregon, August 20–22, 1990.

[26] J. D. Giorgini, E. J. Graat, T.-H. You, M. S. Ryne, S. K. Wong, and J. B. McNamee, "Magellan Navigation Using X-Band Differenced Doppler During Venus Mapping Phase," paper AIAA-92-4521, AIAA/AAS Astrodynamics Conference, Hilton Head, South Carolina, August 10–12, 1992.

[27] W. Kizner, *A Method of Describing Miss Distances for Lunar and Interplanetary Trajectories*, JPL Publication 674, Jet Propulsion Laboratory, Pasadena, California, August 1, 1959.

# Appendix A

# Sensitivity of Planetary Orbiter Navigation to Earth Orientation—A Case Study for Differential Data Types

For a spacecraft in orbit about another planet, Doppler data can be used to determine all components of its orbit except for a few particular geometries [23]. The accuracy with which the orbit is determined by means of Earth-based Doppler tracking depends upon several factors, including data accuracy and the accuracy of the spacecraft force models, particularly those due to the planet's gravitational field. Using the Magellan radar mapping mission of Venus as an example, the uncertainty of the gravity field was such that the expected orbit uncertainty during the prime mission (for daily orbit solutions) was about 15 km with two-way Doppler tracking alone [24]. Mars Global Surveyor (MGS) plans to achieve much better accuracy by solving for an improved gravity field based on an initial data set. The MGS strategy could not have been utilized for Magellan since the gravity field of Venus could not be sampled with a few weeks of radio metric data because of Venus' slow rotation rate.

The orbit determination accuracy achievable with Earth-based Doppler tracking in a two-way coherent mode is very insensitive to Earth orientation errors since the dominant signature in the Doppler data is due to the orbit of the spacecraft about the planet. This is in contrast to planetary approach navigation, where much of the information content in Doppler tracking data is influenced by the Earth's rotation. To first order, Doppler data are insensitive to a rotation about the line of sight from the Earth-based tracking station to the spacecraft. The rotation about the line of sight can be determined by changes in the geometry due to the relative orbits of the Earth and the target planet about which the spacecraft is orbiting. Rotation about the line of sight is measured by the node angle, $\Omega$, with respect to the plane of the sky, which is defined in Fig. A-1 as the plane normal to the Earth–spacecraft direction. In the figure, the orbit inclination, $i$, is the angle between the normal to the spacecraft orbit and the Earth–spacecraft direction; the line of nodes is the intersection of the orbit plane and the plane of the sky; the node with respect to the plane of the sky, $\Omega$, is measured in the plane of the sky from a reference direction to the line of nodes; and the argument of periapsis, $\omega$, is the angle between the line of nodes and periapsis measured in the orbit plane.

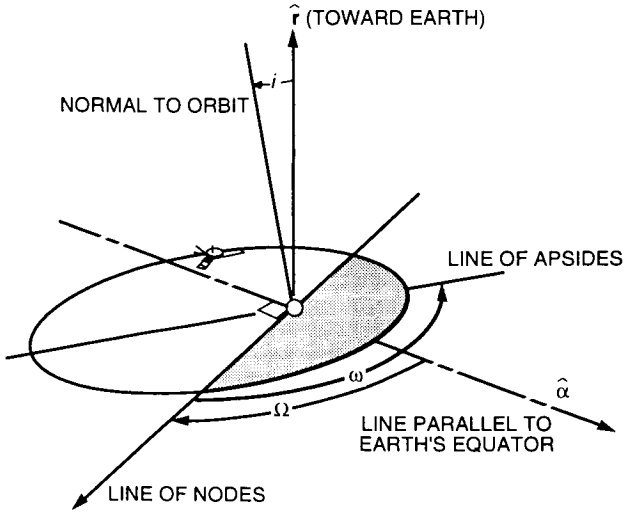


**Fig. A-1. Planetary orbiter geometry.**

At times, the desired orbit accuracy is greater than what can achieved with two-way Doppler tracking alone. In the case of Magellan, the desired orbit accuracy was about 1 km for purposes of aligning radar images. This level of accuracy was better than what could be achieved using Doppler data alone. Differential data types such as differenced one-way Doppler (DOD), delta-differenced one-way Doppler ($\Delta$DOD), or two-way minus three-way Doppler (2DM3D) have been shown to improve orbit determination accuracy [26]; the latter was used successfully for Magellan operations [27]. These differential data types are sensitive to Earth orientation errors. An assessment of the characteristic sensitivity to Earth orientation errors for planetary orbiter navigation when using differential data types is described below. The planetary orbiter scenario is based on the radar mapping phase of the Magellan mission.

Consider the geometry drawn schematically in Fig. A-2. The plane of the figure is taken to be the Earth's equatorial plane. (Note that Venus need not lie in the equatorial plane for the analysis to be valid.) Two stations at different complexes are located at the ends of the baseline vector with equatorial projection, $b_e$. For illustration purposes, an orbit is considered with the orbit plane perpendicular to the Earth's equatorial plane and with the normal to the orbit plane perpendicular to the Earth–Venus direction.

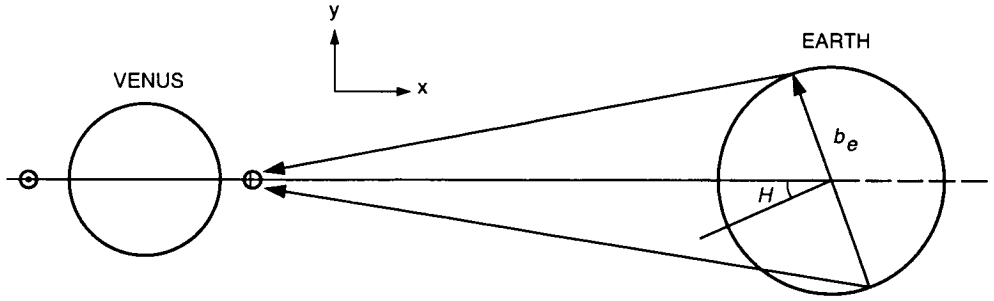


**Fig. A-2. Differenced Doppler measurement geometry used in the case study.**

DOD measurements are formed by differencing the one-way Doppler signals received by two tracking stations separated by large distances [26]. These measurements give the difference in spacecraft line-of-sight velocity as observed by the two stations. (The 2DM3D measurements exhibit the same information content except for a slight difference resulting from use of a DSN uplink signal rather than the spacecraft onboard oscillator as the reference frequency.) For spacecraft at interplanetary distances, the DOD observable can be approximated as

$$DOD \approx \frac{d}{dt}\left(\mathbf{b}\cdot\frac{\mathbf{r}}{r}\right) \approx \frac{1}{r}\left[\mathbf{b}\cdot\left(\mathbf{v}-\dot{r}\frac{\mathbf{r}}{r}\right)+\left(\frac{\mathbf{r}}{r}\times\boldsymbol{\omega}_e\right)\cdot\mathbf{b}\right] \tag{A-1}$$

where $\mathbf{b}$ is the baseline vector between the two tracking stations, $\mathbf{r}$ is the vector from the center of the Earth to the spacecraft with magnitude $r$, $\dot{r}$ is the rate of change of distance between the Earth and the spacecraft, $\mathbf{v}$ is the spacecraft velocity vector with magnitude $v$, and $\boldsymbol{\omega}_e$ is the Earth's rotation rate vector. By considering a DOD measurement for this special case, at the time when the spacecraft velocity is parallel to the Earth's pole, the DOD observable can be further approximated as

$$DOD \approx \frac{1}{r}v_y b_e \cos H + \frac{1}{r}v_z b_z + \omega_e b_e \cos H \tag{A-2}$$

where $v_y$ is the component of the spacecraft velocity in the equatorial plane and perpendicular to the Earth–Venus direction (and nominally zero at the measurement time), $b_e$ is the equatorial baseline length,

$v_z$ is the component of the spacecraft velocity parallel to the Earth's pole, $b_z$ is the length of the projection of the baseline length onto the pole direction, and $H$ is the hour angle between the baseline and the spacecraft. A rotation of the orbit about the Earth–Venus line by an angle $\delta\Omega$ changes $v_y$ from zero to $v\delta\Omega$, which is directly observable in the DOD measurement. If the measurement occurred earlier (or later) in the orbit, where the spacecraft velocity vector was along the spacecraft–Earth direction, there would be no change in spacecraft velocity for a change in the orbit node. In this case, the DOD measurement would not be useful. The importance of performing differenced Doppler measurements at optimum times has been well documented in the literature (see, e.g., [25]).

An error in UT1 introduces a bias in the hour angle, $H$, and, hence, in the DOD measurement. This can affect the determination of the spacecraft node angle. A change in the measurement due to a calibration error, $\delta UT1$, is approximately given by

$$\delta DOD \approx -b_e \omega_e^2 \sin H \delta UT1 \tag{A-3}$$

This change will cause an error to be inferred in the rotation about the line of sight by an amount

$$\delta\Omega \approx \frac{r}{v}\omega_e^2 \tan H \ \delta UT1 \tag{A-4}$$

For DSN baselines (Goldstone–Madrid and Goldstone–Canberra), $H$ can vary from about $-30$ to $+30$ deg, outside of which the spacecraft will fall below the horizon of one of the complexes. DSN baselines have a mean equatorial length of about 8000 km. For the worst case where $H = 30$ deg, a 1-ms error in UT1 will bias the DOD measurements by about 0.02 mm/s. For an orbiter characteristic of Magellan during its mapping phase, with an average orbital velocity, **v** of about 5.5 km/s, and a line-of-sight distance of 1 AU, a 1-ms timing error in UT1 would lead to a node error of up to 0.08 mrad. With a semimajor axis of 10,000 km, this corresponds to an orbit error of about 0.8 km. (Since this is comparable to the desired orbit accuracy for Magellan, it was necessary to have UT1 calibrated with submillisecond accuracy in order to support the generation of daily orbit determination solutions.)

In general, the maximum sensitivity of the differenced Doppler data to Earth orientation errors is of nearly the same magnitude as for the special case studied here. Sensitivity to Earth orientation errors can be an order of magnitude smaller if the data are acquired at times where the baseline hour angle is near zero and the spacecraft velocity at that time is in a direction where the data are sensitive to the spacecraft node. The size of the orbit errors also depends on (among a number of other factors) the shape of the orbit, the uncertainty in the gravity field, and the amount of Doppler data to be used in the "fit" (i.e., the data filtering process).

# Appendix B
# Definition of Aiming Plane (*B*-Plane) Coordinates

Planetary approach trajectories are typically described in aiming plane coordinates, often referred to as "$B$-plane" coordinates (see Fig. B-1). This coordinate system was originally conceived to simplify the targeting of a hyperbolic flyby trajectory and is defined by three orthogonal unit vectors, $\hat{\mathbf{S}}$, $\hat{\mathbf{T}}$, and $\hat{\mathbf{R}}$, with the system origin taken to be the gravitational center of mass of the target planet [27]. The $\hat{\mathbf{S}}$ is directed parallel to the incoming spacecraft asymptotic velocity vector relative to the target planet, while $\hat{\mathbf{T}}$ is normally specified to lie in either the ecliptic plane (the mean plane of the Earth's orbit) or the equatorial plane of the target planet.[27] In addition, $\hat{\mathbf{T}}$ is directed perpendicular to $\hat{\mathbf{S}}$. The unit vector $\hat{\mathbf{R}}$ completes an orthogonal triad with $\hat{\mathbf{S}}$ and $\hat{\mathbf{T}}$, thus, $\hat{\mathbf{R}} = \hat{\mathbf{S}} \times \hat{\mathbf{T}}$.

The aim point for a planetary encounter is defined by the impact parameter, $\mathbf{B}$, which approximates where the point of closest approach would be if the target planet had no mass and did not deflect the flight path. The impact parameter $\mathbf{B}$ is directed perpendicular to $\hat{\mathbf{S}}$; therefore, it lies in the $\hat{\mathbf{T}} - \hat{\mathbf{R}}$ plane. To gain insight into targeting accuracy, orbit determination errors are often characterized by the 1-sigma or 3-sigma uncertainty in the respective "miss components" of $\mathbf{B}$, namely, $\mathbf{B} \cdot \hat{\mathbf{R}}$ and $\mathbf{B} \cdot \hat{\mathbf{T}}$. These quantities are analogous to elevation and azimuth when specifying the impact point for terrestrial targets.

The time from encounter is defined by the linearized time of flight (LTOF), a quantity which is a measure of the "time-to-go" from the current spacecraft position to the intersection of its asymptotic flight path and the aiming plane. LTOF provides a convenient time-to-go parameter because LTOF is not affected by changes in the $\mathbf{B} \cdot \hat{\mathbf{R}}$ and $\mathbf{B} \cdot \hat{\mathbf{T}}$ miss components.[28] Orbit determination errors are also characterized by the 1-sigma or 3-sigma uncertainty in LTOF.

In lieu of using $\mathbf{B} \cdot \hat{\mathbf{R}}$ and $\mathbf{B} \cdot \hat{\mathbf{T}}$ uncertainties to measure targeting accuracy, a 1-sigma or 3-sigma $B$-plane dispersion ellipse (also shown in Fig. B-1) is often used. The semimajor (SMAA) and semiminor (SMIA) axes of the dispersion ellipse are related in quadrature to the uncertainties of $\mathbf{B} \cdot \hat{\mathbf{R}}$ and $\mathbf{B} \cdot \hat{\mathbf{T}}$. The angle $\theta_T$ gives the angle clockwise from $\hat{\mathbf{T}}$ to the SMAA.

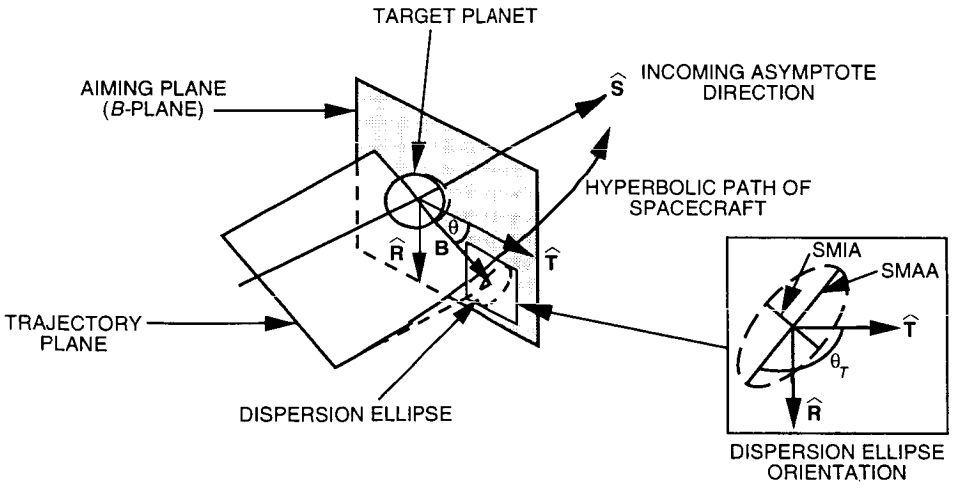


**Fig. B-1. The aiming plane (*B*-plane) coordinate system.**

[27] For the analysis presented in this article, $\hat{\mathbf{T}}$ was specified to lie in the Earth's equatorial plane.

[28] R. A. Jacobson, "Linearized-Time-of-Flight Revisited," JPL Engineering Memorandum 391-680 (Revised) (internal document), Jet Propulsion Laboratory, Pasadena, California, September 22, 1975.

# Wind Gust Models Derived from Field Data

W. Gawronski
Communications Ground Systems Section

*Wind data measured during a field experiment were used to verify the analytical model of wind gusts. Good coincidence was observed; the only discrepancy occurred for the azimuth error in the front and back winds, where the simulated errors were smaller than the measured ones. This happened because of the assumption of the spatial coherence of the wind gust model, which generated a symmetric antenna load and, in consequence, a low azimuth servo error. This result indicates a need for upgrading the wind gust model to a spatially incoherent one that will reflect the real gusts in a more accurate manner.*

*In order to design a controller with wind disturbance rejection properties, the wind disturbance should be known at the input to the antenna rate loop model. The second task, therefore, consists of developing a digital filter that simulates the wind gusts at the antenna rate input. This filter matches the spectrum of the measured servo errors. In this scenario, the wind gusts are generated by introducing white noise to the filter input.*

## I. Introduction

The steady-state wind pressure distribution on scaled antenna models was measured during wind tunnel experiments,[1,2,3] and their validity for actual field antennas was unknown. Wind field data collected recently at the DSS-13 antenna were used to evaluate the accuracy of the steady wind pressure measured in the wind tunnel [2]. A similar evaluation can be done for the time-varying part (gusts) of the wind.

The wind gust analytical model, as developed in [1], is used to simulate the pointing errors of the DSN antennas. The model was developed using the wind tunnel data (as in Footnotes 1 through 3) and the Davenport spectra, but its accuracy was unverified. In this article, the wind measurements of servo errors obtained on January 24, 1994, at the DSS-13 antenna site, c.f. [2], were compared with the simulated servo errors. In most cases, the comparison shows satisfactory coincidence between the measured and the simulated data.

[1] N. L. Fox and B. Layman, Jr., "Preliminary Report on Paraboloidal Reflector Antenna Wind Tunnel Tests," JPL Interoffice Memorandum CP-3 (internal document), Jet Propulsion Laboratory, Pasadena, California, 1962.

[2] N. L. Fox, "Load Distributions on the Surface of Paraboloidal Reflector Antennas," JPL Interoffice Memorandum CP-4 (internal document), Jet Propulsion Laboratory, Pasadena, California, 1962.

[3] R. B. Blaylock, "Aerodynamic Coefficients for a Model of a Paraboloidal Reflector Directional Antenna Proposed for a JPL Advanced Antenna System," JPL Interoffice Memorandum CP-6 (internal document), Jet Propulsion Laboratory, Pasadena, California, 1964.

Recently, the linear quadratic Gaussian (LQG) controller for the DSS-13 antenna was designed and tested (see [4]). This model-based controller used the identified DSS-13 antenna model based on field experiments [3]. This antenna model does not include the wind disturbances, which are necessary for the design of an improved LQG controller with wind disturbance rejection properties. For this purpose, the wind-measured data were used to create the wind disturbance input into the antenna rate-loop model and will serve as a base for the design of an improved controller with wind disturbance rejection properties.

## II. Evaluation of the Analytical Model

In the field experiment, the servo errors due to wind gusts were measured for the elevation angles from 11 to 89 deg and for the the yaw angles (antenna azimuth position with respect to the wind direction) from 0 to 360 deg. The servo errors from the analytical model are available for elevation angles of 60 and 90 deg and for yaw angles of 0 (front wind), 90 (side wind), and 180 deg (back wind). The results are obtained in the form of the standard deviations of the measured servo error, typically of the length of 8000 samples collected at a sampling time of 0.02 s. The results of field measurements and simulations are shown in Figs. 1 through 4, where "x" denotes field data and "o" denotes the analytical results. For the elevation servo error measurements, there were multiple collections of the field data for each elevation position. Thus, in this case, the maximal and minimal root-mean-square sums of the measured error are plotted with the gray area between them (see Figs. 1 through 3). For the azimuth errors, there was one collection of data, so the field errors do not include the gray area.

The elevation servo error plots indicate that the analytical error lies within the gray area of the min–max measurements, while the results of the azimuth error show very close relationship between the measured and simulated standard deviations of the servo error for the side wind and a discrepancy for the front and back winds. In the latter case, the analysis underestimates the error because of the symmetry

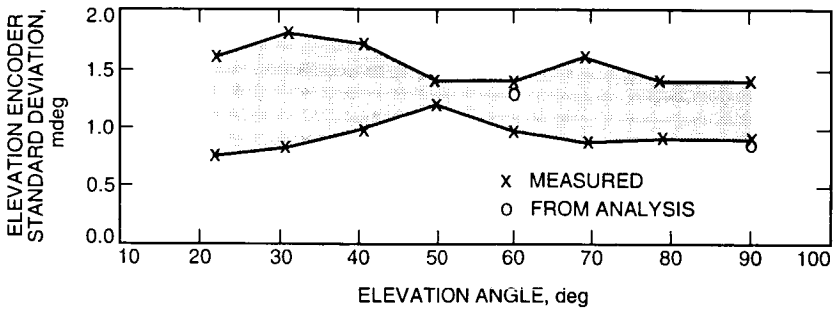


**Fig. 1. Standard deviation of the elevation encoder output due to 40-km/h wind front gusts.**

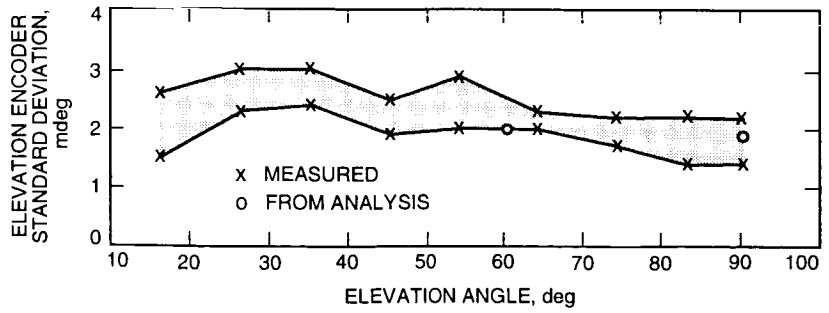


**Fig. 2. Standard deviation of the elevation encoder output due to 40-km/h wind side gusts.**

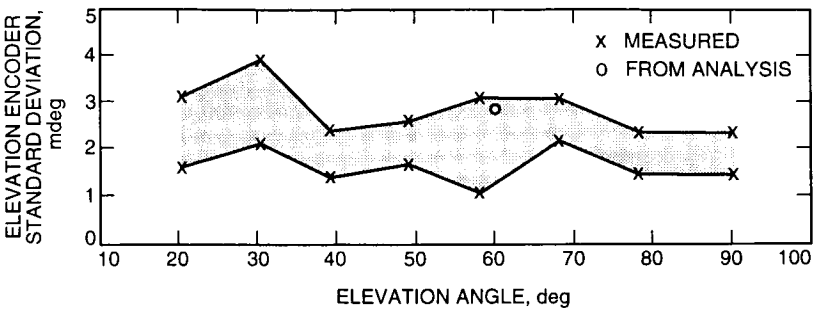


**Fig. 3. Standard deviation of the elevation encoder output due to 40-km/h wind back gusts.**

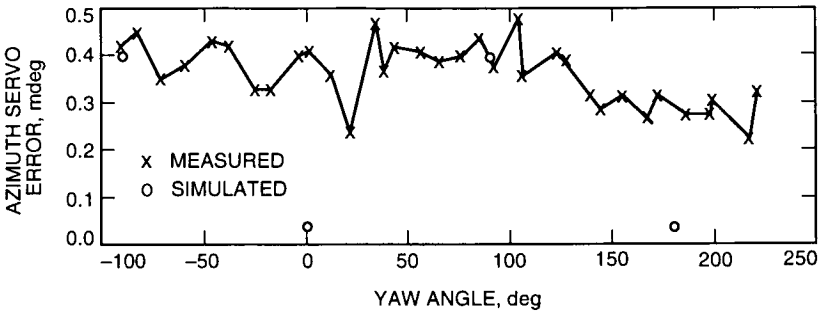


**Fig. 4. Standard deviation of the elevation encoder output due to 32-km/h wind gusts.**

and the coherence of the wind loading model. In reality, wind gusting is significantly unsymmetrical and is spatially uncorrelated. This discrepancy can be corrected by introducing an incoherent wind model using cross-spectra (see [5]).

## III. Wind Gust Model Derived From the Field Data

LQG controllers developed for the DSN antennas (see [4]) are based on the antenna model obtained from the field testing rather than on its analytical model. In order to evaluate the controller wind disturbance rejection properties as well as to improve these properties, one has to develop a wind disturbance model compatible with the antenna-identified model.

The antenna rate-loop model was identified for the azimuth and elevation loops separately. The cross-coupling between the azimuth and elevation axes, and vice versa, was low and was, therefore, ignored. The input to the model is the rate command, and the output is the encoder reading. The rate command creates difficulties in implementation of the analytical wind gust model because the wind is modeled as pressure at the antenna structure and is not readily transformed into the rate command disturbance, but this can be done by using the measured servo errors due to the wind gust disturbance.

Further, only the antenna model in azimuth is considered (the elevation model is developed similarly). The filter at the rate input will model the wind gusts (see Fig. 5). The white noise disturbance, $w(t)$, of unit intensity at the input to the wind filter is assumed. The filter transfer function, $G(s)$, is to be determined. The filter output, $w_r(t)$, adds to the rate command, and it serves as the wind gust model.

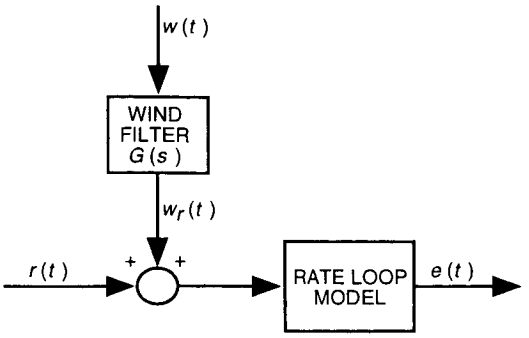


**Fig. 5. Wind filter configuration.**

Let the servo error due to the wind gusts be $e(t)$ and its spectrum be $e(\omega)$. The servo error due to disturbance $w(t)$ is $e_s(t)$, and its spectrum is $e_s(\omega)$. The filter transfer function, $G(\omega)$, is determined such that the difference between the simulated and the measured power spectrum is minimized as follows:

$$G(\omega) \text{ such that } \| e(\omega) - e_s(\omega) \| \text{ is minimal} \tag{1}$$

The filter that satisfies Condition (1) is called the wind filter.

Let $G_r(\omega)$ be the antenna rate-loop transfer function from the rate input to the encoder servo error. Then the simulated error due to wind gusts is obtained as

$$e_s(\omega) = G_r(\omega)G(\omega)w(\omega) \tag{2}$$

In Eq. (2), the spectrum $w(\omega)$ is constant (independent of frequency), and the transfer function $G_r(\omega)$ is dominated by the antenna resonance frequencies. In this case, the magnitude of the filter transfer function can be assumed to be a smooth curve in the form of a shaped integrator, that is,

$$G(s) = \frac{k}{s} \frac{(T_1 s + 1)^2}{(T_2 s + 1)(T_3 s + 1)^2} \tag{3a}$$

where the time constants are

$$\left.\begin{aligned} T_1 &= \frac{1}{2\pi f_1} \\ T_2 &= \frac{1}{2\pi f_2} \\ T_3 &= \frac{1}{2\pi f_3} \end{aligned}\right\} \tag{3b}$$

and $f_1 = 2.2$ Hz, $f_2 = 7.0$ Hz, and $f_3 = 12.0$ Hz are the frequencies where the magnitude of the integrator $k/s$ is shaped. The frequencies $f_1$ and $f_2$ determine the bandwidth of resonance frequencies of the antenna, and the frequency $f_3$ is the cut-off frequency for the wind disturbances. In this transfer function, the only unknown parameter is the gain, $k$. The plot of $G(\omega)$ is shown in Fig. 6 for $k = 1$. The

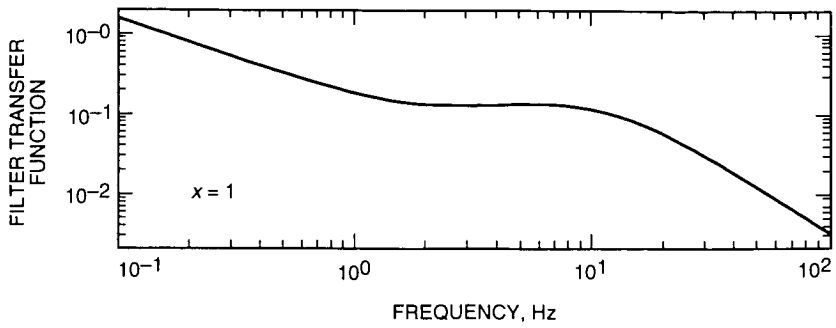


**Fig. 6. Magnitude of the wind filter transfer function for $k = 1$.**

choice of the transfer function shape as in Eq. (3) was done after the investigation of the more general case, where $G(s)$ was a rational function of polynomials of order 5 or less. The performance errors for the polynomials were almost the same as for $G(s)$ in Eq. (3).

The wind model for the azimuth rate loop was determined for two antenna elevation positions: 60 and 11 deg. For each elevation position, the wind from the front, side, and back was considered. The spectra of the azimuth encoder output, measured and simulated, are shown in Fig. 7 for an elevation angle of 60 deg and a wind direction from the back of the antenna. The measured spectrum shows two resonances, at 1.7 and 4.2 Hz, and the spectrum from simulations has an additional resonance peak at 3.1 Hz. The spectra are coincidental at the first three frequencies. The time series of measured and simulated encoder outputs are shown in Figs. 8(a) and 8(b), respectively. They show the similarity; the difference between their standard deviations was less than 7 percent. The gain, $k$, in this case was 0.0095.

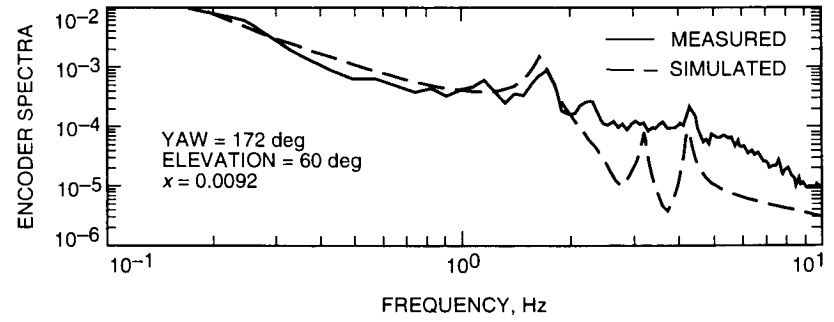


**Fig. 7. Azimuth encoder spectra.**

Similar results were obtained for other cases. Gain $k$ for an 11-deg elevation angle was as follows:

| Front wind | Side wind | Back wind |
|---|---|---|
| 0.0075 | 0.0079 | 0.0075 |

Gain $k$ for a 60-deg elevation angle was as follows:

| Front wind | Side wind | Back wind |
|---|---|---|
| 0.0095 | 0.0096 | 0.0092 |

The tables show that for a given elevation angle the gains for front, back, and side winds are almost the same. Therefore, the wind filter is independent of wind direction; however, it depends on the antenna elevation angle.

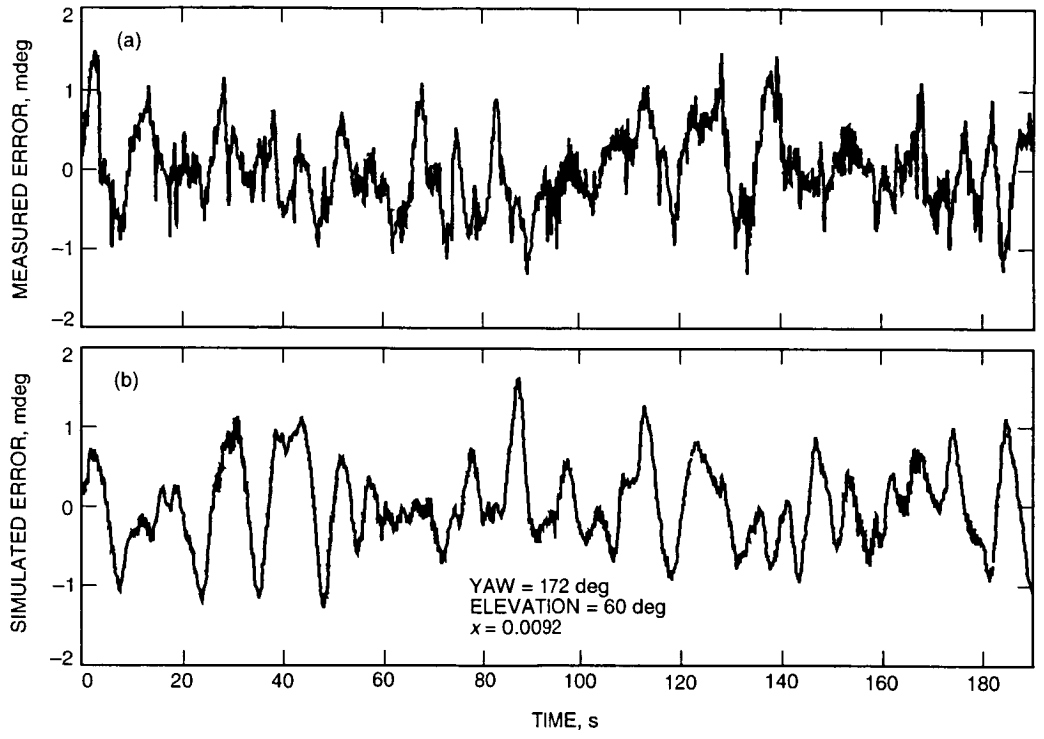


**Fig. 8. Azimuth encoder error due to 32-km/h wind gusts from the back at a 60-deg elevation angle: (a) measured and (b) simulated.**

## IV. Conclusions

The measured wind data at the DSS-13 site were used to verify the analytical wind model. The comparison showed that servo errors from the analytical model fall within the measured servo error boundaries. However, for the front and back winds, the simulated azimuth errors were smaller than the measured ones. This occurred because of the assumption of the spatial coherence of the wind gust model. The coherence caused a symmetric antenna load and, in consequence, a low azimuth servo error. This shortcoming indicates a need for upgrading the analytical wind model so that the spatially incoherent wind gust model reflects the real gusts in a more accurate manner.

The measured wind data were also used to generate a new wind model more suitable for the design of a control system with wind disturbance rejection properties. The wind filter was obtained for the antenna azimuth model for different elevation angles and different wind directions such that the simulated servo error is close to the measured one.

# References


[1] W. Gawronski, B. Bienkiewicz, and R. E. Hill, "Wind-Induced Dynamics of a Deep Space Network Antenna, *Journal of Sound and Vibration*, vol. 174, no. 5, pp. 67–77, 1994.

[2] W. Gawronski and J. A. Mellstrom, "Field Verification of the Wind Tunnel Coefficients," *The Telecommunications and Data Acquisition Progress Report 42-119, July–September 1994*, Jet Propulsion Laboratory, Pasadena, California, pp. 210–220, November 15, 1994, URL http://edms-www.jpl.nasa.gov/tda/progress_report/42-119/119G.pdf.

[3] C. S. Racho and W. Gawronski, "Experimental Modification and Identification of the DSS-13 Antenna Control System," *The Telecommunications and Data Acquisition Progress Report 42-115, July–September 1993*, Jet Propulsion Laboratory, Pasadena, California, pp. 42–53, November 15, 1993.

[4] W. Gawronski, C. S. Racho, and J .A. Mellstrom, "Linear Quadratic Gaussian and Feedforward Controllers for the DSS-13 Antenna," *The Telecommunications and Data Acquisition Progress Report 42-118, April–June 1994*, Jet Propulsion Laboratory, Pasadena, California, pp. 37–55, August 15, 1994, URL http://edms.jpl.nasa.gov/tda/progress_report/42-118/118D.pdf.

[5] E. Simiu and R. H. Scanlan, *Wind Effects on Structures*, New York: Wiley–Interscience, 1978.

TDA Progress Report 42-123 November 15, 1995



# A New Model for Yaw Attitude of Global Positioning System Satellites

Y. E. Bar-Sever
Tracking Systems and Applications Section

*Proper modeling of the Global Positioning System (GPS) satellite yaw attitude is important in high-precision applications. A new model for the GPS satellite yaw attitude is introduced that constitutes a significant improvement over the previously available model in terms of efficiency, flexibility, and portability. The model is described in detail, and implementation issues, including the proper estimation strategy, are addressed. The performance of the new model is analyzed, and an error budget is presented. This is the first self-contained description of the GPS yaw attitude model.*

## I. Introduction

On June 6, 1994, the U.S. Air Force implemented a yaw bias on most Global Positioning System (GPS) satellites. By January 1995, the implementation was extended to all the satellites except SVN 10. The yaw bias was introduced as a way to make modeling of the yaw attitude of the GPS satellites during shadow crossings possible [2]. The yaw attitude of a biased GPS satellite during eclipse seasons is markedly different from the yaw attitude of a noneclipsing satellite or from that of an unbiased satellite. The yaw attitude of the GPS satellite has a profound effect on precise applications. Mismodeling the satellite attitude can cause decimeter-level error in the positioning of ground stations with certain GPS-based techniques and skew media calibrations. This required the development of a special attitude model for biased GPS satellites. In addition to the yaw bias effects, that model also corrected other mismodeling that existed in the old model, namely, that of the "noon turn."

The first attitude model written for the biased constellation was made freely available to the GPS community in the form of a collection of FORTRAN routines [1]. For simplicity, this model is referred to in this article as GYM94 (for GPS Yaw Attitude Model—94). GYM94 was implemented in JPL's GIPSY software and, in various forms, in other high-precision geodetic packages. The model was successfully used within JPL's routine processing of daily GPS orbits and ground station coordinates for the International Global Positioning System Service (IGS). The model had some drawbacks, though. Mainly, it was cumbersome to implement and very demanding of computer resources, namely, memory and central processing unit (CPU) time.

In this article, we describe a new model for the GPS satellite attitude, referred to as GYM95. The model is analytic, in contrast to the numerical nature of GYM94, which required sequential processing in time. A time series of yaw rates estimated by the routine GPS processing at JPL will be analyzed to demonstrate the need to estimate the yaw rates.

## II. Background

The analysis that led to the implementation of the yaw bias on GPS satellites is described in Bar-Sever et al. [2]. A general description of the first yaw attitude model can also be found there. For completeness, we give here a brief summary.

The nominal yaw attitude of a GPS satellite is determined by satisfying two constraints: first, that the navigation antennas point toward the geocenter and, second, that the normal to the solar array surface will be pointing at the Sun. To meet these two conditions, the satellite has to yaw constantly. The resulting yaw attitude algorithm is singular at two points—the intersections of the orbit with the Earth–Sun line. At these points, the yaw attitude is not single-valued, as any yaw angle allows optimal view of the Sun. In the vicinity of these singular points, the yaw rate of the spacecraft, required to keep track of the Sun, is unbounded. This singularity problem was largely ignored prior to the release of GYM94. While this mismodeling problem could be fixed easily through the realization of a finite limit on the spacecraft yaw rate, a bigger problem existed that could only be addressed by changing the attitude control subsystem (ACS) on board the spacecraft. The ACS determines the yaw attitude of the satellite by using a pair of solar sensors mounted on the solar panels. As long as the Sun is visible, the signal from the solar sensors is a true representation of the yaw error. During shadow, in the absence of sunlight, the output from the sensors is essentially zero and the ACS is driven in an open-loop mode by the noise in the system. It turns out that even a small amount of noise can be enough to trigger a yaw maneuver at maximum rate. To make it possible to model the yaw attitude of the GPS satellites, the ACS had to be biased by a small but fixed amount. Biasing the ACS means that the Sun sensor's signal is superposed with another signal (the bias) equivalent to an observed yaw error of 0.5 deg (the smallest bias possible). As a result, during periods when the Sun is observed, the satellite yaw attitude will be about 0.5 deg in error with respect to the nominal orientation. During shadow, this bias dominates the open-loop noise and will yaw the satellite at full rate in the direction of the bias. Upon shadow exit, the yaw attitude of the satellite can be calculated, and the Sun recovery maneuver can also be modeled.

GYM94 accounted for the yaw bias as well as the limit on the yaw rate. It computed the satellite yaw angle through numerical integration of a control law. Its output was a large file containing the yaw attitude history and, optionally, partial derivatives of the yaw attitude with respect to the yaw rate parameter. This file could later be interpolated to retrieve a yaw angle at the requested time. This process required relatively large amounts of computer memory and CPU time. In addition, the model's complex control law—a simulation of the onboard attitude determination algorithm—did not allow much physical insight into the problem and was hard to tune. To overcome all these deficiencies, the GYM95 model was created. GYM94 was used in studies of GPS calibration for the DSN since September 1994, and the design of the new attitude model drew on the experience accumulated with GYM94. GYM95 is simple enough to be described by a small set of formulas, allowing easy implementation in different computing environments. Its analytic nature, as opposed to the numerical nature of GYM94, allows queries at arbitrary time points with great savings in computer resources. Finally, it allows more flexibility in tuning and adapting it to the changing conditions of the GPS constellation.

## III. The New Yaw Attitude Model (GYM95)

### A. Overview

The yaw attitude of a GPS satellite can be divided into four regimes: nominal attitude, shadow crossing, postshadow maneuver, and noon turn. Most of the time (and for noneclipsing satellites all the time), the satellite is in the nominal attitude regime. The postshadow maneuver begins immediately after emerging from the Earth's shadow and lasts until the satellite has regained its nominal attitude. This phase can last from 0 to 40 min. The noon-turn maneuver does not occur until the beta angle goes below about 5 deg and can last between 0 and 40 min.

We will start by defining a few important terms in Table 1 and the notation used, and then describe the yaw attitude during each of the four regimes, including the governing formulas. Finally, we will describe how to tie all the regimes together into one functional model and analyze any built-in errors.

**Table 1. Definition of terms.**

| Term | Definition |
|---|---|
| Orbit midnight | The point on the orbit furthest from the Sun. |
| Orbit noon | The point on the orbit closest to the Sun. |
| Orbit normal | The unit vector along the direction of the satellite's angular momentum, treating the satellite as a point mass (equals position × velocity, where the order of the cross-product is important). |
| Sun vector | The direction from the spacecraft to the Sun. |
| Beta angle | The acute angle between the Sun vector and the orbit plane. It is defined as positive if the Sun vector forms an acute angle with the orbit normal and negative otherwise. |
| Orbit angle | The angle formed between the spacecraft position vector and orbit midnight, growing with the satellite's motion. |
| Yaw origin | A unit vector that completes the spacecraft position vector to form an orthogonal basis for the orbit plane and is in the general direction of the spacecraft velocity vector. |
| Spacecraft-fixed z-axis | The direction of the GPS navigation antennas. |
| Nominal spacecraft-fixed x-axis | A unit vector orthogonal to the spacecraft-fixed z-axis such that it lies in the Earth–spacecraft–Sun plane and points in the general direction of the Sun (note that this definition is not single valued when the Earth, spacecraft, and Sun are collinear). |
| Spacecraft-fixed x-axis | A spacecraft-fixed vector, rotating with the spacecraft, such that far enough from orbit noon and orbit midnight, it coincides with the nominal spacecraft-fixed x-axis. Elsewhere, it is a rotation of the nominal spacecraft-fixed x-axis around the spacecraft-fixed z-axis. |
| Nominal yaw angle | The angle between the nominal spacecraft-fixed x-axis and the yaw-origin direction, restricted to be in [−180,180]. It is defined to have a sign opposite to that of the beta angle. |
| Yaw angle | The angle between the spacecraft-fixed x-axis and the yaw-origin direction, restricted to be in [−180,180], also termed "actual yaw angle." |
| Yaw error | The difference between the yaw angle and the nominal yaw angle, restricted to be in [−180,180]. |
| Midnight turn | The yaw maneuver the spacecraft is conducting from shadow entry until it resumes nominal attitude sometime after shadow exit. |
| Noon turn | The yaw maneuver the spacecraft is conducting in the vicinity of orbit noon when the nominal yaw rate would be higher than the yaw rate the spacecraft is able to maintain. It ends when the spacecraft resumes nominal attitude. |
| Spin-up/down time | The time it takes for the spacecraft to spin up or down to its maximal yaw rate. The spacecraft is spinning down when it has to reverse its yaw rate. |

The notation used is as follows:

$\mu$ = orbit angle

$\beta$ = beta angle

$E$ = Earth-spacecraft-Sun angle

$b$ = yaw bias inserted in the satellite ACS

$B$ = actual yaw angle induced by $b$

$\Psi$ = actual yaw angle

$\Psi_n$ = nominal yaw angle

$t$ = current time, s

$t_i$ = time of shadow entry

$t_e$ = time of shadow exit

$t_n$ = start time of the noon-turn maneuver

$t_1$ = spin-up/-down time

$\Psi_i$ = yaw angle upon shadow entry

$\Psi_e$ = yaw angle upon shadow exit

$R$ = maximal yaw rate of the satellite

$RR$ = maximal yaw-rate rate of the satellite

Angle units, i.e., radians or degrees, will be implied by context. Radians will usually be used in formulas, and degrees will usually be used in the text. FORTRAN function names are used whenever possible with the implied FORTRAN functionality, e.g., ATAN2(a,b) is used to denote arc-tangent($a/b$) with the usual FORTRAN sign convention.

### B. The Nominal Attitude Regime

The realization of the two requirements for satellite orientation mentioned above yields the following formula for the nominal yaw angle:

$$\Psi_n = \mathrm{ATAN2}(-\mathrm{TAN}(\beta), \mathrm{SIN}(\mu)) + B(b, \beta, \mu) \tag{1}$$

where $\beta$ is the beta angle, $\mu$ is the orbit angle, measured from orbit midnight in the direction of motion, and $B$ is the yaw bias (see below). It follows from this formula that the sign of the yaw angle is always opposite that of the beta angle.

Ignoring the time variation of the slow-changing beta angle leads to the following formula for the yaw rate (there are simpler formulas, but they contain removable singularities that are undesirable for computer codes):

$$\dot{\Psi}_n = \mathrm{TAN}(\beta) \times \mathrm{COS}(\mu) \times \frac{\dot{\mu}}{\mathrm{SIN}(\mu)^2 + \mathrm{TAN}(\beta)^2} + \dot{B}(b, \beta, \mu) \tag{2}$$

where $\dot{\mu}$ varies little in time and can safely be replaced by 0.0083 deg/s. Notice that the sign of the nominal yaw rate is the same as the sign of the beta angle in the vicinity of orbit midnight ($\mu = 0$).

The singularity of these two formulas when $\beta = 0$ and $\mu = 0, 180$ is genuine and cannot be removed.

### C. The Yaw Bias

Like any medicine, the yaw bias has its side effects. Outside shadow, it introduces yaw errors that are actually larger that 0.5 deg. To fully understand this, we have to describe the ACS hardware, which

is beyond the scope of this article. The underlying reason is that the output of the solar sensor is proportional not to the yaw error but to its sine, and it is also proportional to the sine of the Earth–spacecraft–Sun angle, $E$. So, in order to offset a bias of $b$ deg inserted in the ACS, the satellite has to actually yaw $B$ deg, where $B$ is given by:

$$B(b,\beta,\mu) = B(b,E) = \mathrm{ASIN}\frac{0.0175 \times b}{\mathrm{SIN}(E)} \tag{3}$$

The hardware-dependent proportionality factor is 0.0175, and the Earth–spacecraft–Sun angle, $E$, the beta angle, $\beta$, and the orbit angle, $\mu$, satisfy the following approximate relationship:

$$\mathrm{COS}(E) = \mathrm{COS}(\beta) \times \mathrm{COS}(\mu) \tag{4}$$

and $E$ is restricted to [0,180]. Equation (3) becomes singular for $E$ less then 0.5013 deg. This has no effect on the actual yaw because a small value of $E$ implies that the spacecraft is in the middle of a midnight turn or a noon turn and is already yawing at full rate. The value of $B$ does have a significant effect, though, on the timing of noon-turn entry and on the yaw angle shortly before that. For example, for $E = 5$ deg, which is the typical threshold value for noon-turn entry, the actual yaw bias is $B \approx 6$ deg.

The bias rate, $\dot{B}$, is given by

$$\dot{B}(b,\beta,\mu) = -0.0175 \times b \times \mathrm{COS}(E) \times \mathrm{COS}(\beta) \times \mathrm{SIN}(\mu) \times \frac{\dot{\mu}}{\mathrm{COS}(B) \times \mathrm{SIN}(E)^3} \tag{5}$$

The ACS bias, $b$, can be $\pm 0.5$ deg or 0 deg. With few exceptions, to be discussed below, the bias is always set to $b = -\mathrm{SIGN}(0.5, \beta)$ since this selection was found to expedite the Sun recovery time after shadow exit.

### D. The Shadow-Crossing Regime

As soon as the Sun disappears from view, the yaw bias alone is steering the satellite. On most satellites, the yaw bias has a sign opposite to that of the beta angle. To correct for the bias-induced error, the satellite has to reverse its yaw rate upon shadow entry. For those satellites with bias of equal sign to that of the beta angle, there is no yaw reversal. The bias is large enough to cause the satellite to yaw at full rate until shadow exit, when the bias can be finally compensated. The yaw angle during shadow crossing depends, therefore, on three parameters: the yaw angle upon shadow entry, $\Psi_i$, the yaw rate upon shadow entry, $\dot{\Psi}_i$, and the maximal yaw rate, $R$. Let $t_i$ be the time of shadow entry and let $t$ be the current time, and define

$$t_1 = \frac{\mathrm{SIGN}(R,b) - \dot{\Psi}_i}{\mathrm{SIGN}(RR,b)} \tag{6}$$

to be the spin-up/-down time. Then the yaw angle during shadow crossing is given by

$$\Psi = \begin{cases} \Psi_i + \dot{\Psi}_i \times (t - t_i) + 0.5 \times \mathrm{SIGN}(RR,b) \times (t - t_i)^2 & t < t_i + t_1 \\ \Psi_i + \dot{\Psi}_i \times t_1 + 0.5 \times \mathrm{SIGN}(RR,b) \times t_1^2 + \mathrm{SIGN}(R,b) \times (t - t_i - t_1) & \text{else} \end{cases} \tag{7}$$

Using this formula, we avoid the singularity problem of the nominal attitude at midnight.

### E. The Postshadow Maneuver

This is the trickiest part of the yaw attitude model. The postshadow maneuver depends critically upon the yaw angle at shadow exit. The ACS is designed to reacquire the Sun in the fastest way possible. Upon shadow exit, the ACS has two options: One is to continue yawing at the same rate until the nominal attitude is resumed; the second is to reverse the yaw rate and yaw at full rate until the nominal attitude is resumed. In this model, we assume that the decision is based on the difference between the actual yaw angle and the nominal yaw angle upon shadow exit, and we denote this difference by $D$. If $t_e$ is the shadow-exit time, then

$$D = \Psi_n(t_e) - \Psi(t_e) - \mathrm{NINT}\left(\frac{\Psi_n(t_e) - \Psi(t_e)}{360}\right) \times 360 \tag{8}$$

and the yaw rate during the postshadow maneuver will be $\mathrm{SIGN}(R, D)$.

Given the yaw angle upon shadow exit, the yaw rate upon shadow exit, $\mathrm{SIGN}(R, b)$, and the yaw rate during the postshadow maneuver, we can compute the actual yaw angle during the postshadow maneuver by using Eq. (7) with the appropriate substitutions. This yields

$$t_1 = \frac{\mathrm{SIGN}(R, D) - \mathrm{SIGN}(R, b)}{\mathrm{SIGN}(RR, D)} \tag{9}$$

$$\Psi = \begin{cases} \Psi(t_e) + \mathrm{SIGN}(R, b) \times (t - t_e) + 0.5 \times \mathrm{SIGN}(RR, D) \times (t - t_e)^2 & t < t_e + t_1 \\ \Psi(t_e) + \mathrm{SIGN}(R, b) \times t_1 + 0.5 \times \mathrm{SIGN}(RR, D) \times t_1^2 + \mathrm{SIGN}(R, D) \times (t - t_e - t_1) & \text{else} \end{cases} \tag{10}$$

The postshadow maneuver ends when the actual yaw attitude, derived from Eq. (10), becomes equal to the nominal yaw attitude. The time of this occurrence is computed in GYM95 by an iterative process that brackets the root of the equation $\Psi(t) = \Psi_n(t)$, where the time dependence of $\Psi_n(t)$ is introduced by substituting $\mu = \mu_e + 0.0083 \times (t - t_e)$ in Eq. (1). This equation can be solved as soon as the satellite emerges from shadow. Once the time of resuming nominal yaw is reached, the satellite switches back to that regime.

### F. The Noon-Turn Regime

The noon-turn regime starts in the vicinity of orbit noon, when the nominal yaw rate reaches its maximal allowed value, and ends when the actual yaw attitude catches up with the nominal regime. First, we have to identify the starting point, and this can be done by finding the root, $t_n$, of the equation $\dot{\Psi}_n(t) = -\mathrm{SIGN}(R, \beta)$, where $\dot{\Psi}_n(t)$ is the nominal yaw rate from Eq. (2). After the start of the noon turn, the yaw angle is governed by Eq. (7), again with the proper substitutions. This yields

$$\Psi = \Psi_n(t_n) - \mathrm{SIGN}(R, \beta) \times (t - t_n) \tag{11}$$

The end time is found by the same procedure that is used to find the end time of the postshadow maneuver.

### G. The Complete Model

Satellite position and velocity, as well as the timing of shadow crossings, are required inputs to GYM95. The model is able to bootstrap, though, if these input values are unavailable far enough into the past. For example, if the satellite is potentially in the postshadow regime upon first query, there is a need to know the shadow-entry time so that all the inputs to Eqs. (9) and (10) be known. If this shadow-entry

time is missing from the input, the model can compute it approximately, as well as the shadow-exit time. Once all the timing information is available, yaw angle queries can be made at arbitrary time points. The model will decide the relevant yaw regime and compute the yaw angle using the correct formula. Given the above formulas, it is an easy matter to compute the partials of the yaw angle with respect to any parameter of the problem, the most important of which is the maximal yaw rate, $R$.

### H. Model Fidelity

The fidelity of the model is a measure of how accurately it describes the true behavior of the satellite. This is hard to measure because there is no high-quality telemetry from the satellite and because the estimated value of the main model parameter, namely, the yaw rate, depends on many other factors beside the attitude model itself: data, estimation strategy, and other models for the orbit and the radiometric measurements. Nevertheless, based on the experience accumulated thus far with this model and its predecessor, GYM94, it is possible to come up with an educated guess of the inaccuracy of GYM95.

The nominal attitude regime is believed to be very accurate. The only source of error is mispointing of the satellite, which is poorly understood and relatively small (of the order of 1 deg around the pitch, yaw, and roll axes). Compensations for the dynamic effect of this error source are discussed in [3] and [4], where they are treated, properly, within the context of the solar pressure model.

Modeling the midnight turn accurately is difficult. Inherent uncertainties like the exact shadow-entry and -exit times are a constant error source. Inaccuracies in shadow-entry time are more important than inaccuracies in shadow-exit time because errors in the former are propagated by the model throughout the midnight-turn maneuver. In contrast, error in the shadow-exit time will affect the postshadow maneuver only. Either way, the inaccuracy will be manifested through a constant error in the yaw angle, something that can be partially compensated through the estimation of the yaw rate. The length of the penumbra region is usually about 60 s. Sometime during this period, the yaw bias kicks in. GYM95 puts that time midway into penumbra. The maximum timing error is, therefore, less than 30 s. A worst-case scenario, ignoring the short spin-up/-down period and using a yaw rate of 0.13 deg/s, will give rise to a constant yaw error of $30 \times 0.13 \approx 4$ deg throughout the midnight turn. A more realistic estimate is 3 deg, even before applying yaw rate compensation, after which the rms error will remain the same, but the mean is expected to vanish. Another error source is the uncertainty in the value of the maximal yaw-rate rate, $RR$. This parameter is weakly observable and, therefore, hard to estimate. The nominal value used in GYM95 is 0.00165 deg/s$^2$ for Block IIA satellites and 0.0018 deg/s$^2$ for Block II satellites, and it should be at least 70-percent accurate. The long-term effects of a yaw-rate rate error can be computed from the second part of Eq. (7) as

$$\Psi(RR) = \frac{\dot{\Psi}_i \times \mathrm{SIGN}(R,b) - 0.5 \times \dot{\Psi}_i^2 - 0.5 \times \mathrm{SIGN}(R,b)^2}{\mathrm{sign}(RR,b)}$$

A worst-case scenario assuming $\dot{\Psi}_i = -\mathrm{SIGN}(R,b) = 0.13$ and 30-percent error in the yaw-rate rate would give rise to a yaw error of about 5 deg. These assumptions also imply a very short shadow duration so the error will not be long lasting. For long shadow events, $\dot{\Psi}_i \approx 0$, and the resulting yaw error is about 1 deg. Again, this error can be partially offset by estimating the yaw rate.

The main error source for the noon turn is the timing uncertainty of the onset of the maneuver. This uncertainty is not expected to be larger than 2 min. A 2-min error will cause a constant yaw error of about 15 deg, assuming a yaw rate of 0.13 deg/s. The relatively short duration of the noon turn diminishes somewhat the effects of such a large error. Estimating the yaw rate will decrease the error further.

The value of the yaw rate is not considered here as an error source. Any nominal value stands to be at least 10 percent in error (see below). Since errors due to yaw rate grow in time, this parameter must

be estimated or, alternatively, a previously estimated value should be used. For example, an error of 0.01 deg/s in the yaw rate will give rise to a 30-deg error in yaw at the end of a 50-min shadow event.

Although unlikely, errors from different sources can add up. In that case, the maximal error for each regime is as follows: 2 deg for the nominal yaw regime, 9 deg for the midnight-turn regime, and 15 deg for the noon-turn regime. Typical errors are expected to be less than half of these values.

## IV. The Estimated Yaw Rates

As part of the implementation of the GYM models at JPL, the yaw rates of all eclipsing satellites are estimated for every midnight turn and every noon turn. In JPL's GIPSY software, this is done by treating the yaw rate as a piece-wise constant parameter for each satellite. The parameter value is allowed to change twice per revolution, midway between noon and midnight. Since a small error in the yaw rate can cause a large yaw error over time and, since our a priori knowledge of the yaw rate is not sufficiently accurate, we found it necessary to iterate on the yaw rate value. JPL routinely publishes the final estimates for the yaw rates as daily text files. Unfortunately, due to a software bug, the archived yaw rates for dates prior to February 16, 1995, were in error. This leaves a period of about 2 months when the estimated yaw rates are available. Figure 1 depicts the estimated yaw rates for each eclipsing satellite, for each midnight turn, and for each noon turn, from February 16 to April 26, 1995. The accuracy of the estimates depends on the amount of data available during each maneuver and this, in turn, is proportional to the duration of the maneuver. The longer the maneuver, the better the estimate. The effect of a reduced estimation accuracy during short maneuvers is mitigated by the fact that the resulting yaw error is also proportional to the duration of the maneuver. For long maneuvers, e.g., a midnight turn at the middle of the eclipse season, the estimates are good to 0.002 deg/s, which leads to a maximal yaw error of about 6 deg. A similar error level is expected for short maneuvers. Noon turns occur only during the middle part of the eclipse season. In Fig. 1, they can be distinguished from midnight-turn rates by the larger formal error associated with them, since they are typically short events of 15- to 30-min duration. As a result, the scatter of the noon-turn rates is larger than that of the midnight-turn rates. Toward the edges of the eclipse season, the quality of the yaw rate estimates drops, again because of the short duration of the shadow events. The most striking feature in Fig. 1 is the discontinuity of the estimated yaw rates in the middle of the eclipse season, corresponding to the beta angle crossing zero. No plausible explanation is currently available for this jump. SVN 29 is the only satellite that does not have a jump discontinuity; this is also the only satellite that does not undergo a bias switch in the middle of the eclipse season. SVN 31 is the only satellite with a jump from high yaw rates to low yaw rates as the beta angle transitions from positive to negative. There is nothing otherwise special about SVN 31. The ratio of the high yaw-rate values to the low yaw-rate values is about 1.3 for all satellites.

Within each half of the eclipse season, the midnight yaw rates are fairly constant, varying by 10 percent or less. The noon-turn yaw rates seem to be more variable. This is not only a consequence of the weak observability but also of the fact that the spacecraft is subject to a varying level of external torque during the noon turn as the eclipse season progresses.

The modeling of the postshadow maneuver is a problem for which a satisfactory solution has not yet been found. The source of the problem is that the presence of the postshadow regime makes the estimation of the yaw rate into a nonlinear problem. There is always a critical value of the yaw rate such that, for higher values, the spacecraft will reverse its yaw upon shadow exit and, for lower values, the spacecraft will retain its yaw rate until the end of the midnight turn. If this critical value falls in the range of feasible yaw rates—which it often does—it becomes very hard to figure out what kind of maneuver the satellite is doing upon shadow exit. To avoid this postshadow ambiguity, we have been rejecting measurement data from shadow exit until about 30 min thereafter.

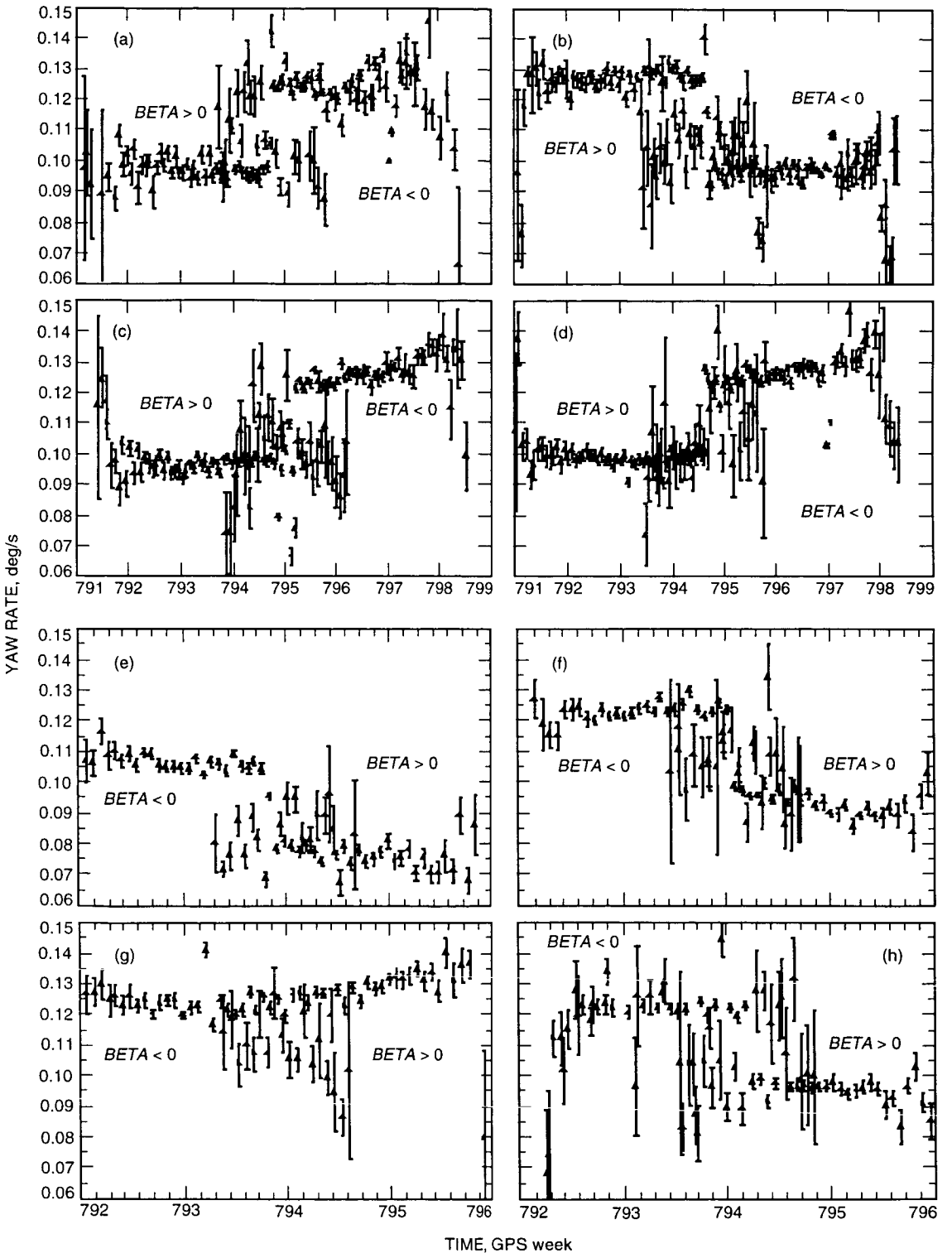


**Fig. 1. Estimated yaw rates with their formal errors versus GPS week for coplanar (C-plane) (a) SVN 28, (b) SVN 31, (c) SVN 36, and (d) SVN 37 and for F-plane (e) SVN 18, (f) SVN 26, (g) SVN 29, and (h) SVN 32.**

## Acknowledgments

The help and cooperation of the U.S. Air Force 2SOPS at Falcon Air Force Base throughout this project was essential to its success and is greatly appreciated. Thanks are also due to the IGS processing team at JPL for its help in producing some of the results reported here.

## References

[1] Y. Bar-Sever, "Improvement to the GPS Attitude Control Subsystem Enables Predictable Attitude During Eclipse Seasons," *International Global Positioning System Service*, mail no. 0591, May 1994.

[2] Y. Bar-Sever, J. Anselmi, W. Bertiger, and E. Davis, "Fixing the GPS Bad Attitude: Modeling GPS Satellite Yaw During Eclipse Seasons," *Proceedings of the ION National Technical Meeting*, Anaheim, California, pp. 827–834, January 1995.

[3] G. Beutler, E. Brockmann, W. Gurtner, U. Hugentobler, L. Mervart, M. Rothacher, and A. Verdun, "Extended Orbit Modeling Technique at the CODE Processing Center of the International GPS Service for Geodynamics (IGS): Theory and Initial Results," *Manuscripta Geodaetica*, vol. 19, no. 6, pp. 367–385, 1994.

[4] D. Kuang, H. J. Rim, B. E. Schutz, and P. A. M. Abusali, "Modeling GPS Satellite Attitude Variation for Precise Orbit Determination," to appear in *Manuscripta Geodaetica.*

TDA Progress Report 42-123 November 15, 1995

# A Light-Induced Microwave Oscillator

X. S. Yao and L. Maleki
Tracking Systems and Applications Section

*We describe a novel oscillator that converts continuous light energy into stable and spectrally pure microwave signals. This light-induced microwave oscillator (LIMO) consists of a pump laser and a feedback circuit, including an intensity modulator, an optical fiber delay line, a photodetector, an amplifier, and a filter. We develop a quasilinear theory and obtain expressions for the threshold condition, the amplitude, the frequency, the line width, and the spectral power density of the oscillation. We also present experimental data to compare with the theoretical results. Our findings indicate that the LIMO can generate ultrastable, spectrally pure microwave reference signals up to 75 GHz with a phase noise lower than −140 dBc/Hz at 10 kHz.*

## I. Introduction

Oscillators are devices that convert energy from a continuous source to a periodically varying signal. They represent the physical realization of a fundamental basis of all physics, the harmonic oscillator, and they are perhaps the most widely used devices in modern day society. Today a variety of oscillators—mechanical [1] (such as the pendulum), electromagnetic (such as LC circuit [2,3] and cavity-based [4]), and atomic (such as masers [5] and lasers [6])—provides a diverse range in the approximation to the realization of the ideal harmonic oscillator. The degree of the spectral purity and stability of the output signal of the oscillator is the measure of the accuracy of this approximation and is fundamentally dependent on the energy storage ability of the oscillator, determined by the resistive loss (generally frequency dependent) of the various elements in the oscillator.

An important type of oscillator widely used today is the electronic oscillator. The first such oscillator was invented by L. De Forest [2] in 1912, shortly after the development of the vacuum tube. In this triode-based device known as the van der Pol oscillator [3], the flux of electrons emitted by the cathode and flowing to the anode is modulated by the potential on the intervening grid. This potential is derived from the feedback of the current in the anode circuit containing an energy storage element (i.e., the frequency-selecting LC filter) to the grid, as shown in Fig. 1(a). Today the solid-state counterparts of these valve oscillators based on transistors are pervasive in virtually every application of electronic devices, instruments, and systems. Despite their widespread use, electronic oscillators, whether of the vacuum-tube or the solid-state variety, are relatively noisy and lack adequate stability for applications where very high stability and spectral purity are required. The limitation to the performance of electronic oscillators is due to ohmic and dispersive losses in various elements in the oscillator, including the LC resonant circuit.

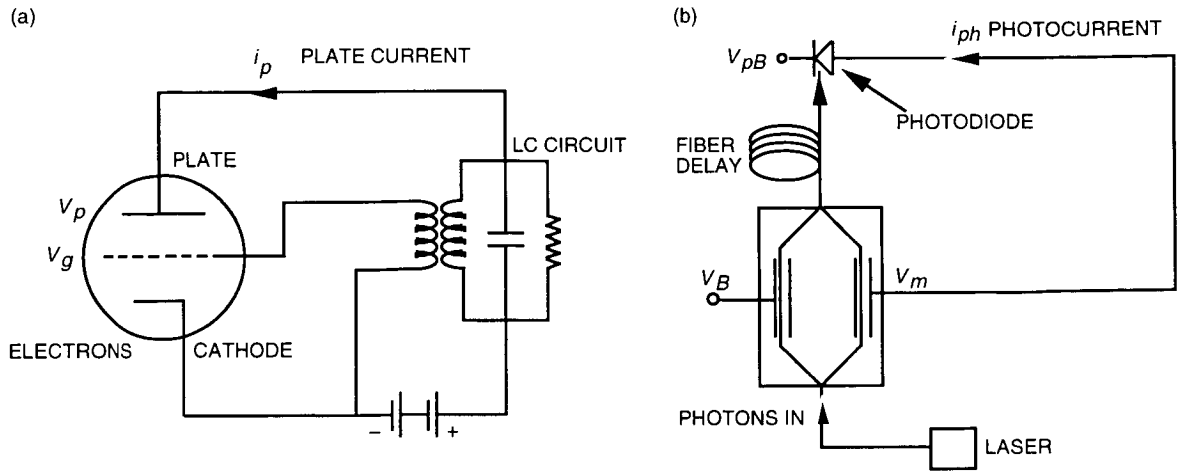


**Fig. 1. Comparison of two types of oscillators: (a) van der Pol oscillator and (b) light-induced microwave oscillator.**

For approximately the past 50 years, the practice of reducing the noise in the electronic oscillator by combining it with a high-quality factor (Q) resonator has been followed to achieve improved stability and spectral purity. The Q is a figure of merit for the resonator given by $Q = 2\pi f \tau_d$, where $\tau_d$ is the energy decay time that measures the energy storage ability of the resonator and $f$ is the resonant frequency. High-Q resonators used for stabilization of the electronic oscillator include mechanical resonators, such as quartz crystals [7,8]; electromagnetic resonators, such as dielectric cavities [9]; and acoustic [10] and electrical delay lines, where the delay time is equivalent to the energy decay time, $\tau_d$, and determines the achievable Q. This combination with a resonator results in hybrid-type oscillators referred to as electromechanical, electromagnetic, or electro-acoustic, depending on the particular resonator used with the oscillator circuit. The choice of the particular resonator is generally determined by a variety of factors, but for the highest achievable Q's at room temperatures, the crystal quartz is the resonator of choice for the stabilization of the electronic oscillator. However, because quartz resonators have only a few high-Q resonant modes at low frequencies [7,8], they have a limited range of frequency tunability and cannot be used to directly generate high-frequency signals.

In this article, we introduce a novel photonic oscillator [11,12] characterized by spectral purity and frequency stability rivaling the best crystal oscillators. This oscillator, shown schematically in Fig. 1(b), is based on converting the continuous light energy from a pump laser to radio frequency (RF) and microwave signals, and thus we refer to it with the acronym LIMO, for the light-induced microwave oscillator. The LIMO is fundamentally similar to the van der Pol oscillator, with photons replacing the function of electrons, an electro-optic (E/O) modulator replacing the function of the grid, and a photodetector replacing the function of the anode. The energy storage function of the LC circuit in the van der Pol oscillator is replaced with a long fiber-optic delay line in the LIMO.

Despite this close similarity, the LIMO is characterized by significantly lower noise and very high stability, as well as by other functional characteristics that are not achieved with the electronic oscillator. The superior performance of the LIMO results from the use of electro-optic and photonic components that are generally characterized by high efficiency, high speed, and low dispersion in the microwave frequency regime. Specifically, currently there are photodetectors available with quantum efficiency as high as 90 percent that can respond to signals with frequencies as high as 110 GHz [13]. Similarly, E/O modulators with a 75-GHz frequency response are also available [14]. Finally, the commercially available optical fiber, which has a small loss of 0.2 dB/km for 1550 nm light, allows long storage time of the optical energy with negligible dispersive loss (loss dependent on frequency) for the intensity modulations at microwave frequencies.

The LIMO may also be considered as a hybrid oscillator in so far as its operation involves both light energy and microwave signals. Nevertheless, as a hybrid oscillator, the LIMO is unique in that its output may be obtained both directly as a microwave signal or as intensity modulation of a optical carrier. This property of the LIMO is quite important for applications involving optical elements, devices, or systems [11].

The ring configuration, consisting of an electro-optic modulator that is fed back with a signal from the detected light at its output, has been previously studied by a number of investigators interested in the nonlinear dynamics of bistable optical devices [15–19]. The use of this configuration as a possible oscillator was first suggested by A. Neyer and E. Voges [20]. Their investigations, however, were primarily focused on the nonlinear regime and the chaotic dynamics of the oscillator. This same interest persisted in the work of T. Aida and P. Davis [21], who used a fiber wave guide as a delay line in the loop. Our studies, by contrast, are specifically focused on the stable oscillation dynamics and the noise properties of the oscillator. The sustainable quasilinear dynamics, both in our theoretical and experimental demonstrations, are arrived at by the inclusion of a filter in the feedback loop to eliminate harmonics generated by the nonlinear response of the E/O modulator. This approach yields stable, low-noise oscillations that closely support the analytical formulation presented here.

In this article, we first describe the oscillator and identify the physical basis for its operation. We then develop a quasilinear theory for the oscillator dynamics and the oscillator noise. Results of the theory are then compared with experimental results.

## II. Description of the Oscillator

The LIMO utilizes the transmission characteristics of a modulator together with a fiber-optic delay line to convert light energy into stable, spectrally pure RF/microwave reference signals. A detailed view of the construction of the oscillator is shown schematically in Fig. 2. In this depiction, light from a laser is introduced into an E/O modulator, the output of which is passed through a long optical fiber, and detected with a photodetector. The output of the photodetector is amplified and filtered and fed back to the electric port of the modulator. This configuration supports self-sustained oscillations at a frequency determined by the fiber delay length, bias setting of the modulator, and the bandpass characteristics of the filter. It also provides for both electrical and optical outputs, a feature which would be of considerable advantage to photonics applications.

We use a regenerative feedback model to analyze the spectral properties of the LIMO. Similar methods have been successfully used to analyze lasers [5] and surface acoustic wave oscillators [22]. The conditions for self sustained oscillations include coherent addition of partial waves each way around the loop and a loop gain exceeding losses for the circulating waves in the loop. The first condition implies that all signals that differ in phase by some multiple of $2\pi$ from the fundamental signal may be sustained. Thus, the oscillation frequency is limited only by the characteristic frequency response of the modulator and the setting of the filter, which eliminates all other sustainable oscillations. The second condition implies that with adequate light input power, self sustained oscillations may be obtained, without the need for the RF/microwave amplifier in the loop. These characteristics, which are expected based on the qualitative analysis of the oscillator dynamics, are mathematically derived in the following sections.

## III. Quasilinear Theory of the LIMO

In the following sections, we introduce a quasilinear theory to study the dynamics and noise of a LIMO. In the discussion, we assume that the E/O modulator in the oscillator is of the Mach–Zehnder type. However, the analysis of oscillators made with different E/O modulators may follow the same procedure. The flow of the theory is as follows: First, the open-loop characteristics of a photonic link consisting of a laser, a modulator, a fiber delay, and a photodetector are determined. We then close the

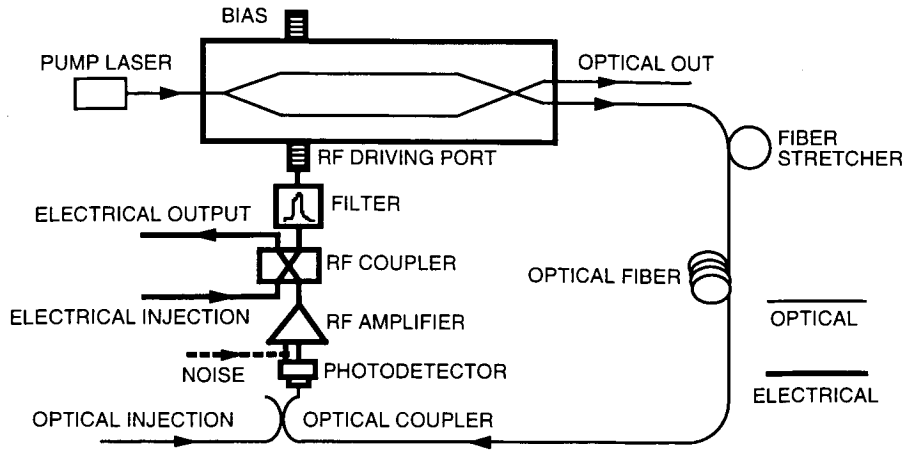


**Fig. 2. Detailed construction of a LIMO. Optical injection and RF injection ports are supplied for synchronizing the oscillator with an external reference by either optical injection locking or electrical injection locking [12]. The bias port and the fiber stretcher can be used to fine tune the oscillation frequency [12]. Noise in the oscillator can be viewed as being injected from the input of the amplifier.**

loop back into the modulator and invoke a quasilinear analysis by including a filter in the loop. This approach leads us to a formulation for the amplitude and the frequency of the oscillation. In the next step, we consider the influence of the noise in the oscillator, again assisted by the presence of the filter, which limits the number of circulating Fourier components. We finally arrive at an expression for the spectral density of the LIMO that would be suitable for experimental investigations.

### A. Oscillation Threshold

The optical power from the E/O modulator's output port that forms the loop is related to an applied voltage $V_{in}(t)$ by

$$P(t) = \left(\frac{\alpha P_o}{2}\right)\left\{1 - \eta \sin \pi \left[\frac{V_{in}(t)}{V_\pi} + \frac{V_B}{V_\pi}\right]\right\} \tag{1}$$

where $\alpha$ is the fractional insertion loss of the modulator, $V_\pi$ is its half-wave voltage, $V_B$ is its bias voltage, $P_o$ is the input optical power, and $\eta$ determines the extinction ratio of the modulator by $(1+\eta)/(1-\eta)$.

If the optical signal $P(t)$ is converted to an electric signal by a photodetector, the output electric signal after an RF amplifier is

$$V_{out}(t) = \rho P(t) R G_A = V_{ph}\left\{1 - \eta \sin \pi \left[\frac{V_{in}(t)}{V_\pi} + \frac{V_B}{V_\pi}\right]\right\} \tag{2}$$

where $\rho$ is the responsivity of the detector, $R$ is the load impedance of the photodetector, $G_A$ is the amplifier's voltage gain, and $V_{ph}$ is the photovoltage, defined as

$$V_{ph} = \left(\frac{\alpha P_o \rho}{2}\right) R G_A = I_{ph} R G_A \tag{3}$$

with $I_{ph} \equiv \alpha P_o \rho / 2$ as the photocurrent. The LIMO is formed by feeding the signal of Eq. (2) back to the RF input port of the E/O modulator. Therefore, the small signal open-loop gain, $G_S$, of the LIMO is

$$G_S \equiv \left.\frac{dV_{out}}{dV_{in}}\right|_{V_{in}=0} = -\frac{\eta\pi V_{ph}}{V_\pi}\cos\left(\frac{\pi V_B}{V_\pi}\right) \tag{4}$$

The highest small signal gain is obtained when the modulator is biased at quadrature, that is, when $V_B = 0$ or $V_\pi$. From Eq. (4), one may see that $G_S$ can be either positive or negative, depending on the bias voltage. The modulator is said to be positively biased if $G_S > 0$; otherwise it is negatively biased. Therefore, when $V_B = 0$, the modulator is biased at negative quadrature, while when $V_B = V_\pi$, the modulator is biased at positive quadrature. Note that in most externally modulated photonic links the E/O modulators can be biased at either positive or negative quadrature without affecting their performance. However, as will be seen next, the biasing polarity will have an important effect on the operation of the LIMO.

In order for the LIMO to oscillate, the magnitude of the small signal open-loop gain must be larger than unity. From Eq. (4), we immediately obtain the oscillation threshold of the LIMO to be

$$V_{ph} = \frac{V_\pi}{\pi\eta|\cos(\pi V_B/V_\pi)|} \tag{5}$$

For the ideal case in which $\eta = 1$ and $V_B = 0$ or $V_\pi$, Eq. (5) becomes

$$V_{ph} = \frac{V_\pi}{\pi} \tag{6}$$

It is important to notice from Eqs. (3) and (6) that the amplifier in the loop is not a necessary condition for oscillation. So long as $I_{ph}R \geq V_\pi/\pi$ is satisfied, no amplifier is needed ($G_A = 1$). It is the optical power from the pump laser that actually supplies the necessary energy for the photonic oscillator. This property is of practical significance because it enables the LIMO to be powered remotely using an optical fiber. Perhaps more significantly, however, the elimination of the amplifier in the loop also eliminates the amplifier noise, resulting in a more stable oscillator. For a modulator with a $V_\pi$ of 3.14 V and an impedance, $R$, of 50 Ω, a photocurrent of 20 mA is required for sustaining the photonic oscillation without an amplifier. This corresponds to an optical power of 25 mW, assuming the responsivity, $\rho$, of the photodetector to be 0.8 A/W.

### B. Linearization Of the E/O Modulator's Response Function

In general, Eq. (2) is nonlinear. If the electrical input signal $V_{in}(t)$ to the modulator is a sinusoidal wave with an angular frequency of $\omega$, an amplitude of $V_o$, and an initial phase of $\beta$,

$$V_{in}(t) = V_o\sin(\omega t + \beta) \tag{7}$$

then the output at the photodetector, $V_{out}(t)$, can be obtained by substituting Eq. (7) in Eq. (2) and expanding the left-hand side of Eq. (2) with Bessel functions:

$$V_{out}(t) = V_{ph}\left\{1 - \eta\sin\left(\frac{\pi V_B}{V_\pi}\right)\left[J_o\left(\frac{\pi V_o}{V_\pi}\right) + 2\sum_{m=1}^{\infty} J_{2m}\left(\frac{\pi V_o}{V_\pi}\right)\cos(2m\omega t + 2m\beta)\right]\right.$$

$$\left. - 2\eta\cos\left(\frac{\pi V_B}{V_\pi}\right)\sum_{m=0}^{\infty} J_{2m+1}\left(\frac{\pi V_o}{V_\pi}\right)\sin\left[(2m+1)\omega t + (2m+1)\beta\right]\right\} \tag{8}$$

It is clear from Eq. (8) that the output contains many harmonic components of $\omega$.

The output can be linearized if it passes through an RF filter with a bandwidth sufficiently narrow to block all harmonic components. The linearized output can be obtained easily from Eq. (8):

$$V_{out}(t) = G(V_o)V_{in}(t) \tag{9}$$

where the voltage gain coefficient, $G(V_0)$, is defined as

$$G(V_o) = G_s \frac{2V_\pi}{\pi V_o} J_1\left(\frac{\pi V_o}{V_\pi}\right) \tag{10a}$$

It can be seen that the voltage gain, $G(V_o)$, is a nonlinear function of the input amplitude, $V_o$, and its magnitude decreases monotonically with $V_o$. However, for a small enough input signal ($V_o << V_\pi$ and $J_1(\pi V_o/V_\pi) = \pi V_o/2V_\pi$), we can recover from Eq. (10a) the small signal gain: $G(V_o) = G_S$.

If we expand the left-hand side of Eq. (2) with Taylor series, the gain coefficient can be obtained as

$$G(V_o) = G_S\left[1 - \frac{1}{2}\left(\frac{\pi V_o}{2V_\pi}\right)^2 + \frac{1}{12}\left(\frac{\pi V_o}{2V_\pi}\right)^4\right] \tag{10b}$$

It should be kept in mind that, in general, $G(V_o)$ is also a function of the frequency, $\omega$, of the input signal, because $V_{ph}$ is linearly proportional to the gain of the RF amplifier and the responsivity of the photodetector, which are all frequency dependent. In addition, the $V_\pi$ of the modulator is also a function of the input RF frequency. Furthermore, the frequency response of the RF filter in the loop can also be lumped into $G(V_o)$. In the discussions below, we will introduce a unitless complex filter function, $\tilde{F}(\omega)$, to explicitly account for the combined effect of all frequency-dependent components in the loop while treating $G(V_o)$ to be frequency independent:

$$\tilde{F}(\omega) = F(\omega)e^{i\phi(\omega)} \tag{11}$$

where $\phi(\omega)$ is the frequency-dependent phase caused by the dispersive component in the loop and $F(\omega)$ is the real normalized transmission function. Now Eq. (9) can be rewritten in complex form as

$$\tilde{V}_{out}(t) = \tilde{F}(\omega)G(V_o)\tilde{V}_{in}(\omega, t) \tag{12}$$

where $\tilde{V}_{in}$ and $\tilde{V}_{out}$ are complex input and output voltages. Note that although Eq. (12) is linear, the nonlinear effect of the modulator is not lost—it is contained in the nonlinear gain coefficient, $G(V_o)$.

### C. Oscillation Frequency and Amplitude

In this section, we derive the expressions for the amplitude and frequency of the LIMO. Like other oscillators, the oscillation of a LIMO starts from noise transient, which is then built up and sustained with feedback at the level of the oscillator output signal. We derive the amplitude of the oscillating signal by considering this process mathematically. The noise transient can be viewed as a collection of sine waves with random phases and amplitudes. To simplify our derivation, we use this noise input with the linearized expression of Eq. (12) for the loop response. Because Eq. (12) is linear, the superposition

principle holds, and we can analyze the response of the LIMO by first inspecting the influence of a single-frequency component of the noise spectrum:

$$\tilde{V}_{in}(\omega, t) = \tilde{V}_{in}(\omega)e^{i\omega t} \tag{13}$$

where $\tilde{V}_{in}(\omega)$ is a complex amplitude of the frequency component.

Once the noise component of Eq. (13) is in the oscillator, it would circulate in the loop, and the recurrence relation of the fields from Eq. (12) is

$$\tilde{V}_n(\omega, t) = \tilde{F}(\omega)G(V_o)\tilde{V}_{n-1}(\omega, t - \tau') \tag{14}$$

where $\tau'$ is the time delay resulting from the physical length of the feedback and $n$ is the number of times the field has circulated around the loop, with $\tilde{V}_{n=0}(\omega, t) = \tilde{V}_{in}(\omega, t)$. In Eq. (14), the argument $V_o$ in $G(V_o)$ is the amplitude of the total field (the sum of all circulating fields) in the loop.

The total field at any instant of time is the summation of all circulating fields. Therefore, with the input of Eq. (13) injected in the oscillator, the signal measured at the RF input to the modulator for the case when the open-loop gain is less than unity can be expressed as

$$\tilde{V}(\omega, t) = G_a\tilde{V}_{in}(\omega)\sum_{n=0}^{\infty}\left[\tilde{F}(\omega)G(V_o)\right]^n e^{i\omega(t-n\tau')} = \frac{G_a\tilde{V}_{in}e^{i\omega t}}{1 - \tilde{F}(\omega)G(V_o)e^{-i\omega\tau'}} \tag{15}$$

For a loop gain below threshold and with a small $V_o$, $G(V_o)$ is essentially the small signal gain, $G_S$, given by Eq. (4).

The corresponding RF power of the circulating noise at frequency $\omega$ is, therefore,

$$P(\omega) = \frac{|\tilde{V}(\omega, t)|^2}{2R} = \frac{G_A^2|\tilde{V}_{in}(\omega)|^2/(2R)}{1 + |F(\omega)G(V_o)|^2 - 2F(\omega)|G(V_o)|\cos[\omega\tau' + \phi(\omega) + \phi_0]'} \tag{16}$$

where $\phi_0 = 0$ if $G(V_o) > 0$ and $\phi_0 = \tau$ if $G(V_o) < 0$.

For a constant $\tilde{V}_{in}(\omega)$, the frequency response of a LIMO has equally spaced peaks similar to those of a Fabry-Perot resonator, as shown in Fig. 3. These peaks are located at the frequencies determined by

$$\omega_k\tau' + \phi(\omega_k) + \phi_o = 2k\pi \qquad k = 0, 1, 2, \cdots \tag{17}$$

where $k$ is the mode number. In Fig. 3, each peak corresponds to a frequency component resulting from the coherent summation of all circulating fields in the loop at that frequency. As the open-loop gain increases, the magnitude of each peak becomes larger and its shape becomes sharper. These peaks are the possible oscillation modes of the LIMO. When the open-loop gain is larger than unity, each time a noise component at a peak frequency travels around the loop, it is amplified and its amplitude increases geometrically—an oscillation is started from noise.

Because an RF filter is placed in the loop, the gain of only one mode is allowed to be larger than unity, thus selecting the mode that is allowed to oscillate. Because of the nonlinearity of the E/O modulator or the RF amplifier, the amplitude of the oscillation mode cannot increase indefinitely. As the amplitude

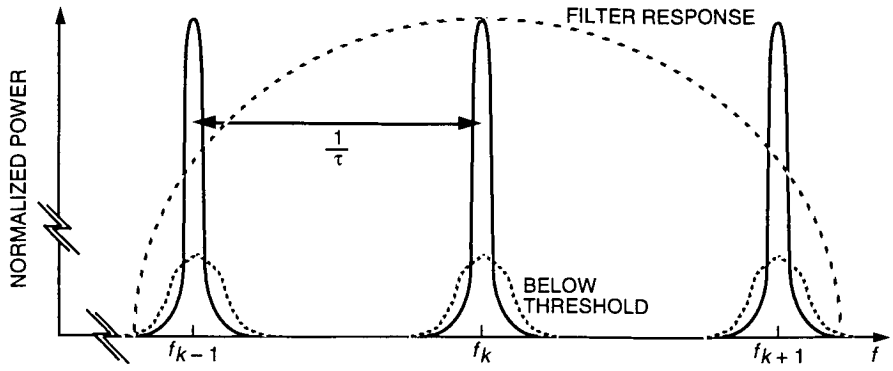


**Fig. 3. Illustration of the oscillator's output spectra below and above the threshold.**

increases, higher harmonics of the oscillation will be generated by the nonlinear effect of the modulator or the amplifier, at the expense of the oscillation power, and these higher harmonics will be filtered out by the RF filter. Effectively, the gain of the oscillation mode is decreased according to Eq. (10) until the gain is, for all practical measures, equal to unity, and the oscillation is stable. As will be shown later, because of the continuous presence of noise, the closed-loop gain of an oscillating mode is actually less than unity by a tiny amount on the order of $10^{-10}$, which ensures that the summation in Eq. (15) converges.

In the discussion that follows, only one mode $k$ is allowed to oscillate, and so we will denote the oscillation frequency of this mode as $f_{osc}$ or $\omega_{osc}(\omega_{osc} = 2\pi f_{osc})$, its oscillation amplitude as $V_{osc}$, and its oscillation power as $P_{osc}(P_{osc} = V_{osc}^2/2R)$. In this case, the amplitude, $V_o$, of the total field in Eq. (16) is just the oscillation amplitude, $V_{osc}$, of the oscillating mode. If we choose the transmission peak of the filter to be at the oscillation frequency, $\omega_{osc}$, and so $F(\omega_{osc}) = 1$, the oscillation amplitude can be solved by setting the gain coefficient, $|G(V_{osc})|$, in Eq. (16) at unity. From Eq. (10a), this leads to

$$\left| J_1\left(\frac{\pi V_{osc}}{V_\pi}\right)\right| = \frac{1}{2|G_s|}\,\frac{\pi V_{osc}}{V_\pi} \tag{18a}$$

In deriving Eq. (18a), we have assumed that the RF amplifier in the loop is linear enough that the oscillation power is limited by the nonlinearity of the E/O modulator. The amplitude of the oscillation can be obtained by solving Eq. (18a) graphically, and the result is shown in Fig. 4(a). Note that this result is the same as that obtained by Neyer and Voges [20] using a more complicated approach.

If we use Eq. (10b), we can obtain the approximated solutions of the oscillation amplitude:

$$V_{osc} = \frac{2\sqrt{2}V_\pi}{\pi}\sqrt{1-\frac{1}{|G_S|}} \qquad \text{third-order expansion} \tag{18b}$$

$$V_{osc} = \frac{2\sqrt{3}V_\pi}{\pi}\left(1-\frac{1}{\sqrt{3}}\sqrt{\frac{4}{|G_S|}-1}\right)^{1/2} \qquad \text{fifth-order expansion} \tag{18c}$$

The threshold condition of $|G_S| \geq 1$ is clearly indicated in Eqs. (18b) and (18c). Figure 4(a) shows the normalized oscillation amplitude as a function of $|G_S|$, obtained from Eqs. (18a), (18b) and (18c), respectively. Comparing the three theoretical curves, one can see that, for $|G_S| \leq 1.5$, the third-order expansion result is a good approximation. For $|G_S| \leq 3$, the fifth-order expansion result is a good approximation.

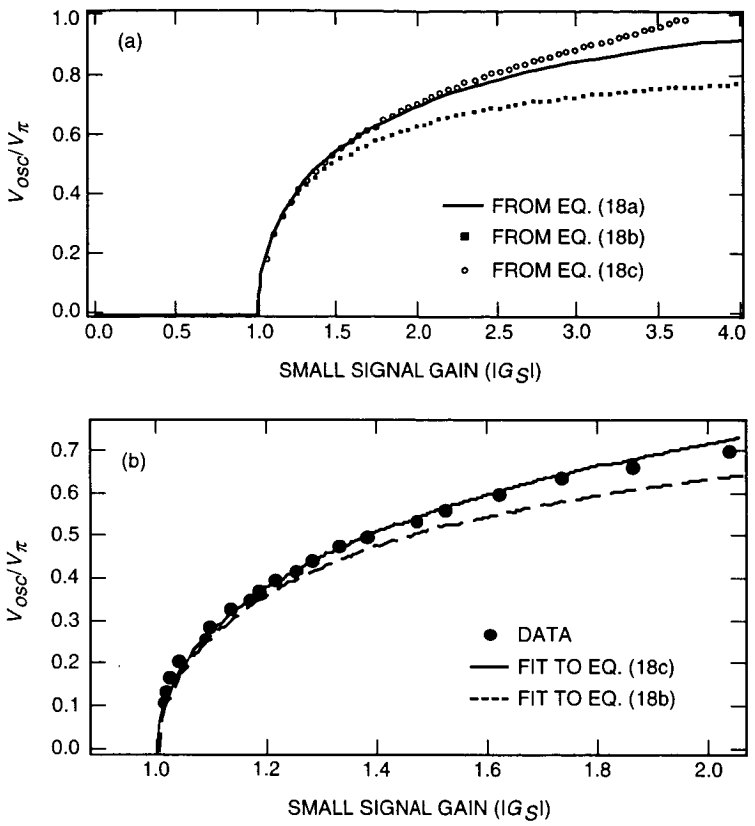


**Fig. 4. The normalized oscillation amplitude of a LIMO as a function of small signal gain,$|G_S|$: (a) theoretical calculation using Eqs. (18a), (18b), and (18c) and (b) experimental data and curve fitting to Eqs. (18b) and (18c).**

The corresponding oscillation frequency, $f_{osc} \equiv f_k = \omega_k/2\pi$, can be obtained from Eq. (17) as

$$f_{osc} \equiv f_k = \frac{(k+1/2)}{\tau} \qquad \text{for } G(V_{osc}) < 0 \tag{19a}$$

$$f_{osc} \equiv f_k = \frac{k}{\tau} \qquad \text{for } G(V_{osc}) > 0 \tag{19b}$$

where $\tau$ is the total group delay of the loop, including the physical length delay, $\tau'$, of the loop and the group delay resulting from dispersive components (such as an amplifier) in the loop, and it is given by

$$\tau = \tau' + \left.\frac{d\phi(\omega)}{d\omega}\right|_{\omega=\omega_{osc}} \tag{20}$$

For all practical purposes, $J_1(\pi V_{osc}/V_\pi) \geq 0$ or $V_{osc}/V_\pi \leq 1.21$, and the sign of $G(V_{osc})$ is determined by the small signal gain, $G_s$. It is interesting to notice from Eqs. (19a) and (19b) that the oscillation frequency depends on the biasing polarity of the modulator. For negative biasing ($G_s < 0$), the fundamental frequency is $1/(2\tau)$, while for positive biasing ($G_s > 0$), the fundamental frequency is doubled to $1/\tau$.

### D. The Spectrum

The fundamental noise in a LIMO consists of the thermal noise, the shot noise, and the laser's intensity noise, which for the purpose of analysis can be viewed as all originating from the photodetector. Since the photodetector is directly connected to the amplifier, the noise can be viewed as entering the oscillator at the input of the amplifier, as shown in Fig. 2.

We compute the spectrum of the oscillator signal by determining the power spectral density of noise in the oscillator. Let $\rho_N(\omega)$ be the power density of the input noise at frequency $\omega$; we have

$$\rho_N(\omega)\Delta f = \frac{|\tilde{V}_{in}(\omega)|^2}{2R} \tag{21}$$

where $\Delta f$ is the frequency bandwidth. Substituting Eq. (21) in Eq. (16) and letting $F(\omega_{osc}) = 1$, we obtain the power spectral density of the oscillating mode $k$:

$$S_{RF}(f') = \frac{P(f')}{\Delta f P_{osc}} = \frac{\rho_N G_A^2 / P_{osc}}{1 + |F(f')G(V_{osc})|^2 - 2F(f')|G(V_{osc})|\cos(2\pi f'\tau)} \tag{22}$$

where $f' \equiv (\omega - \omega_{osc})/2\pi$ is the frequency offset from the oscillation peak, $f_{osc}$. In deriving Eq. (22), both Eqs. (17) and (20) are used.

By using the normalization condition

$$\int_{-\infty}^{\infty} S_{RF}(f')df' \approx \int_{-1/2\tau}^{1/2\tau} S_{RF}(f')df' = 1 \tag{23}$$

we obtain

$$1 - |G(V_{osc})|^2 \approx 2\,[1 - |G(V_{osc})|] = \frac{\rho_N G_A^2}{\tau P_{osc}} \tag{24}$$

Note that in Eq. (23) we have assumed that the spectral width of the oscillating mode is much smaller than the mode spacing, $1/\tau$, of the oscillator, so that the integration over $1/\tau$ is sufficiently accurate. In addition, in the derivation, we have assumed that $|F(f')| \approx 1$ in the frequency band of integration.

Typically, $\rho_N \sim 10^{-17}$ mW/Hz, $P_{osc} \sim 10$ mW, $G_A^2 \sim 100$, and $\tau \sim 10^{-6}$ s. From Eq. (24), one can see that the closed-loop gain, $|G(V_{osc})|$, of the oscillating mode is less than unity by an amount of $10^{-10}$. Therefore, the equation $|G(V_{osc})| = 1$ is sufficiently accurate for calculating the oscillation amplitude, $V_{osc}$, as in Eqs. (18a), (18b), and (18c).

Finally, substituting Eq. (24) in Eq. (22), we obtain the RF spectral density of the LIMO:

$$S_{RF}(f') = \frac{\delta}{(2 - \delta/\tau) - 2\sqrt{1 - \delta/\tau}\cos(2\pi f'\tau)} \tag{25}$$

where $\delta$ is defined as

$$\delta \equiv \frac{\rho_N G_A^2}{P_{osc}} \tag{26}$$

As mentioned before, $\rho_N$ is the equivalent input noise density injected into the oscillator from the input port of the amplifier, and $P_{osc}/G_A^2$ is the total oscillating power measured before the amplifier. Therefore, $\delta$ is the input noise-to-signal ratio to the oscillator.

For the case where $2\pi f'\tau << 1$, we can simplify Eq. (25) by expanding the cosine function in Taylor series:

$$S_{RF}(f') = \frac{\delta}{(\delta/2\tau)^2 + (2\tau)^2(\tau f')^2} \tag{27}$$

Equation (27) is a good approximation even for $2\pi f'\tau = 0.7$, at which value the error resulting from neglecting the higher-order terms in Taylor expansion is less than 1 percent. It can be seen from Eq. (27) that the spectral density of the oscillating mode is a Lorentzian function of frequency. Its full width at half maximum (FWHM), $\Delta f_{FWHM}$, is

$$\Delta f_{FWHM} = \frac{1}{2\pi}\frac{\delta}{\tau^2} = \frac{1}{2\pi}\frac{G_A^2\rho_N}{\tau^2 P_{osc}} \tag{28a}$$

It is evident from Eq. (28a) that $\Delta f_{FWHM}$ is inversely proportional to the square of the loop delay time and linearly proportional to the input noise-to-signal ratio, $\delta$. For a typical $\delta$ of $10^{-16}$/Hz and a loop delay of 100 ns (20 m), the resulting spectral width is submilli-Hertz. The fractional power contained in $\Delta f_{FWHM}$ is $\Delta f_{FWHM} S_{RF}(0) = 64$ percent.

From Eq. (28a), one can see that for fixed $\rho_N$ and $G_A$, the spectral width of a LIMO is inversely proportional to the oscillation power, similar to the famous Schawlow–Townes formula [23,35] for describing the spectral width, $\Delta\nu_{laser}$, of a laser:

$$\Delta\nu_{laser} = \frac{1}{2\pi}\frac{\rho_s}{\tau_{laser}^2 P_{laser}} \tag{28b}$$

where $\rho_s = h\nu$ is the spontaneous emission noise density of the laser, $P_{laser}$ is the laser oscillation power, and $\tau_{laser}$ is the decay time of the laser cavity. However, because, as will be shown in Section III.E, both $P_{osc}$ and $\rho_N$ are functions of the photocurrent, the statement that the spectral width of a LIMO is inversely proportional to the oscillation power is valid only when thermal noise dominates in the oscillator at low photocurrent levels.

From Eq. (28a), the quality factor, Q, of the oscillator is

$$Q = \frac{f_{osc}}{\Delta f_{FWHM}} = Q_D\frac{\tau}{\delta} \tag{29}$$

where $Q_D$ is the quality factor of the loop delay line and is defined as

$$Q_D = 2\pi f_{osc}\tau \tag{30}$$

From Eq. (27), we easily obtain

$$S_{RF}(f') = \frac{4\tau^2}{\delta} \qquad |f'| << \frac{\Delta f_{FWHM}}{2} \tag{31a}$$

$$S_{RF}(f') = \frac{\delta}{(2\pi)^2(\tau f')^2} \qquad |f'| >> \frac{\Delta f_{FWHM}}{2} \tag{31b}$$

It can be shown [24] that for an oscillator with a phase fluctuation much less than unity, its power spectral density is equal to the sum of the single side-band phase noise density and the single side-band amplitude noise density. In most cases in which the amplitude fluctuation is much less than the phase fluctuation, the power spectral density is just the single side-band phase noise. Therefore, it is evident from Eq. (31b) that the phase noise of the LIMO decreases quadratically with the frequency offset, $f'$. For a fixed $f'$, the phase noise decreases quadratically with the loop delay time. The larger the $\tau$, the smaller the phase noise. However, the phase noise cannot decrease to zero no matter how large $\tau$ is, because at large enough $\tau$, Eqs. (27) and (31b) are not valid anymore. From Eq. (27), the minimum phase noise is $S_{RF}^{\min} \approx \delta/4$ at $f' = 1/2\tau$. For the frequency offset, $f'$, outside of the passband of the loop filter (where $F(f') = 0$), the phase noise is simply the noise-to-signal ratio, $\delta$, as can be seen from Eq. (22).

Equations (25), (27), and (31b) also indicate that the oscillator's phase noise is independent of the oscillation frequency, $f_{osc}$. This result is significant because it allows the generation of high-frequency and low phase-noise signals with the LIMO. The phase noise of a signal generated using frequency-multiplying methods generally increases quadratically with the frequency.

### E. The Noise-to-Signal Ratio

As mentioned before, the total noise density input to the oscillator is the sum of the thermal noise, $\rho_{thermal} = 4k_BT(NF)$; the shot noise, $\rho_{shot} = 2eI_{ph}R$; and the laser's relative intensity noise (RIN), $\rho_{RIN} = N_{RIN}I_{ph}^2R$, densities [25,26]:

$$\rho_N = 4k_BT(NF) + 2eI_{ph}R + N_{RIN}I_{ph}^2R \tag{32}$$

where $k_B$ is Boltzmann's constant, $T$ is the ambient temperature, $NF$ is the noise factor of the RF amplifier, $e$ is the electron charge, $I_{ph}$ is the photocurrent across the load resistor of the photodetector, and $N_{RIN}$ is the RIN of the pump laser.

From Eqs. (26) and (32), one can see that if the thermal noise is dominant, then $\delta$ is inversely proportional to the oscillating power, $P_{osc}$, of the oscillator. In general, $P_{osc}$ is a function of photocurrent, $I_{ph}$, and amplifier gain, $G_A$, as determined by Eqs. (18a), (18b), and (18c), and the noise-to-signal ratio from Eq. (26) is thus,

$$\delta = \frac{|G_S|^2}{1 - 1/|G_S|} \frac{4kT(NF) + 2eI_{ph}R + N_{RIN}I_{ph}^2R}{4\eta^2\cos^2(\pi V_B/V_\pi)D_{ph}^2R} \tag{33}$$

In deriving Eq. (33), Eqs. (4) and (18b) are used. From Eq. (33), one can see that $\delta$ is a nonlinear function of the small signal gain of the oscillator. As shown in Fig. 5(a), it reaches the minimum value at $|G_S| = 3/2$:

$$\delta_{\min} = \frac{4kT(NF) + 2eI_{ph}R + N_{RIN}I_{ph}^2R}{(16/27)I_{ph}^2R} \tag{34}$$

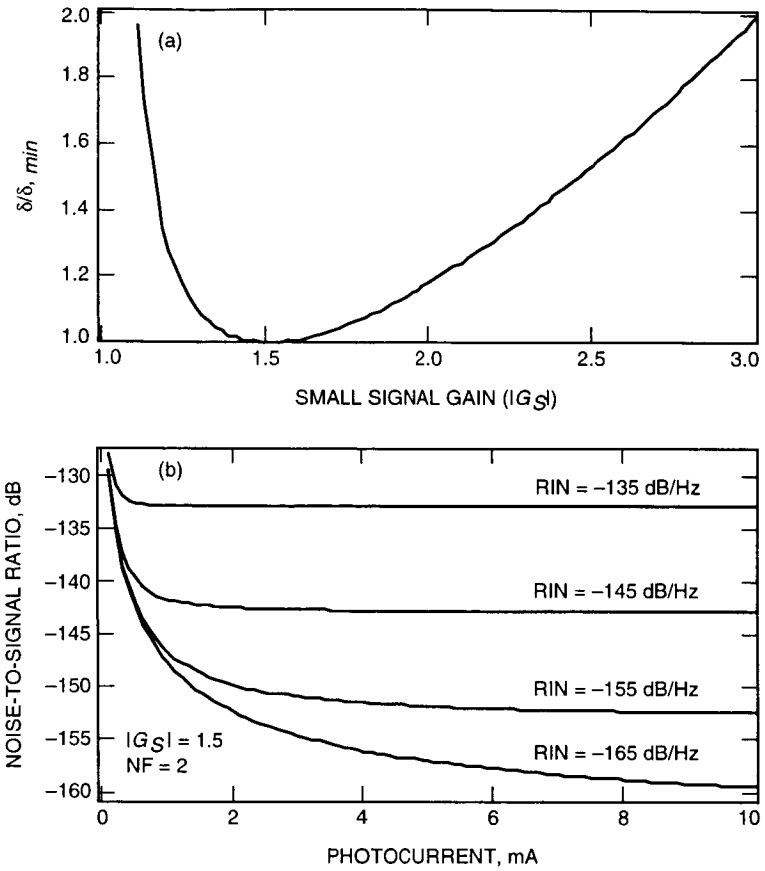


**Fig. 5. Calculations of the LIMO's input noise-to-signal ratio (a) as a function of small signal gain, $|G_S|$, showing a minimum value at $|G_S|$= 1.5 and (b) as a function of photocurrent for different values of the laser's RIN noise.**

where $\eta = 1$ and $\cos(\pi V_B/V_\pi) = 1$ are assumed. The oscillation amplitude at $|G_S| = 3/2$ can be obtained from Eq. (18b) as

$$V_{osc} = \frac{2\sqrt{2}V_\pi}{(\pi\sqrt{3})} \approx 0.52V_\pi \tag{35a}$$

and the corresponding RF power is

$$P_{osc} = \frac{4V_\pi^2}{(3\pi^2 R)} = \frac{10P_m^{1\text{dB}}}{3} \tag{35b}$$

where $P_m^{1\text{dB}}$ is the input 1-dB compression power of the E/O modulator [26,27]. From Eq. (35b), one can conclude that, in order to have minimum noise, the oscillation power measured at the input of the E/O modulator should be 5 dB above the 1-dB compression power of the modulator. Equation (35a) indicates that the noise of the oscillator is at minimum when the oscillating amplitude is roughly half of $V_\pi$ or the voltage in the oscillator is varying between the peak and the trough of the sinusoidal transmission curve of the E/O modulator. This makes sense because the modulator has its minimum sensitivity to voltage

variations at the maximum and the minimum of the transmission curve, and the most likely cause of voltage variations in a LIMO is the noise in the loop.

It is evident from Eq. (34) that the higher the photocurrent, the less the noise-to-signal ratio of the oscillator until it flattens out at the laser's RIN level. Therefore, the ultimate noise-to-signal ratio of a LIMO is limited by the pump laser's RIN. If the RIN of the pump laser in a LIMO is $-160$ dB/Hz, the ultimate noise-to-signal ratio of the oscillator is also $-160$ dB/Hz, and the signal-to-noise ratio is 160 dB/Hz. Figure 5(b) shows the noise-to-signal ratio, $\delta$, as a function of photocurrent, $I_{ph}$, for different RIN levels. In the plot, the small signal gain, $G_S$, is chosen to be a constant of 1.5, which implies that when $I_{ph}$ is increased, the amplifier gain, $G_A$, must be decreased to keep $G_S$ constant. From the figure, one can easily see that $\delta$ decreases quadratically with $I_{ph}$ at small $I_{ph}$ and flattens out at the RIN level at large $I_{ph}$.

### F. Effects of Amplifier Nonlinearity

In the discussions above, we have assumed that the nonlinear distortion of a signal from the E/O modulator is more severe than from the amplifier (if any) used in the oscillator, so that the oscillation amplitude or power is limited by the nonlinear response of the E/O modulator. Using an engineering term, this simply means that the output 1-dB compression power of the amplifier is much larger than the input 1-dB compression power of the E/O modulator [26].

For cases in which the output 1-dB compression power of the amplifier is less than the input 1-dB compression power of the modulator, the nonlinearity of the amplifier will limit the oscillation amplitude, $V_{osc}$ (or power $P_{osc}$), of the oscillator, resulting in an oscillation amplitude less than that given by Eqs. (18a), (18b), and (18c). The exact relation between the oscillation amplitude (or power) and the small signal gain, $G_S$, can be determined using the same linearization procedure as that used for obtaining Eqs. (18a), (18b), and (18c) if the nonlinear response function of the amplifier is known. However, all the equations in Section III.D for describing the spectrum of the oscillator are still valid, provided that the oscillation power in those equations is determined by the nonlinearity of the amplifier. For a high enough small signal gain, $G_S$, the oscillation power is approximately a few dB above the output 1-dB compression power of the amplifier.

In all the experiments below, the output 1-dB compression power of the amplifiers chosen is much larger than the input 1-dB compression power of the modulator, so that the oscillation power is limited by the modulator.

## IV. Experiments

### A. Amplitude Versus Open-Loop Gain

We performed measurements to test the level of agreement of the theory described above with experimental results. In all of our experiments, we used a highly stable diode-pumped Nd:YAG ring laser [35] with a built-in RIN reduction circuit [36] to pump the LIMO. The experimental setup for measuring the oscillation amplitude as a function of the open-loop gain is shown in Fig. 6(a). Here an RF switch was used to open and close the loop. While the loop was open, an RF signal from a signal generator with the same frequency as the oscillator was injected into the E/O modulator. The amplitudes of the injected signal and the output signal from the loop were measured with an oscilloscope to obtain the open-loop gain, which was the ratio of the output amplitude to the injected signal amplitude. The open-loop gain was varied by changing the bias voltage of the E/O modulator, by attenuating the optical power of the loop, or by using a variable RF attenuator after the photodetector, as indicated by Eq. (4). When closing the loop, the amplitude of the oscillation was conveniently measured using the same oscilloscope. We measured the oscillation amplitudes of the LIMO for different open-loop gains at an oscillation frequency of 100 MHz, and the data obtained are plotted in Fig. 4(b). It is evident that the experimental data agreed well with our theoretical predictions.

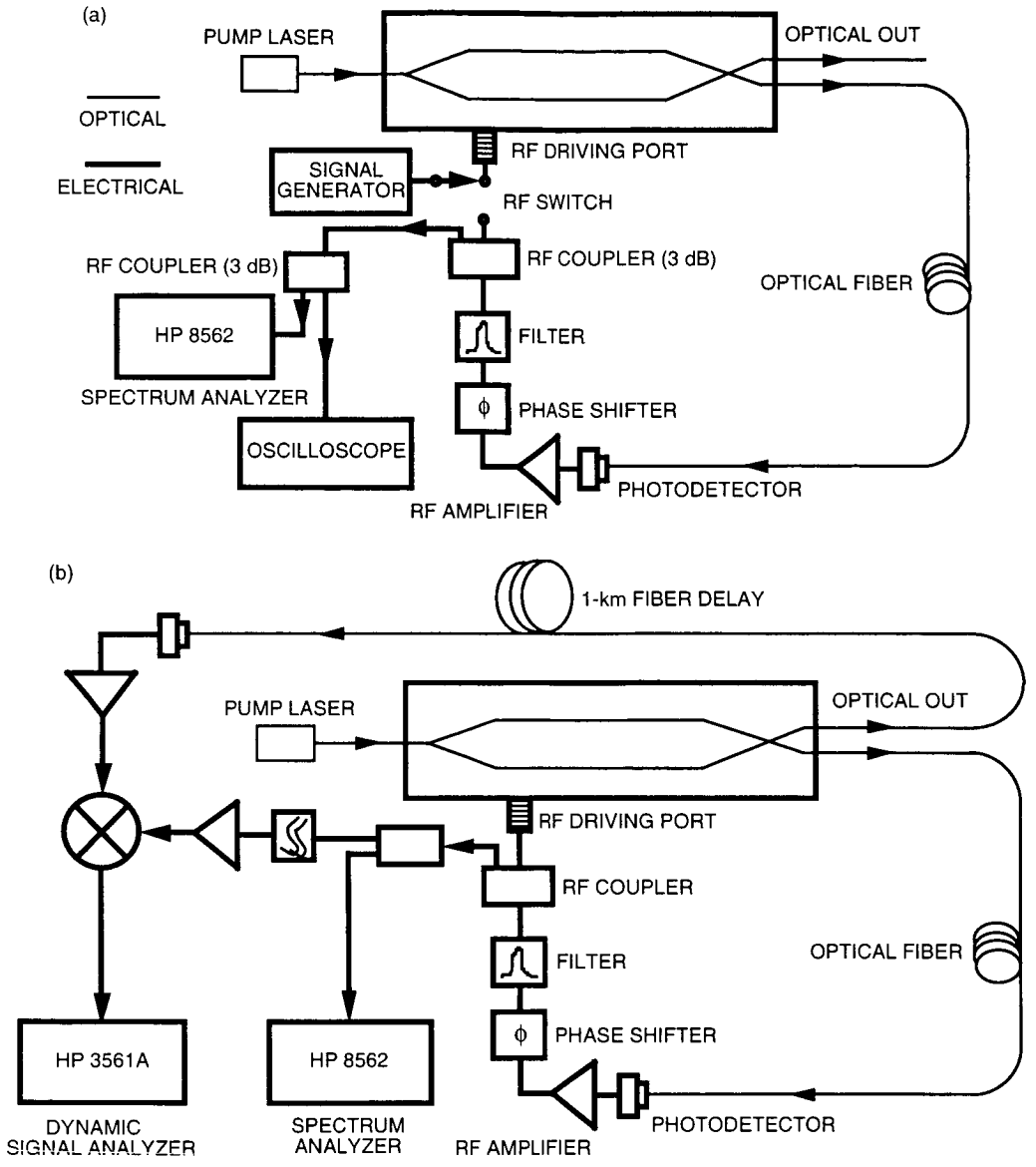


**Fig. 6. Experimental setups for measuring (a) the oscillation amplitude of a LIMO as a function of the small signal gain and (b) the phase noise of a LIMO using the frequency discrimination method.**

## B. Phase Noise Measurement Setup

We used a frequency discriminator method [28] to measure the phase noise of the LIMO, and the experimental setup is shown in Fig. 6(b). The advantage of this method is that it does not require a frequency reference and, hence, can be used to measure an oscillator of any frequency. Using a microwave mixer in the experiment, we compared the phase of a signal from the electrical output port of the LIMO with its delayed replica from the optical output port. The length of the delay line is important because, the longer the delay line, the lower the frequency offset at which the phase noise can be accurately measured. On the other hand, if the delay line is too long, the accuracy of the phase noise at a higher frequency offset will suffer. The length of delay used in our experiment is 1 km, or 5 $\mu$s. Because of this delay, any frequency fluctuation of the LIMO will cause a voltage fluctuation at the output of the mixer. We measured the spectrum of this voltage fluctuation with a high dynamic-range spectrum analyzer and

transferred the spectral data to a computer. Finally, we converted this information into the phase noise spectrum of the LIMO according to the procedures given in [28]. In these experiments, the noise figure of the RF amplifier was 7 dB.

### C. Phase Noise as a Function of Offset Frequency and Loop Delay

Figure 7(a) is the log versus log scale plot of the measured phase noise as a function of the frequency offset, $f'$. Each curve corresponds to a different loop delay time. The corresponding loop delays for curves 1–5 are listed adjacent to each curve, and the corresponding oscillation powers are 16.33, 16, 15.67, 15.67, and 13.33 dBm, respectively. Curve fitting yields the following phase noise relations as a function of frequency offset $f'$: $-28.7 - 20\log(f')$, $-34.84 - 20\log(f')$, $-38.14 - 20\log(f')$, $-40.61 - 20\log(f')$, and $-50.45 - 20\log(f')$. Clearly, the phase noise has a 20-dB-per-decade dependence with the frequency offset, in excellent agreement with the theoretical prediction of Eq. (31b).

Figure 7(b) is the measured phase noise at 30 kHz from the center frequency as a function of loop delay time, with data points extracted from curves 1–5 of Fig. 7(a) and corrected to account for oscillation power differences. Because the loop delay is increased by adding more fiber segments, the open-loop gains of the oscillator with longer loops decrease as more segments are connected, causing the corresponding oscillation power to decrease. From the results of Fig. 8, discussed below, the phase noise of the LIMO decreases

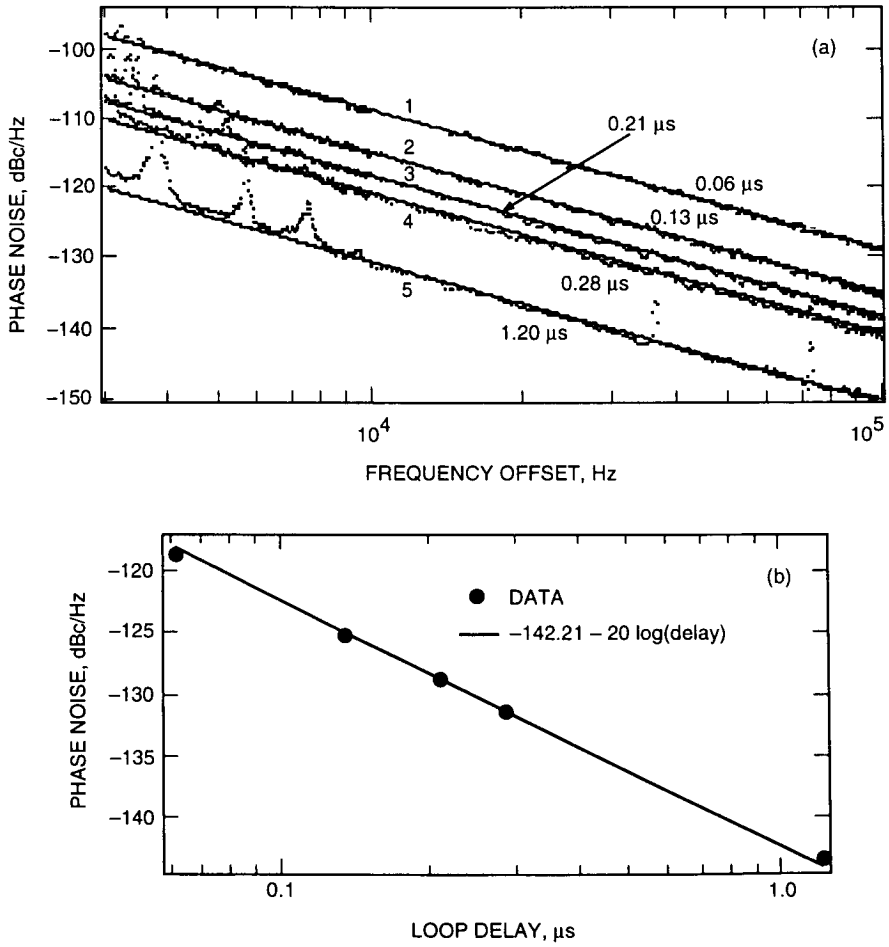


**Fig. 7. Single side-band phase noise of a LIMO measured at 800 MHz: (a) phase noise spectra at different loop delays and their fits to Eq. (31b) and (b) phase noise at a 30-kHz offset from the center frequency as a function of loop delay time.**

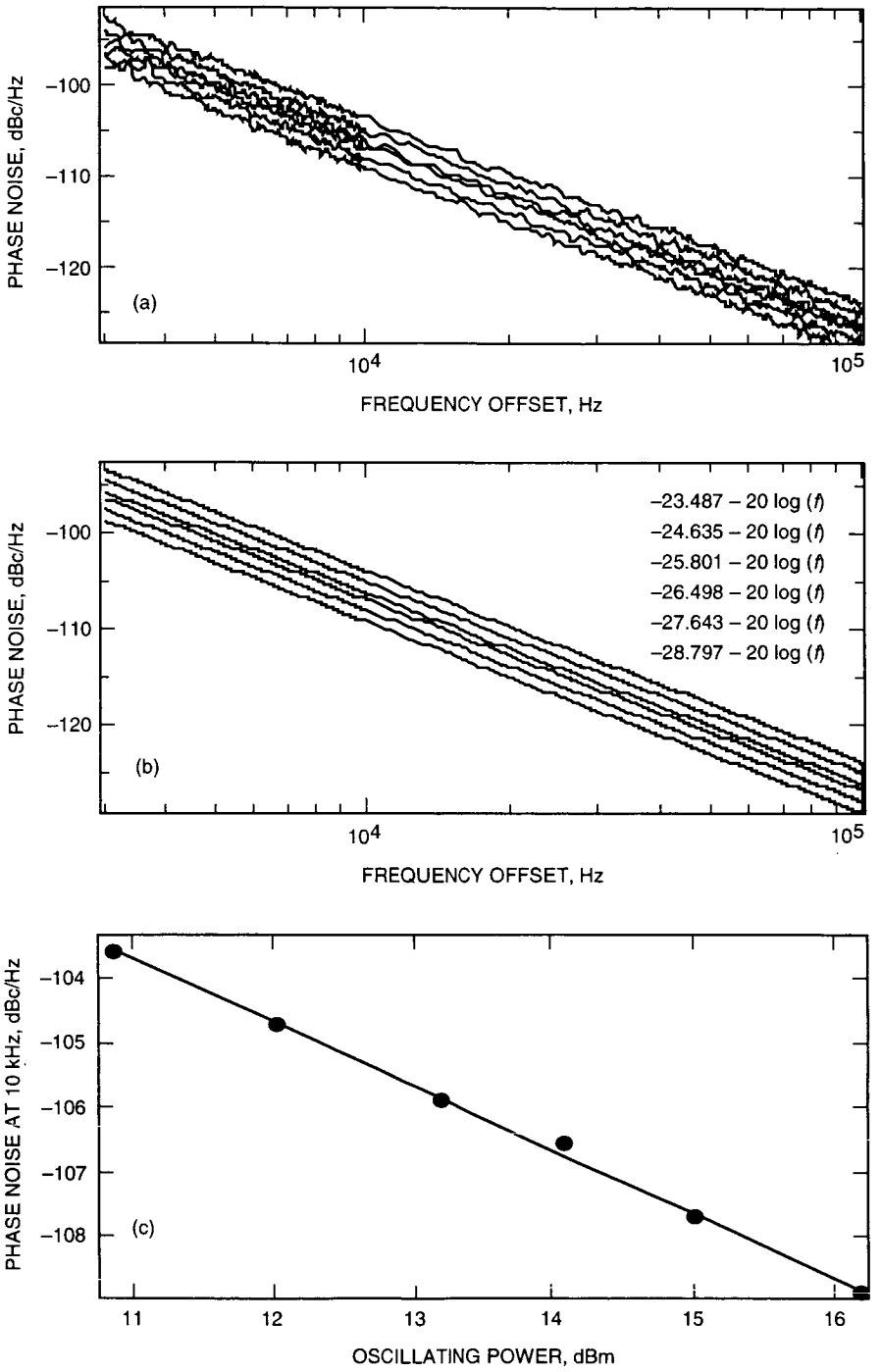


**Fig. 8. Single side-band phase noise spectra as a function of oscillation power measured at 800 MHz: (a) experimental data, (b) the fit to Eq. 31(b), and (c) phase noise at a 10-kHz offset as a function of oscillation power, extracted from (b).**

linearly with the oscillation power. To extrapolate the dependence of the phase noise on the loop delay only from Fig. 7(a), each data point in Fig. 7(b) is calibrated using the linear dependence of Fig. 8, while keeping the oscillation power for all data points at 16.33 mW. Again, the experimental data agree well with the theoretical prediction.

### D. Phase Noise as a Function of Oscillation Power

We have also measured the phase noise spectrum of the LIMO as a function of oscillation power, with the results shown in Fig. 8. In that experiment, the loop delay of the LIMO was 0.06 $\mu$s, the noise figure of the RF amplifier was 7 dB, and the oscillation power was varied by changing the photocurrent, $I_{ph}$, according to Eqs. (3), (4), (18a), (18b), and (18c). With this amplifier and the photocurrent level (1.8 mA – 2.7 mA), the thermal noise in the oscillator dominates. Recall that, in Eqs. (25), (26), and (27), the phase noise of a LIMO is shown to be inversely proportional to the oscillation power. This is true if the gain of the amplifier is kept constant and the photocurrent is low enough to ensure that the thermal noise is the dominant noise term. In Fig. 8(a), each curve is the measurement data of the phase noise spectrum corresponding to an oscillating power, and the curves in Fig. 8(b) are the fits of the data to Eq. (31b). Figure 8(c) is the phase noise of the LIMO at 10 kHz as a function of the oscillation power, extracted from the data of Fig. 8(b). The resulting linear dependence agrees well with the theoretical prediction of Eq. (31b).

### E. Phase Noise Independence of Oscillation Frequency

To confirm our prediction that the phase noise of the LIMO is independent of the oscillation frequency, we measured the phase noise spectrum as a function of the oscillation frequency, and the result is shown in Fig. 9(a). In the experiment, we kept the loop length at 0.28 $\mu$s and varied the oscillation frequency by changing the RF filter in the loop. The frequency was fine tuned using an RF line stretcher. It is evident from Fig. 9(a) that all phase noise curves at frequencies 100, 300, 700, and 800 MHz overlap with one another, indicating a good agreement with the theory. Figure 9(b) is a plot of the phase noise data at 10 kHz as a function of the frequency. As predicted, it is a flat line, in contrast with the case when a frequency multiplier is used to obtain higher frequencies. This result is significant because it confirms that the LIMO can be used to generate high-frequency signals up to 75 GHz with a much lower phase noise than that which can be attained with frequency-multiplying techniques.

## V. Summary

We have introduced a high-frequency, high-stability, high spectral purity, widely tunable electro-optic oscillator, which we have termed a LIMO. The high stability and spectral purity of the LIMO result from the extremely low energy storage loss realization obtained with a long optical fiber. The optical fiber is also virtually free of any frequency-dependent loss, resulting in the same long storage time and high spectral purity signals for both low- and high-frequency oscillation. On the other hand, the oscillation frequency of the LIMO is limited only by the speed of the modulator, which at the present can be as high as 75 GHz. As yet another unique feature, the output of the LIMO may be obtained directly as microwave signals or as intensity modulations on an optical carrier for easy interface with optical systems.

We also analyzed the performance of this oscillator by deriving expressions for the oscillation threshold, Eq. (5); oscillation amplitude, Eqs. (18a), (18b), and (18c); and oscillation frequency, Eqs. (19a) and (19b). These results agree quite well with experimental data obtained with laboratory versions of the LIMO.

We also derived the expression for the spectrum of the output of the LIMO and showed that it has a Lorenzian line shape, given by Eqs. (25) and (27). The spectral width of the output signal was found to be inversely proportional to the square of the loop delay time, given by Eqs. (28a) and (28b). In addition, at low optical pumping levels where thermal noise dominates, the spectral line width was found to be inversely proportional to the oscillation power of the oscillator, similar to the Schawlow–Townes formula

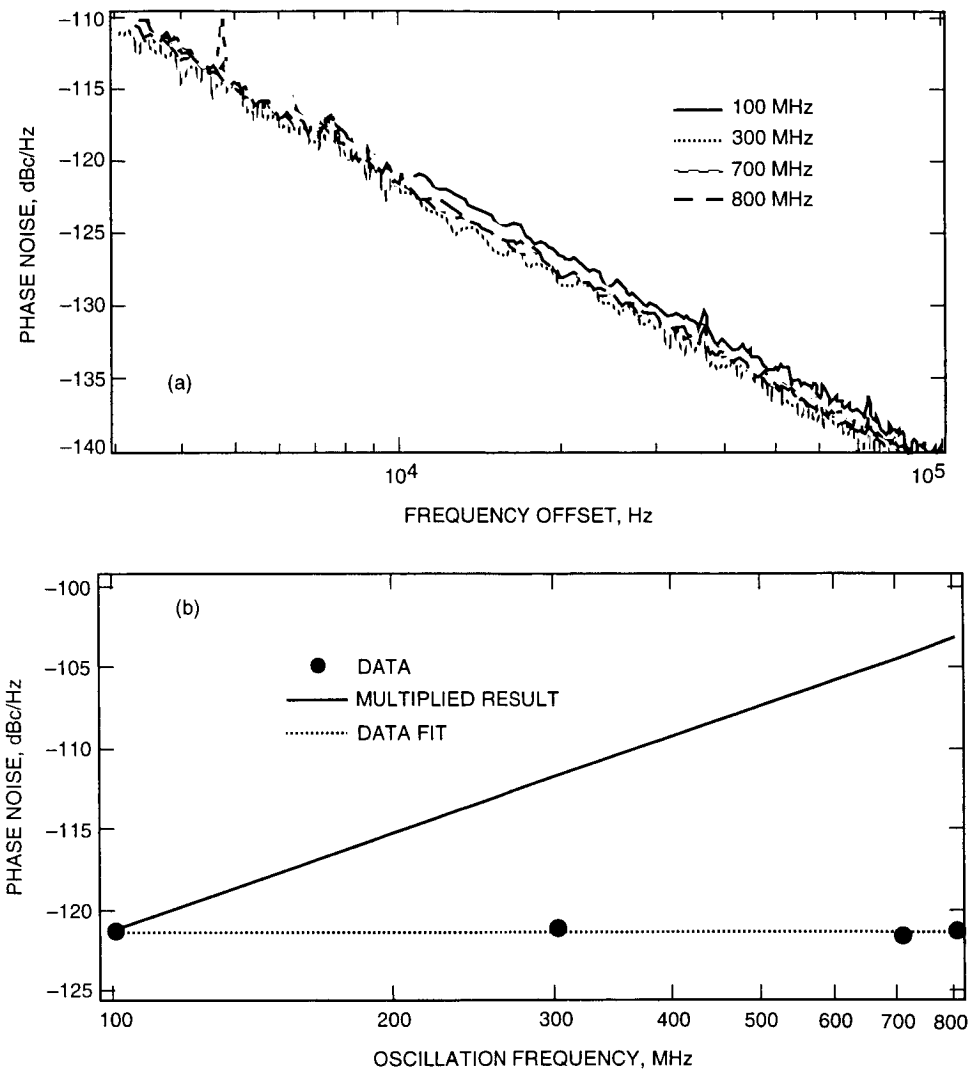


**Fig. 9. Single side-band phase noise measurements of the LIMO at different oscillation frequencies: (a) phase noise spectra and (b) phase noise at a 10-kHz offset frequency as a function of oscillation frequency, extracted from (a). The loop delay for the measurements is 0.28 μs.**

describing the spectral width of a laser. Since increasing the optical pump power increases the oscillation power, in this regime the line width of the LIMO decreases as the optical pump power increases. On the other hand, at high pump powers where the pump laser's relative intensity noise dominates, the spectral width approaches a minimum value determined by the laser's RIN noise, as given by Eqs. (25), (27), and (34). We measured the phase noise spectrum of the LIMO and verified our theoretical findings.

It is important to note that the analysis performed here was for the specific case of the LIMO with a Mach–Zehnder electro-optic modulator. Other modulation schemes, such as with electro-absorptive modulators or by direct modulation of semiconductor lasers [29], will also lead to signals with characteristics similar to those obtained in this work. For these examples, the theoretical approach developed above is still applicable after suitable modifications. The major change required in the analysis is the replacement of Eq. (1), which describes the transmission characteristics of a Mach–Zehnder modulator, with the appropriate equation for the specific modulation scheme. All other equations can then be derived in the same way as described in the theory.

Because of its unique properties, the LIMO may be used in a number of applications. As a voltage-controlled oscillator (VCO) [12], it can perform all the VCO functions in both electronic and photonic applications, including generating, tracking, and cleaning RF carriers. The LIMO has the unique property of actually amplifying injected signals [12] and, thus, may be used in high-frequency carrier regeneration and signal amplification. Other important potential applications of the LIMO include high-speed clock recovery [30,31], comb and pulse generation [12], high-gain frequency multiplication, and photonic signal upconversion and downconversion [32] in photonic RF systems [33,34].

## Acknowledgments

We thank G. Lutes and C. Greenhall for many helpful discussions and M. Calhoun for lending us many key components in experiments.

## References

[1] J. Marion and W. Hornyak, *Physics for Science and Engineering*, Chapter 15, Philadelphia, Pennsylvania: Saunders College Publishing.

[2] B. van der Pol, "A Theory of the Amplitude of Free and Forced Triode Vibrations," *Radio Review*, vol. 7, pp. 701–754, 1920.

[3] B. van der Pol, "The Nonlinear Theory of Electric Oscillations," *Proceedings of the Institute of Radio Engineers*, vol. 22, no. 9, pp. 1051–1086, 1934.

[4] O. Ishihara, T. Mori, H. Sawano, and M. Nakatani, "A Highly Stabilized GaAs FED Oscillator Using a Dielectric Resonator Feedback Circuit in 9–14 GHz," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-28, no. 8, pp. 817–824, 1980.

[5] A. Siegman, *Microwave Solid State Masers*, New York: McGraw-Hill, 1964.

[6] A. Siegman, *Lasers*, Chapter 11, Mill Valley, California: University Science Books, 1986.

[7] A. Ballato, "Piezoelectric Resonators," *Design of Crystal and Other Harmonic Oscillators*, edited by B. Parzen, New York: John Wiley and Sons, pp. 66–122, 1983.

[8] W. L. Smith, "Precision Oscillators," *Precision Frequency Control*, vol. 2, edited by E. A. Gerber and A. Ballato, New York: Academic Press, pp. 45–98, 1985.

[9] J. K. Plourde and C. R. Ren, "Application of Dielectric Resonators," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-29, no. 8, pp. 754–769, 1981.

[10] M. W. Lawrence, "Surface Acoustic Wave Oscillators," *Wave Electronics*, vol. 2, pp. 199–218, 1976.

[11] X. S. Yao and L. Maleki, "High Frequency Optical Subcarrier Generator," *Electron. Letters*, vol. 30, no. 18, pp. 1525–1526, 1994.

[12] X. S. Yao and L. Maleki, "A Novel Photonic Oscillator," *The Telecommunications and Data Acquisition Progress Report 42-122, April–June 1995*, Jet Propulsion Laboratory, Pasadena, California, pp. 32–42, August 15, 1995, URL http://edms-www.jpl.nasa.gov/tda/progress_report/42-122/122K.pdf. Also in *1995 Digest of the LEOS Summer Topical Meetings, RF Optoelectronics*, IEEE catalog no. 95TH8031, Institute of Electrical and Electronics Engineering, Piscataway, New Jersey, pp. 17–18.

[13] M. Rodwell, J. E. Bowers, R. Pullela, K. Gilboney, J. Pusl, and D. Nguyen, "Electronic and Optoelectronic Components for Fiber Transmission at Bandwidths Approaching 100 GHz," *1995 Digest of the LEOS Summer Topical Meetings, RF Optoelectronics*, IEEE catalog no. 95TH8031, Institute of Electrical and Electronics Engineering, Piscataway, New Jersey, pp. 21–22.

[14] K. Noguchi, H. Miyazawa, and O. Mitomi, "75 GHz Broadband Ti:$LiNbO_3$ Optical Modulator With Ridge Structure," *Electron. Letters*, vol. 30, no. 12, pp. 949–951, 1994.

[15] A. Neyer and E. Voges, "Nonlinear Electrooptic Oscillator Using an Integrated Interferometer," *Optics Communications*, vol. 37, pp. 169–174, 1980.

[16] A. Neyer and E. Voges, "Dynamics of Electrooptic Bistable Devices With Delayed Feedback," *IEEE Journal Quantum Electron.*, vol. QE-18, no. 12, pp. 2009–2015, 1982.

[17] H. F. Schlaak and R. T. Kersten, "Integrated Optical Oscillators and Their Applications to Optical Communication Systems," *Optics Communications*, vol. 36, no. 3, pp. 186–188, 1981.

[18] H. M. Gibbs, F. A. Hopf, D. L. Kaplan, M. W. Derstine, and R. L. Shoemaker, "Periodic Oscillation and Chaos in Optical Bistability: Possible Guided Wave All Optical Square-Wave Oscillators," *SPIE Proceedings, Integrated Optics and Millimeter and Microwave Integrated Circuits*, vol. 317, pp. 297–304, 1981.

[19] E. Garmire, J. H. Marburger, S. D. Allen, and H. G. Winful, "Transient Response of Hybrid Bistable Optical Devices," *Appl. Phys. Letters*, vol. 34, no. 6, pp. 374–376, 1979.

[20] A. Neyer and E. Voges, "High-Frequency Electro-Optic Oscillator Using an Integrated Interferometer," *Appl. Phys. Letters*, vol. 40, no. 1, pp. 6–8, 1982.

[21] T. Aida and P. Davis, "Applicability of Bifurcation to Chaos: Experimental Demonstration of Methods for Switching Among Multistable Modes in a Nonlinear Resonator," *OSA Proceedings on Nonlinear Dynamics in Optical Systems*, vol. 7, edited by N. B. Abraham, E. Garmire, and P. Mandel, Washington, DC: Optical Society of America, pp. 540–544, 1990.

[22] M. F. Lewis, "Some Aspects of Saw Oscillators," *Proceedings of 1973 Ultrasonics Symposium*, IEEE, pp. 344–347, 1973.

[23] A. L. Schawlow and C. H. Townes, "Infrared and Optical Masers," *Phys. Rev.*, vol. 112, no. 6, pp. 1940–1949, 1958.

[24] L. S. Culter and C. L. Searle, "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards," *Proceedings of the IEEE*, vol. 54, no. 2, pp. 136–154, 1966.

[25] A. Yariv, *Introduction to Optical Electronics*, 2nd ed., Chapter 10, New York: Holt, Rinehart and Winston, 1976.

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

[27] X. S. Yao and L. Maleki, "Field Demonstration of X-Band Photonic Antenna Remoting in the Deep Space Network," *The Telecommunications and Data Acquisition Progress Report 42-117, January-March 1994*, Jet Propulsion Laboratory, Pasadena, California, pp. 29–34, May 15, 1994.

[28] *Phase Noise Characterization of Microwave Oscillators—Frequency Discriminator Method*, Hewlett-Packard Co. Product Note 117 29C-2, Palo Alto, California.

[29] M. F. Lewis, "Novel RF Oscillator Using Optical Components," *Electron. Letters*, vol. 28, no. 1, pp. 31–32, 1992.

[30] X. S. Yao and G. Lutes, "A High Speed Photonic Clock and Carrier Regenerator," *The Telecommunications and Data Acquisition Progress Report 42-121, January-March 1995*, Jet Propulsion Laboratory, Pasadena, California, pp. 202–209, May 15, 1995, URL http://edms-www.jpl.nasa.gov/tda/progress_report/42-121/121O.pdf.

[31] X. S. Yao, G. Lutes, L. Maleki, and S. Cao, "A Novel Photonic Clock and Carrier Recovery Device," *Wireless Communications*, SPIE Proceedings vol. 2556, San Diego, California, pp. 118–127, 1995.

[32] G. K. Gopalakrishnan, W. K. Burns, and C. H. Bulmer, "Microwave-Optical Mixing in $LiNbO_3$ Modulators," *IEEE Trans. Microwave Theory Techn.*, vol. 41, no. 12, pp. 2383–2391, 1993.

[33] H. Ogawa, D. Polifko, and S. Banba, "Millimeter-Wave Fiber Optics Systems for Personal Radio Communication," *IEEE Trans. Microwave Theory Techn.*, vol. 40, no. 12, pp. 2285–2293, 1992.

[34] P. Herczfeld and A. Daryoush, "Fiber Optic Feed Network for Large Aperture Phased Array Antennas," *Microwave Journal*, pp. 160–166, August 1987.

[35] R. L. Byer, "Diode Laser-Pumped Solid-State Lasers," *Science*, vol. 239, pp. 742–747, 1988.

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

TDA Progress Report 42-123

November 15, 1995

6340

P 30

# Optimum Detection of Tones Transmitted by a Spacecraft

M. K. Simon and M. M. Shihabi
Communications Systems and Research Section

T. Moon[1]
Utah State University, Logan, Utah

*The performance of a scheme proposed for automated routine monitoring of deep-space missions is presented. The scheme uses four different tones (sinusoids) transmitted from the spacecraft (S/C) to a ground station with the positive identification of each of them used to indicate different states of the S/C. Performance is measured in terms of detection probability versus false alarm probability with detection signal-to-noise ratio as a parameter. The cases where the phase of the received tone is unknown and where both the phase and frequency of the received tone are unknown are treated separately. The decision rules proposed for detecting the tones are formulated from average-likelihood ratio and maximum-likelihood ratio tests, the former resulting in optimum receiver structures.*

## I. Introduction

It has been proposed that automated routine monitoring of deep-space missions be provided by transmitting one out of $n$ (typically $n = 4$) different subcarriers (tones) from the spacecraft (S/C) and then using a small automated terminal (for example, a 6-m low Earth orbiter terminal (LEO-T)-class) ground station to detect the presence or absence of each possible tone. The positive identification of each of the tones at the receiver will indicate different stages of the S/C, for example, S/C healthy, S/C needs help, S/C is going to transmit telemetry, etc. Since each of these tones is transmitted from the S/C to the ground over an additive white Gaussian noise (AWGN) channel along with the added possibility of Doppler distortion, the above-mentioned detection problem to be solved at the receiver can be formulated as a binary hypotheses test of signal plus noise versus noise only. In the most general case, the signal is modeled as a constant power sinusoid with unknown [i.e., uniformly distributed on $(-\pi, \pi)$] phase and unknown (i.e., uniformly distributed in some interval $(f_1, f_2)$ governed by the amount of Doppler) frequency.

The optimum solution to problems of this nature is based upon maximum-likelihood (ML) considerations of the type discussed in VanTrees [1]. In particular, the solution takes the form of a binary hypothesis likelihood ratio test against a threshold whose value depends on the specified false alarm and detection probabilities, the available signal power-to-noise spectral density ratio, and the duration of

[1] This work was performed under a NASA Summer Faculty Fellowship at the Jet Propulsion Laboratory, Communications Systems and Research Section.

the observation of the hypotheses. We shall see that there are, in principle, two detection/estimation philosophies suggested by the ML approach, corresponding respectively to what is commonly known as *noncoherent* detection, wherein no attempt is made to estimate the unknown parameters prior to detection, and *pseudocoherent* detection, wherein an attempt is made to first estimate the parameters (using an ML approach) and then to use these estimates to aid in the detection process [2]. Since there appears to be some question about which is the better approach, we shall consider both approaches, discuss their philosophical differences, and compare their performances.

This article is organized in two parts. In Part 1, we consider the problem of optimally detecting a sinusoidal signal of known amplitude (power) and frequency but of unknown phase [i.e., uniformly distributed on $(-\pi, \pi)$] transmitted from a S/C to the ground over an AWGN channel. In so far as the optimum receiver design is concerned, the problem will be formulated as a binary hypothesis test of signal plus noise versus noise only with a single unknown parameter (i.e., carrier phase). In Part 2, we consider the added possibility of Doppler distortion, which produces an uncertainty in the received carrier frequency. Once again, the problem can be formulated as a binary hypothesis test of signal plus noise versus noise only, where now the signal is modeled as a constant power sinusoid with unknown phase and unknown frequency. Unfortunately, however, the theory for this case is not as well developed in [1] as for the case where frequency is known. Nevertheless, other researchers [3–6] have examined problems of this type in the context of frequency-hopped (FH) or direct sequence (DS) spread spectrum communication systems, and we shall make use of their results wherever appropriate.

# Part 1. Known Frequency and Unknown Phase

## II. The Average-Likelihood Ratio Test

### A. Derivation of the Optimum Decision Rule and Associated Receiver Structure

Consider the transmission of a fixed (known) amplitude sinusoid with known frequency and unknown phase over an AWGN channel. As such, the received signal can be modeled over the interval of observation $0 \leq t \leq T$ as corresponding to either of two hypotheses, namely,

$$r(t) = s(t, \theta) + n(t) = \sqrt{2P}\cos(\omega_c t + \theta) + n(t) \tag{1a}$$

when indeed the signal was sent (hypothesis $H_1$) or

$$r(t) = n(t) \tag{1b}$$

when the signal was not sent (hypothesis $H_0$). In Eq. (1a), $P, \omega_c$ respectively denote the known signal power and radian carrier frequency, and $\theta$ denotes the unknown carrier phase assumed to be uniformly distributed in the interval $(-\pi, \pi)$. Also, $n(t)$ denotes the AWGN with single-sided power spectral density $N_0$ W/Hz.

The optimum detection of a signal transmitted over an AWGN channel is the solution to the problem of finding the likelihood ratio (LR), defined as the ratio of the conditional probability density functions (pdf's) of the received signal under the two hypotheses, namely,

$$\Lambda(r(t)) \triangleq \frac{p\,(r(t)|H_1)}{p\,(r(t)|H_0)} \tag{2}$$

and then comparing this ratio to a suitable chosen threshold to decide between $H_1$ and $H_0$, i.e.,

$$\Lambda(r(t)) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{3}$$

In the case where *all* parameters of the signal are known, the evaluation of the numerator and denominator of Eq. (2) is straightforward, namely,

$$\begin{aligned} p\left(r(t)|H_1\right) &= \frac{1}{\sqrt{\pi N_0}} \exp\left\{-\frac{1}{N_0}\int_0^T (r(t)-s(t))^2 dt\right\} \\ p\left(r(t)|H_0\right) &= \frac{1}{\sqrt{\pi N_0}} \exp\left\{-\frac{1}{N_0}\int_0^T r^2(t) dt\right\} \end{aligned} \tag{4}$$

When the signal has an unknown parameter, e.g., the phase $\theta$, then to compute the numerator of Eq. (3), we must first condition the pdf $p\left(r(t)|H_1\right)$ on the unknown parameter ($\theta$) and then average over this parameter, i.e.,

$$p\left(r(t)|H_1\right) = \int_{-\pi}^{\pi} p\left(r(t)|H_1,\theta\right) p_\theta(\theta) d\theta \tag{5}$$

where $p_\theta(\theta)$ denotes the pdf of the unknown parameter $\theta$. In our situation, the phase is assumed to be completely unknown and, thus, $p_\theta(\theta)$ is a uniform distribution. Also note that this conditioning on the unknown parameter is now necessary in the denominator of Eq. (3) since the signal does not explicitly appear in $p(r(t)|H_0)$ [see Eq. (4)]. Hence, combining Eqs. (3) through (5), the average-likelihood ratio (ALR)[2] becomes

$$\begin{aligned} \Lambda(r(t)) &= \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi} p(r(t)|H_1,\theta) d\theta}{p(r(t)|H_0)} = \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{\sqrt{\pi N_0}} \exp\left\{-\frac{1}{N_0}\int_0^T (r(t)-s(t,\theta))^2 dt\right\} d\theta}{\frac{1}{\sqrt{\pi N_0}} \exp\left\{-\frac{1}{N_0}\int_0^T r^2(t) dt\right\}} \\ &= \exp\left\{-\frac{PT}{N_0}\right\} \frac{1}{2\pi}\int_{\pi}^{\pi} \exp\left\{\frac{2}{N_0}\int_0^T r(t)s(t,\theta) dt\right\} d\theta \\ &= \exp\left\{-\frac{PT}{N_0}\right\} \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left\{\frac{2\sqrt{2P}}{N_0}\int_0^T r(t)\cos(\omega_c t+\theta) dt\right\} d\theta \end{aligned} \tag{6}$$

[2] We shall refer to this formulation as an *average*-likelihood ratio (ALR) test to distinguish it from another (in general, less optimum) approach to be discussed shortly, in which a best (maximum-likelihood) estimate of the unknown parameter is obtained first and then used in the detection process. We shall refer to the latter approach as a *maximum*-likelihood ratio (MLR) test. This vernacular is not standard in the literature. What is important to understand here is that the words *average* and *maximum* refer to the manner in which the unknown parameter is handled, i.e., the *estimation* part of the problem and not the manner in which the decision on the hypothesis is made, i.e., the *detection* part of the problem. We shall be more explicit and mathematically precise about these differences later on in the article.

In arriving at the final result in Eq. (6), we have noted that the term $\exp\left\{-\int_0^T r^2(t)dt/N_0\right\}$ is common to both the numerator and denominator and, thus, cancels, and also that

$$\exp\left\{-\frac{1}{N_0}\int_0^T s^2(t,\theta)dt\right\} = \exp\left\{-\frac{2P}{N_0}\int_0^T \cos^2(\omega_c t+\theta)dt\right\} = \exp\left\{-\frac{PT}{N_0}\right\} \tag{7}$$

assuming $\omega_c T >> 1$, as is typically the case. Defining the in-phase (I) and quadrature (Q) correlations

$$L_c \triangleq \int_0^T r(t)\sqrt{2}\cos\omega_c t dt$$

$$L_s \triangleq \int_0^T r(t)\sqrt{2}\sin\omega_c t dt$$

then Eq. (6) can be rewritten as

$$\Lambda(r(t)) = \exp\left\{-\frac{PT}{N_0}\right\}\frac{1}{2\pi}\int_{-\pi}^{\pi}\exp\left\{\frac{2\sqrt{P}}{N_0}L\cos(\theta+\alpha)\right\}d\theta = \exp\left\{-\frac{PT}{N_0}\right\}I_0\left(\frac{2\sqrt{P}}{N_0}L\right) \tag{8}$$

where

$$\begin{aligned} L &\triangleq \sqrt{L_c^2+L_s^2} \\ \alpha &\triangleq \tan^{-1}\frac{L_s}{L_c} \end{aligned} \tag{9}$$

Comparing $\Lambda(r(t))$ to a threshold $\eta$ is equivalent to comparing $\ln\Lambda(r(t))$ to $\ln\eta$. Thus, taking the natural logarithm of Eq. (8), we obtain the equivalent decision rule

$$\ln I_0\left(\frac{2\sqrt{P}}{N_0}L\right) \underset{H_0}{\overset{H_1}{\gtrless}} \ln\eta + \frac{PT}{N_0} \tag{10}$$

Finally, since $\ln I_0(x)$ is a monotonic function of its argument, $x$, and since $PT/N_0$ can be absorbed into the decision threshold, then the decision rule of Eq. (10) can be further simplified to

$$\frac{2\sqrt{P}}{N_0}L \underset{H_0}{\overset{H_1}{\gtrless}} \xi \tag{11}$$

or, equivalently,

$$L^2 \underset{H_0}{\overset{H_1}{\gtrless}} \xi^2 \frac{N_0^2}{4P} \triangleq \gamma \tag{12}$$

i.e., the optimum decision of signal present versus signal absent is determined from a comparison of the output of a square-law envelope detector with a normalized threshold, $\gamma$, whose value is determined from the specifications on false alarm probability and detection probability (see the next section). An implementation of the decision rule in Eq. (12) is illustrated in Fig. 1.

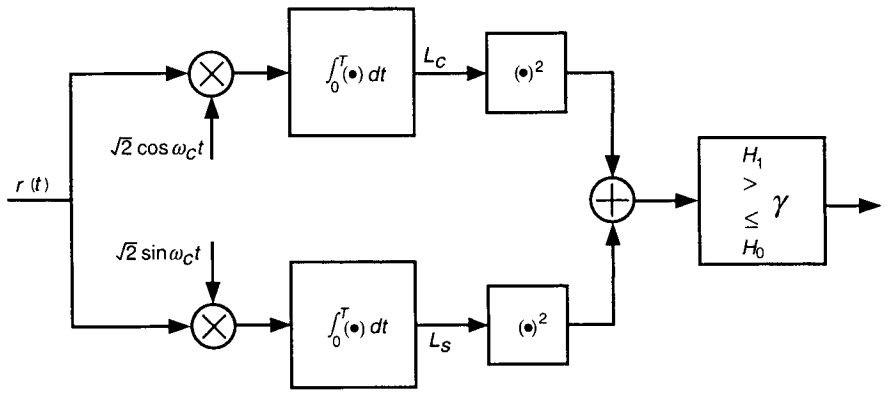


**Fig. 1. Average-likelihood (noncoherent) detector for detection of a single sinusoidal tone with known frequency and unknown phase in AWGN.**

## B. Performance (Receiver Operating Characteristic)

The performance of the receiver in Fig. 1 is described in terms of its *false alarm probability* ($P_F$), defined as the probability of deciding $H_1$ (signal is present) when indeed $H_0$ is true (signal is absent), and its *probability of detection* ($P_D$), defined as the probability of deciding $H_0$ (signal is absent) when indeed $H_1$ is true (signal is present). These probabilities are readily computed from knowledge of the first and second moments of the Gaussian random variables $L_c$ and $L_s$ [see Eq. (8)] under the two hypotheses, namely,

$$\left.\begin{aligned} H_0: \quad & E\{L_c\} = E\{L_s\} = 0 \\ & \text{var}\,\{L_c\} = \text{var}\,\{L_s\} = \frac{N_0 T}{2} \\ H_1: \quad & E\{L_c|\theta\} = \sqrt{PT}\cos\theta \\ & E\{L_s|\theta\} = -\sqrt{PT}\sin\theta \\ & \text{var}\,\{L_c\} = \text{var}\,\{L_s\} = \frac{N_0 T}{2} \end{aligned}\right\} \tag{13}$$

To compute $P_F$, we observe that, under hypotheses $H_0$, $L$ is a Rayleigh random variable ($L^2$ is a central chi-squared random variable). Thus,

$$P_F = Pr\{H_1|H_0\} = Pr\{L^2 > \gamma|H_0\} = \int_0^{2\pi}\int_{\sqrt{\gamma}}^{\infty} \frac{1}{2\pi(N_0T/2)} L\exp\left(-\frac{L^2}{N_0T}\right) dLd\theta$$

$$= \int_{\sqrt{\gamma/N_0T}}^{\infty} 2R\exp(-R^2)dR = \exp\left(-\frac{\gamma}{N_0T}\right) \tag{14}$$

Similarly, we observe that, under hypothesis $H_1$, $L$ is a Rician random variable ($L^2$ is a noncentral chi-squared random variable). Thus,

$$\left.\begin{aligned} P_D &= Pr\{H_1|H_1\} = Pr\{L^2 > \gamma|H_1\} = \int_{\sqrt{\gamma}}^{\infty} \frac{1}{N_0T/2} L\exp\left(-\frac{L^2+\beta^2}{N_0T}\right) I_0\left(\frac{2L\beta}{N_0T}\right) dL \\ \beta^2 &\triangleq (E\{L_c|\theta\})^2 + (E\{L_s|\theta\})^2 = PT^2 \\ &= \int_{\sqrt{2\gamma/N_0T}}^{\infty} R\exp\left(-\frac{R^2+d^2}{2}\right) I_0(Rd)dR = Q\left(d, \sqrt{\frac{2\gamma}{N_0T}}\right) \end{aligned}\right\} \tag{15}$$

where

$$d^2 \triangleq \frac{2PT}{N_0} = \frac{2E}{N_0} \tag{16}$$

is the detection signal-to-noise ratio (SNR) and $Q(\alpha, \beta)$ is the Marcum Q-function defined by [1]:

$$Q(\alpha, \beta) = \int_{\beta}^{\infty} z\exp\left(-\frac{z^2+\alpha^2}{2}\right) I_0(\alpha z)dz \tag{17}$$

Combining Eqs. (14) and (15) and eliminating the normalized detection threshold, one obtains the receiver operating characteristic (ROC) given by

$$P_D = Q\left(d, \sqrt{-2\ln P_F}\right) \tag{18}$$

which is illustrated in Fig. 2 for several values of the parameter $d$ (or, equivalently, $E/N_0$). Alternatively, $P_D$ is plotted versus $d^2$ with $P_F$ as a parameter in Fig. 3.

## III. The Maximum-Likelihood Ratio Test

### A. Derivation of the Optimum Decision Rule and Associated Receiver Structure

Although the *exact* evaluation of the numerator of the likelihood ratio in Eq. (2), i.e., $p(r(t)|H_1)$ is obtained from the law of conditional probability as described by Eq. (5), namely, conditioning on the unknown parameter and averaging its distribution, it is also possible to approximate this numerator by first finding the ML estimate of the unknown parameter and then substituting this value into the conditional probability $p(r(t)|H_1, \theta)$. That is, we approximate $p(r(t)|H_1)$ by $p\left(r(t)|H_1, \hat{\theta}_{ML}\right)$, in which case the likelihood ratio test (now referred to as the *maximum*-likelihood ratio (MLR) test) becomes

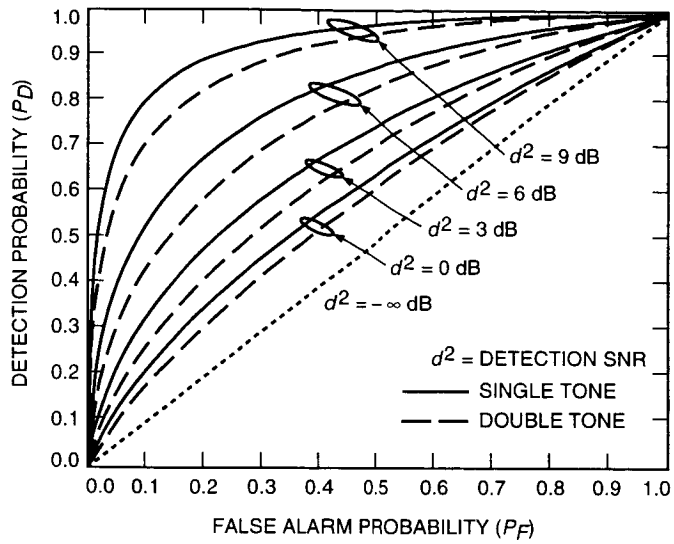


**Fig. 2. ROC: frequency known and phase unknown.**

$$\Lambda(r(t)) \cong \frac{p\left(r(t)|H_1, \hat{\theta}_{ML}\right)}{p\left(r(t)|H_0\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{19}$$

We refer to this approach of first optimally estimating the phase and then using this estimate to aid the detection process as *pseudocoherent* detection. It is important at this point to emphasize that in the general context of problems of this type, i.e., detection of signals with completely unknown parameters, *the performance of a receiver derived from MLR considerations (e.g., a pseudocoherent receiver) is never better than the performance of the ALR receiver (e.g., a noncoherent receiver), which is indeed optimum under the assumed conditions.* Thus, *at best*, one could hope that the MLR receiver would perform equally well as does the ALR receiver. In the next section, we shall indeed reveal the extent to which this equality in performance can be achieved for the problem at hand. First, however, let us derive the ML estimate of phase, namely, $\hat{\theta}_{ML}$, to be used in approximating the numerator of the likelihood ratio.

The ML estimate of $\theta$ is defined as

$$\hat{\theta}_{ML} = \max_{\theta} \frac{p(r(t)|H_1, \theta)}{p(r(t)|H_0)} \tag{20}$$

Using Eq. (4) in Eq. (20), it is straightforward to show that

$$\hat{\theta}_{ML} = \max_{\theta} \exp\left\{\frac{2}{N_0}\int_0^T r(t)s(t,\theta)dt\right\} = \max_{\theta} \exp\left\{\frac{2\sqrt{2P}}{N_0}\int_0^T r(t)\cos(\omega_c t + \theta)\,dt\right\}$$

$$= \max_{\theta} \exp\left\{\frac{2\sqrt{P}}{N_0} L\cos(\theta + \alpha)\right\} \tag{21}$$

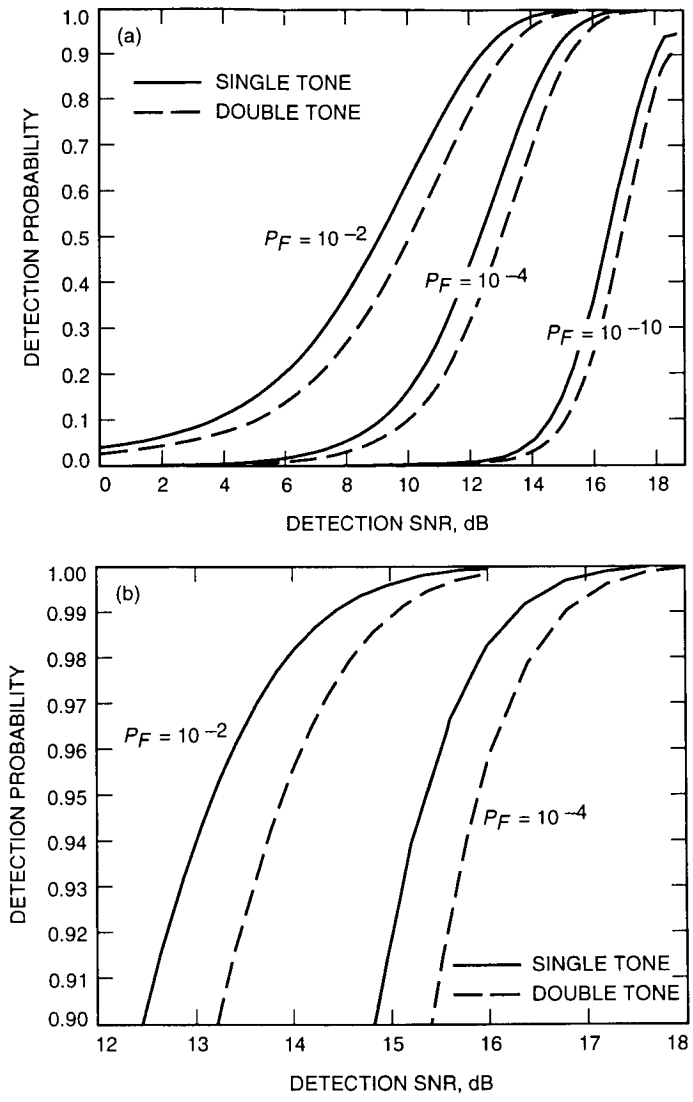


**Fig. 3. Detection probability ($P_D$) versus detection SNR ($d^2$): (a) frequency known and phase unknown and (b) frequency known and phase unknown (expanded view).**

where the envelope, $L$, and the phase, $\alpha$, are defined by Eq. (9) together with Eq. (8). Since $L$ is positive and independent of $\theta$, then the maximization required in Eq. (21) is achieved when the argument of the cosine function is equal to zero. Thus,

$$\hat{\theta}_{ML} = -\alpha \tag{22}$$

An implementation of this ML estimator of the unknown channel phase is illustrated in Fig. 4. Also illustrated in Fig. 4 is the pseudocoherent detector that employs this ML phase estimator, which can be obtained by taking the natural logarithm of Eq. (19). We now find the decision rule based on the MLR test in Eq. (19) and compare it with that of the previously discussed ALR test. Using Eq. (22) in Eq. (19) gives, by analogy with Eq. (8),

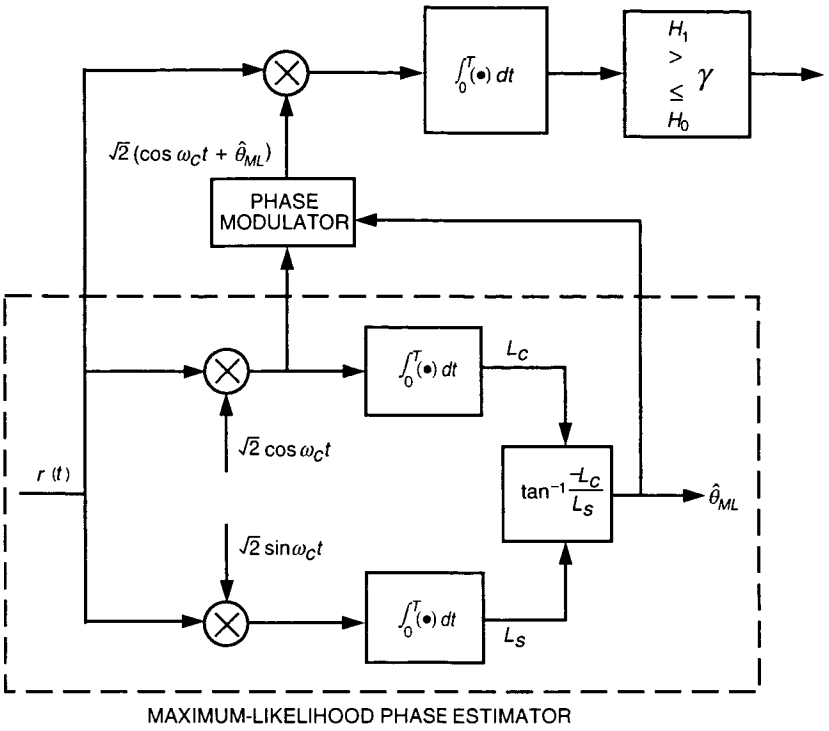


**Fig. 4. Maximum-likelihood phase estimator and pseudocoherent detector.**

$$\Lambda(r(t)) \cong \exp\left\{-\frac{PT}{N_0}\right\} \exp\left\{\frac{2\sqrt{P}}{N} L \cos\left(\hat{\theta}_{ML} + \alpha\right)\right\} = \exp\left\{-\frac{PT}{N_0}\right\} \exp\left\{\frac{2\sqrt{P}}{N_0} L\right\} \tag{23}$$

Taking the natural logarithm of Eq. (23), we then have, by analogy with Eq. (10),

$$\frac{2\sqrt{P}}{N_0} L \underset{H_0}{\overset{H_1}{\gtrless}} \ln\eta + \frac{PT}{N_0} \tag{24}$$

Since, as previously noted, the term $PT/N_0$ can be absorbed into the decision threshold, then an equivalent test to Eq. (24) is

$$\frac{2\sqrt{P}}{N_0} L \underset{H_0}{\overset{H_1}{\gtrless}} \xi \tag{25}$$

*which is identical to Eq. (11)!* Thus, we conclude that in this particular circumstance, *the MLR test (pseudocoherent receiver) and the ALR test (noncoherent receiver) are identical.* Hence, the performance of the pseudocoherent receiver is also described by Figs. 2 and 3. It is to be emphasized again that the equivalence found here between ALR and MLR receivers is not typical and applies only in this very

special case of the detection of a signal with known frequency and unknown phase. More often than not, the receiver derived from the MLR approach will have an inferior performance to the optimum one derived from the ALR approach.

## IV. A More Precise Formulation of the Problem

In reality, the subcarriers that are transmitted to indicate the status of the S/C are continuous square waves that biphase modulate the carrier. Thus, denoting the carrier radian frequency and phase by $\omega_c$ and $\theta_c$ (previously called $\theta$), respectively, and the square-wave subcarrier radian frequency and phase by $\omega_{sc}$ and $\theta_{sc}$, respectively, then the received signal analogous to Eq. (1a) is given by

$$\begin{aligned} r(t) &= s(t,\theta_c,\theta_{sc}) + n(t) = \sqrt{2P}\sin\left(\omega_c t + \theta_c + \frac{\pi}{2}\ \mathrm{Sq}\ (\omega_{sc}t + \theta_{sc})\right) + n(t) \\ &= \sqrt{2P}\ \mathrm{Sq}\ (\omega_{sc}t + \theta_{sc})\cos(\omega_c t + \theta_c) + n(t) \end{aligned} \tag{26}$$

Assuming that the harmonics with frequencies other than the sum and difference of $\omega_c$ and $\omega_{sc}$ are filtered out, then in so far as detection is concerned, we may consider the received signal to be[3]

$$r(t) = s\,(t,\theta_c,\theta_{sc}) + n(t) = \sqrt{P}\,\{\cos[(\omega_c + \omega_{sc})t + (\theta_c + \theta_{sc})] + \cos[(\omega_c - \omega_{sc})t + (\theta_c - \theta_{sc})]\} + n(t) \tag{27}$$

i.e., the problem is to detect the presence or absence of *two* tones in an AWGN background where both $\omega_c$ and $\omega_{sc}$ are assumed to be known but both $\theta_c$ and $\theta_{sc}$ are assumed to be completely unknown. For convenience of notation, we shall rewrite Eq. (27) as

$$r(t) = s(t,\theta_+,\theta_-) + n(t) = \sqrt{P}\,\{\cos[\omega_+ t + \theta_+] + \cos[\omega_- t + \theta_-]\} + n(t) \tag{28}$$

where

$$\left.\begin{aligned} \omega_\pm &\triangleq \omega_c \pm \omega_{sc} \\ \theta_\pm &\triangleq \theta_c \pm \theta_{sc} \end{aligned}\right\} \tag{29}$$

At first glance, it might appear that, because the phases $\theta_c$ and $\theta_{sc}$ appear in the two signal tones as their sum and difference, the detection of these tones cannot be performed independently. Interestingly enough, $\theta_+ \triangleq \theta_c + \theta_{sc}$ and $\theta_- \triangleq \theta_c - \theta_{sc}$ when reduced modulo $2\pi$ can be shown to be independent uniformly distributed random variables (see the Appendix). Thus, as we shall see shortly, the detection of two distinct sinusoidal tones with independent random phases in an AWGN background can be treated by a likelihood ratio approach analogous to that discussed in the previous section for a single tone in the same background.

[3] In reality, the $\sqrt{P}$ amplitude factor in Eq. (27) should be $(2\sqrt{2}/\pi)\sqrt{P} = 0.9003\sqrt{P}$ to account for the amplitude of the first harmonic in the square-wave subcarrier. For simplicity, we shall ignore this minor difference.

### A. The ALR Test

As discussed in Section II, the optimum decision rule is, in general, obtained by applying the *average*-likelihood approach, which in this case means averaging the conditional likelihood function over the *two* random phases $\theta_+$ and $\theta_-$. In particular, the conditional pdf of the received signal under hypothesis $H_1$ is analogous to Eq. (5):

$$p(r(t|H_1) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} p(r(t)|H_1,\theta_+,\theta_-)p_{\theta_+,\theta_-}(\theta_+,\theta_-)d\theta_+ d\theta_- = \left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} p(r(t)|H_1,\theta_+,\theta_-)d\theta_+ d\theta_- \tag{30}$$

and, hence, the ALR becomes

$$\begin{aligned}\Lambda(r(t)) &= \frac{\left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} p(r(t)|H_1,\theta_+,\theta_-)d\theta_+ d\theta_-}{p(r(t)|H_0)} \\ &= \exp\left\{-\frac{PT}{N_0}\right\}\left(\frac{1}{2\pi}\right)^2 \int_{-\pi}^{\pi} \exp\left\{\frac{2\sqrt{P}}{N_0}\int_0^T r(t)\cos(\omega_- t+\theta_-)dt\right\} d\theta_- \\ &\quad \times \int_{-\pi}^{\pi} \exp\left\{\frac{2\sqrt{P}}{N_0}\int_0^T r(t)\cos(\omega_+ t+\theta_+)dt\right\} d\theta_+ \end{aligned} \tag{31}$$

Defining the I and Q correlations for the sum and difference frequencies by

$$\left.\begin{aligned} L_{c\pm} &\triangleq \int_0^T r(t)\sqrt{2}\cos\omega_\pm t dt \\ L_{s\pm} &\triangleq \int_0^T r(t)\sqrt{2}\sin\omega_\pm t dt \end{aligned}\right\} \tag{32}$$

then, the likelihood function of Eq. (30) can be rewritten as

$$\Lambda(r(t)) = \exp\left\{-\frac{PT}{N_0}\right\} I_0\left(\frac{\sqrt{2P}}{N_0}L_-\right) I_0\left(\frac{\sqrt{2P}}{N_0}L_+\right) \tag{33}$$

where, analogous to Eq. (9), the envelopes corresponding to the upper and lower subcarrier tones are given by

$$L_\pm \triangleq \sqrt{L_{c\pm}^2 + L_{s\pm}^2} \tag{34}$$

Alternately, in terms of the log-likelihood function, we arrive at the decision rule

$$\ln I_0\left(\frac{\sqrt{2P}}{N_0}L_-\right) + \ln I_0\left(\frac{\sqrt{2P}}{N_0}L_+\right) \underset{H_0}{\overset{H_1}{\gtrless}} \ln\eta + \frac{PT}{N_0} \tag{35}$$

Note that now, despite the fact that $\ln I_0(x)$ is a monotonic function of its argument, $x$, we cannot directly simplify Eq. (35) to a form analogous to Eq. (11). Rather, to get such a form, one must approximate the $\ln I_0(x)$ function by its series and asymptotic forms for small and large arguments, namely,

$$\ln I_0(x) \cong \begin{cases} \dfrac{x^2}{4}, & \text{small } x \\ |x|, & \text{large } x \end{cases} \tag{36}$$

Thus, for example, if we invoke the small argument approximation of the $\ln I_0(x)$ function in Eq. (35), we get the decision rule (optimum for small SNR)

$$L_-^2 + L_+^2 \triangleq L^2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{37}$$

where $\gamma$ is again a normalized threshold [not necessarily equal to the one defined in Eq. (12)]. The decision rule in Eq. (37) suggests the ALR structure illustrated in Fig. 5, which is analogous to that given in Fig. 1. For the large argument approximation of the $\ln I_0(x)$ function, the implementation of Fig. 5 would require square root devices in each arm entering the final summer prior to the decision device.

### B. The MLR Test

Let us now again compare the noncoherent two-tone detector derived from ALR considerations and specified by the decision rule of Eq. (35) to a pseudocoherent detector that can be derived from MLR considerations. In particular, consider the joint ML estimates $\hat{\theta}_{ML+}, \hat{\theta}_{ML-}$ of $\theta_+, \theta_-$ defined as

$$\hat{\theta}_{ML+}, \hat{\theta}_{ML-} = \max_{\theta_+,\theta_-} \frac{p(r(t)|H_1, \theta_+, \theta_-)}{p(r(t)|H_0)} \tag{38}$$

which, because of the independence of $\theta_+$ and $\theta_-$, is determined as

$$\begin{aligned} \hat{\theta}_{ML+}, \hat{\theta}_{ML-} &= \max_{\theta_+,\theta_-} \exp\left\{\frac{2\sqrt{P}}{N_0}\int_0^T r(t)\cos(\omega_- t + \theta_-)dt\right\} \exp\left\{\frac{2\sqrt{P}}{N_0}\int_0^T r(t)\cos(\omega_+ t + \theta_+)dt\right\} \\ &= \max_{\theta_+,\theta_-} \left\{\frac{\sqrt{2P}}{N_0}L_-\cos(\theta_- + \alpha_-)\right\} \exp\left\{\frac{\sqrt{2P}}{N_0}L_+\cos(\theta_+ + \alpha_+)\right\} \end{aligned} \tag{39}$$

where $\alpha_\pm$ are defined in terms of $L_\pm$, analogous to Eq. (9). The solution to Eq. (39) is

$$\hat{\theta}_{ML\pm} = -\alpha_\pm \tag{40}$$

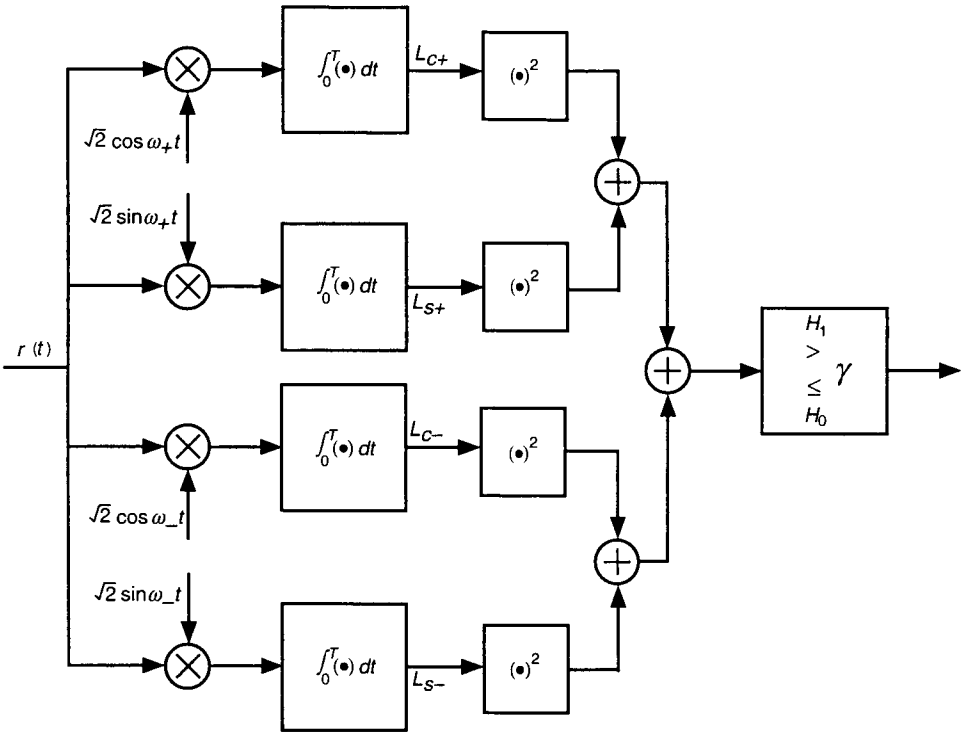


**Fig. 5. Average-likelihood (noncoherent) detector for detection of a pair of independent sinusoidal tones with known frequencies and unknown phases in AWGN.**

which, when substituted in Eq. (39), gives

$$\Lambda(r(t)) \cong \exp\left\{-\frac{PT}{N_0}\right\} \exp\left\{\frac{\sqrt{2P}}{N_0} L_+\right\} \exp\left\{\frac{\sqrt{2P}}{N_0} L_-\right\} = \exp\left\{-\frac{PT}{N_0}\right\} \exp\left\{\frac{2\sqrt{P}}{N_0} L\right\} \tag{41}$$

Taking the natural logarithm of Eq. (40), we get the decision rule

$$\frac{\sqrt{2P}}{N_0}(L_+ + L_-) \underset{H_0}{\overset{H_1}{\gtrless}} \ln\eta + \frac{PT}{N_0} \tag{42}$$

Comparing Eq. (42) with Eq. (35), we observe that, in the two-tone case, the MLR test (which would lead to a pseudocoherent form of detector analogous to Fig. 4) is *not* the same as the ALR test. However, using the large argument approximation of the $\ln I_0(x)$ function as given by Eq. (36), we see that the ALR and MLR tests once again become equivalent. In summary then, we observe that, *for detection of a single tone in AWGN, the ALR (noncoherent) test and MLR (pseudocoherent) test are equivalent for all SNRs, whereas for the detection of a pair of equal power tones in AWGN, the ALR and MLR tests are equivalent only at sufficiently large SNR.*

## C. Performance (Receiver Operating Characteristic)

The performance of the low SNR receiver in Fig. 5 is, as before, described in terms of its false alarm probability ($P_F$) and its probability of detection ($P_D$). These probabilities are readily computed from

knowledge of the first and second moments of the Gaussian random variables $L_{c\pm}$ and $L_{s\pm}$ [see Eq. (31)] under the two hypotheses, namely,

$$\left.\begin{aligned} H_0: \quad & E\{L_{c\pm}\} = E\{L_{s\pm}\} = 0 \\ & \text{var}\,\{L_{c\pm}\} = \text{var}\,\{L_{s\pm}\} = \frac{N_0 T}{2} \\ H_1: \quad & E\{L_{c\pm}|\theta\} = \sqrt{\frac{P}{2}} T \cos\theta_{\pm} \\ & E\{L_{s\pm}|\theta\} = \sqrt{\frac{P}{2}} T \sin\theta_{\pm} \\ & \text{var}\,\{L_{c\pm}\} = \text{var}\,\{L_{s\pm}\} = \frac{N_0 T}{2} \end{aligned}\right\} \tag{43}$$

To compute $P_F$, we observe as before that, under hypothesis $H_0$, $L^2$ is a central chi-squared random variable (now with two more degrees of freedom). Thus,

$$P_F = \Pr\{H_1|H_0\} = \Pr\{L^2 > \gamma|H_0\} = \int_{\gamma/N_0T}^{\infty} r\exp(-r)dr = \left(1 + \frac{\gamma}{N_0T}\right)\exp\left(-\frac{\gamma}{N_0T}\right) \tag{44}$$

Similarly, we observe that, under hypothesis $H_1$, $L^2$ is a noncentral chi-squared random variable (now with two more degrees of freedom). Thus,

$$P_D = \Pr\{H_1|H_1\} = \Pr\{L^2 > \gamma|H_1\} = \int_{\sqrt{2\gamma/N_0T}}^{\infty} R\left(\frac{R}{d}\right)\exp\left(-\frac{R^2+d^2}{2}\right) I_1(Rd)dR = Q_2\left(d, \sqrt{\frac{2\gamma}{N_0T}}\right) \tag{45}$$

where $d^2$ is the detection SNR defined as before [see Eq. (16)] and $Q_M(\alpha,\beta)$ is the generalized Marcum Q-function defined by

$$Q_M(\alpha,\beta) = \int_{\beta}^{\infty} z\left(\frac{z}{\alpha}\right)^{M-1}\exp\left(-\frac{z^2+\alpha^2}{2}\right) I_{M-1}(\alpha z)dz \tag{46}$$

Note that $Q_M(\alpha,\beta)$ can be obtained from $Q(\alpha,\beta) \triangleq Q_1(\alpha,\beta)$ by the relation [2, Appendix 5A]

$$Q_M(\alpha,\beta) = Q(\alpha,\beta) + \exp\left(\frac{\alpha^2+\beta^2}{2}\right)\sum_{j=1}^{M-1}\left(\frac{\beta}{\alpha}\right)^j I_j(\alpha\beta)$$

Unfortunately, the normalized detection threshold cannot be explicitly eliminated in Eqs. (39) and (40) to give a closed-form expression for the receiver operating characteristic (ROC) analogous to Eq. (18). However, for any range of interest, the ROC can be determined numerically. Such numerical results are

superimposed on the single-tone detection in Figs. 2, 3(a), and 3(b). We observe that the performance penalty associated with using a pair of subcarrier tones each with half the total power relative to the full-power single carrier tone case is quite small, e.g., on the order of 0.4 dB or less for $P_F = 10^{-2}$ and on the order of 0.3 dB or less for $P_F = 10^{-4}$. The degradation associated with the true optimum ALR scheme as described by the decision rule of Eq. (35) would be even smaller. Thus, the performance curves of the true optimum ALR scheme for two tones would lie between the solid and dashed curves in Figs. 2, 3(a), and 3(b) since indeed these performance results cannot beat those corresponding to the single-tone case. Because of the small degradations involved, we choose not to simulate the true optimum case.

# Part 2. Unknown Frequency and Unknown Phase

## V. The Average-Likelihood Ratio Test

### A. Derivation of the Optimum Decision Rule and Associated Receiver Structure

Consider the transmission of a fixed (known) amplitude sinusoid with unknown frequency and unknown phase over an AWGN channel. Analogous to Eq. (1), the received signal can be modeled over the interval of observation $0 \leq t \leq T$ as corresponding to either of two hypotheses, namely,

$$r(t) = s(t,\theta) + n(t) = \sqrt{2P}\cos(\omega t + \theta) + n(t) \tag{47a}$$

when indeed the signal was sent (hypothesis $H_1$) or

$$r(t) = n(t) \tag{47b}$$

when the signal was not sent (hypothesis $H_0$). In addition to the previously defined parameters, in Eq. (47a), $f \triangleq \omega/2\pi$ denotes the unknown carrier frequency assumed to be uniformly distributed in the interval $(f_c - B/2, f_c + B/2)$, where as before $f_c$ denotes the nominal carrier frequency (i.e., in the absence of Doppler). When the signal has two unknown parameters, e.g., the phase $\theta$ and frequency $f$, then to compute the numerator of Eq. (3), we must first condition the pdf $p(r(t)|H_1)$ on *both* of the unknown parameters and then average over them, i.e.,

$$p(r(t)|H_1) = \int_{f_c-B/2}^{f_c+B/2}\int_{-\pi}^{\pi} p(r(t)|H_1,\theta,f)p_\theta(\theta)p_f(f)d\theta df \tag{48}$$

where $p_\theta(\theta), p_f(f)$ respectively denote the pdf's of the unknown parameters $\theta$ and $f$. In our situation, the phase and frequency are assumed to be completely unknown, and thus $p_\theta(\theta)$ and $p_f(f)$ are uniform distributions. Hence, combining Eqs. (3) and (48), the average-likelihood ratio (ALR) becomes

$$\begin{aligned}
\Lambda(r(t)) &= \frac{\dfrac{1}{2\pi B}\displaystyle\int_{f_c-B/2}^{f_c+B/2}\int_{-\pi}^{\pi} p(r(t)|H_1,\theta,f)d\theta df}{p(r(t)|H_0)} \\
&= \exp\left\{-\frac{PT}{N_0}\right\}\frac{1}{2\pi B}\int_{f_c-B/2}^{f_c+B/2}\int_{-\pi}^{\pi}\exp\left\{\frac{2\sqrt{2P}}{N_0}\int_0^T r(t)\cos(\omega_c t+\theta)dt\right\}d\theta df \\
&= \exp\left\{-\frac{PT}{N_0}\right\}\frac{1}{B}\int_{f_c-B/2}^{f_c+B/2} I_0\left(\frac{2\sqrt{P}}{N_0}L(f)\right)df
\end{aligned} \tag{49}$$

where

$$L(f) \triangleq \sqrt{L_c^2(f) + L_s^2(f)} \tag{50}$$

with

$$\left.\begin{aligned} L_c(f) &\triangleq \int_0^T r(t)\sqrt{2}\cos 2\pi f t dt \\ L_s(f) &\triangleq \int_0^T r(t)\sqrt{2}\sin 2\pi f t dt \end{aligned}\right\} \tag{51}$$

It should be noted that $L(f)$ is nothing more than the magnitude of the complex Fourier transform (FT) of $r(t)$ in the interval $0 \leq t \leq T$. If $r(t)$ is band limited to $W$ Hz, then for large $WT$, the real and imaginary components of this complex FT, namely, $L_c(f), L_s(f)$ can be approximated by the discrete Fourier transforms (DFTs)

$$\left.\begin{aligned} L_c(f) &= \sqrt{2}\frac{1}{2W}\sum_{n=1}^{2WT} r\left(\frac{n}{2W}\right)\cos\left(2\pi f\frac{n}{2W}\right) \\ L_s(f) &= \sqrt{2}\frac{1}{2W}\sum_{n=1}^{2WT} r\left(\frac{n}{2W}\right)\sin\left(2\pi f\frac{n}{2W}\right) \end{aligned}\right\} \tag{52}$$

Comparing $\Lambda(r(t))$ to a threshold produces (after suitable normalization) the ALR decision rule

$$\int_{f_c-B/2}^{f_c+B/2} I_0\left(\frac{2\sqrt{P}}{N_0}L(f)\right) df \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{53}$$

Since Eq. (53) is overly demanding to implement, one discretizes the frequency uncertainty interval into $G = B/T^{-1} = BT$ subintervals to each of which is associated a candidate frequency $f_i; i = 0, 1, 2, \cdots, G-1$ located at its center. As such, the integration over the continuous uncertainty region in Eq. (53) is approximated by a discrete (Riemann) sum and, hence, the approximate decision rule becomes

$$\sum_{i=0}^{G-1} I_0\left(\frac{2\sqrt{p}}{N_0}L(f_i)\right) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{54}$$

which has the implementation representation of Fig. 6. It is important to understand that spacing the frequencies $f_i; i = 0, 1, 2, \cdots, G-1$ by $1/T$ guarantees independence of the noise components that appear at the output of each spectral estimate channel. However, orthogonality of the signal components of these same outputs depends on the true value of the received frequency relative to the discretized frequencies $f_i; i = 0, 1, 2, \cdots, G-1$ assumed for implementation of the receiver. That is, if the true received frequency happens to fall on one of the $f_i$'s, then a signal component will appear only in the corresponding spectral

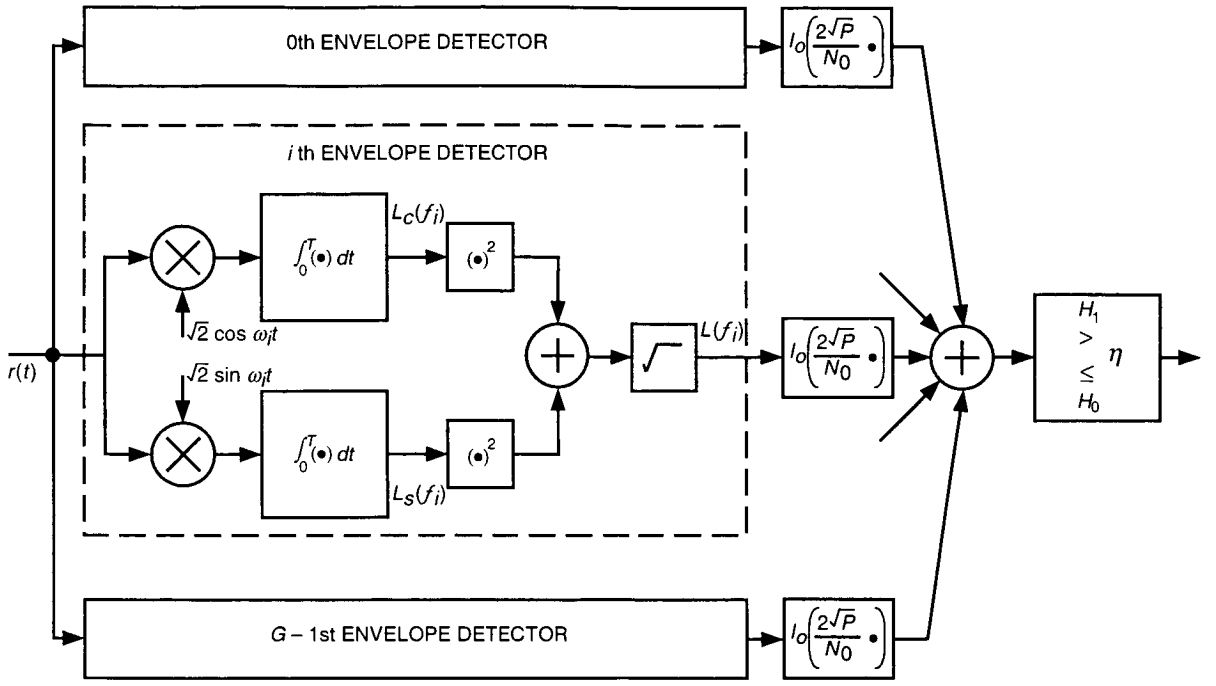


**Fig. 6. Average-likelihood (noncoherent) detector for detection of a single sinusoidal tone with unknown frequency and unknown phase in AWGN.**

estimate channel, i.e., all other channels will be noise only. On the other hand, if the true received frequency falls somewhere between two of the $f_i$'s, then we have a loss of orthogonality in that a spillover of signal energy occurs in the neighboring spectral estimates. The worst-case spillover would occur when the true received frequency is midway between two of the $f_i$'s.

We conclude this section by noting that a decision metric similar to Eq. (54) arises in the study of FH or DS/low probability of intercept (LPI) optimum ALR (noncoherent) detection [3–5], where in the FH case, $f_i; i - 0, 1, 2, \cdots, G - 1$ corresponds to the $G$ possible frequencies that the transmitted signal can hop to and the detection is based on observation of a single hop of duration $T_H = T$, and in the DS case $G$ is the number of possible code sequences that can occur in the observation interval. Many of the results obtained from these works are directly applicable to the problem at hand.

## B. Performance

It is tempting for large values of $G$ (as is typically the case) to apply a central limit theorem argument to the left side of Eq. (11), i.e., approximate it as a Gaussian random variable in so far as computing the receiver operating characteristic associated with this decision rule [4]. Unfortunately, it was shown in [5] that following such an approach is very poor when compared with results obtained from simulation or numerical methods applied to the true decision rule of Eq. (11), even for values of $G$ as large as 1000 or 10,000. In fact, it is stated in [5] that $G$ on the order of "ten thousands is not guaranteed to be large enough to validate the Gaussian approximation." Thus, to obtain the true receiver performance, we too must resort to simulation and/or numerical methods, such as those suggested by Requicha [7], wherein the characteristic function and fast Fourier transforms (FFTs) are used to compute approximate values of the distribution function associated with the left-hand side of Eq. (11). More about this later on.

## VI. The Maximum-Likelihood Ratio Test

### A. Derivation of the Optimum Decision Rule and Associated Receiver Structure

Although the *exact* evaluation of the numerator of the likelihood ratio in Eq. (2), i.e., $p(r(t)|H_1)$ is obtained from the law of conditional probability as described by Eq. (5), namely, conditioning on the unknown parameters and averaging over their distribution, it is also possible to approximate this numerator by first finding the ML estimates of the unknown parameters and then substituting these values into the conditional probability $p(r(t)|H_1, \theta, f)$. That is, we approximate $p(r(t)|H_1)$ by $p(r(t)|H_1, \hat{\theta}_{ML}, \hat{f}_{ML})$, in which case the likelihood ratio test (now referred to as the *maximum*-likelihood ratio (MLR) test) becomes

$$\Lambda(r(t)) \cong \frac{p\left(r(t)|H_1, \hat{\theta}_{ML}, \hat{f}_{ML}\right)}{p(r(t)|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \tag{55}$$

where

$$\hat{\theta}_{ML}, \hat{f}_{ML} \triangleq \max_{\theta, f} \frac{p(r(t)|H_1, \theta, f)}{p(r(t)|H_0)} \tag{56}$$

The maximization over $\theta$ required in Eq. (56) can be performed identically to that in Section III [see Eq. (23)]:

$$\max_{\theta} \frac{p(r(t)|H_1, \theta, f)}{p(r(t)|H_0)} = \exp\left(-\frac{PT}{N_0}\right) \exp\left(\frac{2\sqrt{P}}{N_0} L(f)\right) \tag{57}$$

where $L(f)$ is as defined in Eq. (50). Thus, the optimum maximum a posteriori (MAP) decision rule becomes

$$\max_{f} \exp\left(-\frac{PT}{N_0}\right) \exp\left(\frac{2\sqrt{P}}{N_0} L(f)\right) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{58}$$

Since the exponential is a monotonic function of its argument, we have the equivalent decision rule[4]

$$\max_{f} L(f) \underset{H_0}{\overset{H_1}{\gtrless}} \sqrt{\gamma} \tag{59}$$

which results in a *spectral maximum* form of receiver. Again, because of the excessive demand placed on the implementation by the need to evaluate Eq. (59) over a continuum of frequencies, we again quantize the frequency uncertainty interval into $G = BT$ subintervals, each with an associated candidate frequency

[4] We define the normalized threshold equal to $\sqrt{\gamma}$ to be consistent with the notation used in Part 1. In this way, when $G$ is equated to unity, then our results obtained here will reduce to those given in Part 1 for the MLR decision rule.



C-2.

$f_i; i = 0, 1, 2, \cdots, G-1$ located at its center. Thus, the frequency continuous decision rule of Eq. (59) can be approximated by the decision rule

$$\max_i L(f_i) \underset{H_0}{\overset{H_1}{\gtrless}} \sqrt{\gamma} \tag{60}$$

which suggests the receiver of Fig. 7. Here again, as with Eq. (54), the orthogonality of the spectral estimates is not guaranteed unless the frequency of the received signal falls on one of the $f_i$'s. Also, since $L(f_i); i = 0, 1, 2, \cdots, G$ represents a uniform sampling of $L(f)$, then in view of Eq. (52), we can implement Fig. 7 with FFT techniques.

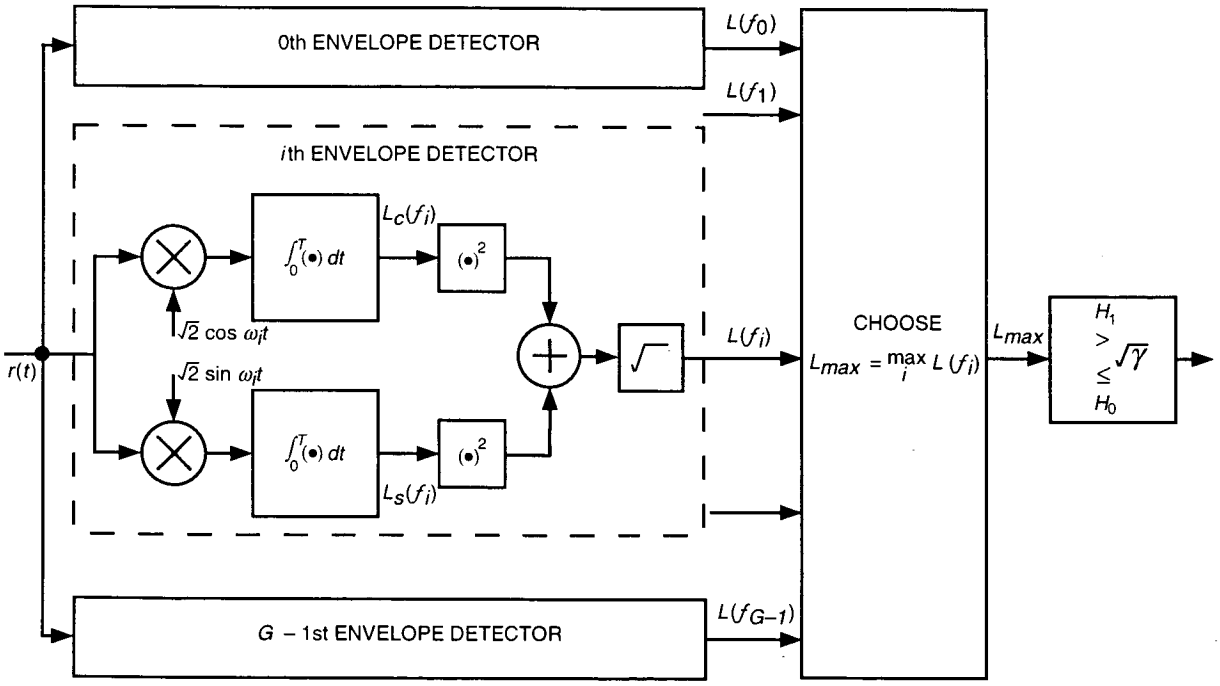


**Fig. 7. Maximum-likelihood detector for detection of a single sinusoidal tone with unknown frequency and unknown phase in AWGN.**

## B. Performance

The performance of the MLR decision rule of Eq. (60) can be obtained analytically since the pdf of $G$ independent random variables can be explicitly written in terms of the pdf's of individual random variables, which in turn are obtained from the results in Part 1. The procedure is as follows.

**1. Best-Case Performance.** Consider first the optimistic (best) case, where the actual received carrier frequency is indeed equal to one of the $G$ frequencies, say $f_k$ used to approximately implement the optimum decision rule as per the discussion following Eq. (59). Under $H_1, G-1$ of the $L(f_i)$'s are Rayleigh distributed with pdf [see Eq. (14)]

$$p_{L(f_i)}(L) = \frac{2}{N_0 T} L \exp\left(-\frac{L^2}{N_0 T}\right); \quad L \geq 0 \tag{61}$$

and the single $L(f_i)$ that is associated with the received signal carrier frequency, namely $L(f_k)$, is Rician distributed with pdf [see Eq. (15)]

$$p_{L(f_k)}(L) = \frac{2}{N_0 T} \exp\left(-\frac{L^2+\beta^2}{N_0 T}\right) I_0\left(\frac{2L\beta}{N_0 T}\right); \quad L \geq 0, \quad \beta^2 = PT^2 \tag{62}$$

Let $P_F^*$ denote the *per frequency channel* false alarm probability, i.e.,

$$P_F^* = \Pr\{L(f_i) > \sqrt{\gamma} | H_0\} = \int_{\sqrt{\gamma}}^{\infty} \frac{2}{N_0 T} L \exp\left(-\frac{L^2}{N_0 T}\right) dL = \exp\left(-\frac{\gamma}{N_0 T}\right) \tag{63}$$

which is independent of $f_i$. Then, the overall false alarm probability, $P_F$, is given by

$$\begin{aligned} P_F &= \Pr\left\{\max_i L(f_i) > \sqrt{\gamma} | H_0\right\} \\ &= 1 - \Pr\{L(f_0) \leq \sqrt{\gamma}, L(f_1) \leq \sqrt{\gamma}, \cdots, L(f_{G-1}) \leq \sqrt{\gamma} | H_0\} \\ &= 1 - \prod_{i=0}^{G-1} \Pr\{L(f_i) \leq \sqrt{\gamma} | H_0\} = 1 - \prod_{i=0}^{G-1} (1 - \Pr\{L(f_i) > \sqrt{\gamma} | H_0\}) \\ &= 1 - (1 - P_F^*)^G = 1 - \left(1 - \exp\left(-\frac{\gamma}{N_0 T}\right)\right)^G \end{aligned} \tag{64}$$

Since, under $H_1$, $G-1$ of the spectral estimates (i.e., the ones containing noise only) have the same pdf, namely Eq. (61), as under $H_0$, and one spectral estimate has the Rician pdf of Eq. (62), then the overall probability of detection, $P_d$, is determined from

$$P_D = \Pr\left\{\max_i L(f_i) > \sqrt{\gamma} | H_1\right\} = 1 - \prod_{i=0}^{G-1} (1 - \Pr\{L(f_i) > \sqrt{\gamma} | H_1\}) = 1 - (1 - P_F^*)^{G-1} (1 - P_D^*) \tag{65}$$

where $P_D^*$ corresponds to the detection probability of the single-frequency channel containing the signal, i.e.,

$$P_D^* = \Pr\{L(f_i) > \sqrt{\gamma} | H_1\} = Q\left(d, \sqrt{\frac{2\gamma}{N_0 T}}\right); \quad d^2 = \frac{2PT}{N_0} \tag{66}$$

Substituting Eqs. (63) and (65) into Eq. (66) gives

$$P_D = 1 - \left(1 - \exp\left(-\frac{\gamma}{N_0 T}\right)\right)^{G-1} \left(1 - Q\left(d, \sqrt{\frac{2\gamma}{N_0 T}}\right)\right) \tag{67}$$

or, equivalently, the overall probability of miss, $P_M$, is

$$P_M \triangleq 1 - P_D = \left(1 - \exp\left(-\frac{\gamma}{N_0 T}\right)\right)^{G-1} \left(1 - Q\left(d, \sqrt{\frac{2\gamma}{N_0 T}}\right)\right) \tag{68}$$

Note that, for $G = 1$, Eqs. (64) and (67) reduce, respectively, to Eqs. (14) and (15).

The ROC can be determined by eliminating the normalized threshold between Eqs. (64) and (67), in which case one obtains

$$P_D = 1 - (1 - P_F)^{(G-1)/G} \left(1 - Q\left(d, \sqrt{-2\ln\left(1 - (1 - P_F)^{1/G}\right)}\right)\right) \tag{69}$$

**2. Worst-Case Performance.** The worst-case performance occurs when the actual received carrier frequency is indeed midway between two of the $G$ frequencies used to approximately implement the optimum decision rule as per the discussion following Eq. (59). Under $H_0$, the false alarm performance is still described by Eq. (64). However, under $H_1$, all $G$ spectral estimates are now Rician distributed with pdf's of the form in Eq. (62), namely,

$$p_{L(f_i)}(L) = \frac{2}{N_0 T} L \exp\left(-\frac{L^2 + \beta_i^2}{N_0 T}\right) I_0\left(\frac{2L\beta_i}{N_0 T}\right); \quad L \geq 0 \tag{70}$$

where the $\beta_i$'s are determined as follows. Since [see Eq. (15)]

$$\beta_i^2 \triangleq \left(E\left\{L_c(f_i)|\theta, f\right\}\right)^2 + \left(E\left\{L_s(f_i)|\theta, f\right\}\right)^2 \tag{71}$$

then, assuming that the actual received carrier frequency, $f$, is situated midway between $f_k$ and $f_{k+1}$, which are separated by $1/T$, i.e., $f = f_k + 1/2T$, Eq. (70) is evaluated as (for simplicity, we ignore the edge effects at the ends of the frequency uncertainty band)

$$\beta_i^2 = PT^2 \left[\frac{\sin\left(\pi\left(k - i + \frac{1}{2}\right)\right)}{\pi\left(k - i + \frac{1}{2}\right)}\right]^2 = \begin{cases} PT^2\left(\frac{2}{\pi}\right)^2; & i = k, k+1 \\ PT^2\left(\frac{2}{\pi}\right)^2 \left(\frac{1}{1 + 2(k-i)}\right)^2; & i \neq k, k+1 \end{cases}$$

$$\triangleq PT^2 \Gamma_i \tag{72}$$

Finally then, analogous to Eq. (70), the detection probability would be given by

$$P_D = 1 - \prod_{i=0}^{G-1} \left(1 - Q\left(\Gamma_i d, \sqrt{\frac{2\gamma}{N_0 T}}\right)\right) \tag{73}$$

which, in general, depends on $f_k$, i.e., the location of $f$ within the uncertainty band.

It has been suggested in [5] that the two nearest spectral estimates (envelopes) to the frequency location of the received signal dominate the performance, i.e., the spillover effect of signal in the other

frequency slots can be ignored to a first-order approximation. When this is done, then, under $H_1$, two of the spectral estimates are identically Rician distributed and the remaining $G-2$ are identically Rayleigh distributed. In this case, Eq. (73) is replaced by an expression somewhat like Eq. (67), namely,

$$P_D = 1 - \left(1 - \exp\left(-\frac{\gamma}{N_0 T}\right)\right)^{G-2} \left(1 - Q\left(\left(\frac{2}{\pi}\right) d, \sqrt{\frac{2\gamma}{N_0 T}}\right)\right)^2 \tag{74}$$

which is now independent of the frequency location of the signal. Combining Eqs. (64) and (74), the ROC is approximately given by

$$P_D = 1 - (1 - P_F)^{(G-2)/G} \left(1 - Q\left(\left(\frac{2}{\pi}\right) d, \sqrt{-2\ln\left(1 - (1 - P_F)^{1/G}\right)}\right)\right)^2 \tag{75}$$

## VII. A More Precise Formulation

As discussed in Section IV of Part 1, the true transmitted signal corresponds to a sinusoidal carrier phase modulated by a square-wave subcarrier of radian frequency $\omega_{sc}$. At the receiver, the harmonics with frequencies other than the sum and difference of $\omega_{sc}$ and $\omega_c$ are filtered out, which means that in so far as detection is concerned, the received signal in the absence of frequency uncertainty can be modeled as

$$r(t) = s(t, \theta_c, \theta_{sc}) + n(t) = \sqrt{P}\left\{\cos\left[(\omega_c + \omega_{sc})t + (\theta_c + \theta_{sc})\right] + \cos\left[(\omega_c - \omega_{sc})t + (\theta_c - \theta_{sc})\right]\right\} + n(t) \tag{76}$$

In the presence of frequency uncertainty due, for example, to Doppler shift, both the upper and lower frequency tones in Eq. (76) will be shifted from their nominal values with the higher-frequency tone experiencing a larger shift than that corresponding to the lower-frequency tone. If, however, the subcarrier frequency is much smaller than the carrier frequency, i.e., $\omega_{sc} << \omega_c$, as is the case of interest, then for all practical purposes, one can associate the frequency uncertainty with the carrier as discussed in Section V.A and assume to a first-order approximation that both upper and lower frequency tones experience the same frequency shift. Stated another way, we can assume that, in so far as detection is concerned, we observe a pair of tones whose frequencies are unknown (but by the same amount), each in a band $B$ Hz centered around its nominal value. Furthermore, the uncertainty band is assumed to be very narrow with respect to the subcarrier frequency, i.e., $B << f_{sc}$.

### A. The ALR Test

Analogous to what was done in Part 1, the conditional pdf of the received signal under hypothesis $H_1$ is given by

$$p(r(t)|H_1) = \left(\frac{1}{2\pi}\right)^2 \frac{1}{B} \int_{f_c - B/2}^{f_c + B/2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} p(r(t)|H_1, \theta_+, \theta_-, f - f_{sc}, f + f_{sc}) d\theta_+ d\theta_- df \tag{77}$$

whereupon the ALR becomes

$$\Lambda(r(t)) = \exp\left\{-\frac{PT}{N_0}\right\} \frac{1}{B} \int_{f_c - B/2}^{f_c + B/2} I_0\left(\frac{\sqrt{2P}}{N_0} L_-(f)\right) I_0\left(\frac{\sqrt{2P}}{N_0} L_+(f)\right) df \tag{78}$$

In Eqs. (77) and (78), the spectral envelopes at the lower and upper tones are defined by

$$L_{\pm}(f) \triangleq \sqrt{L_{c\pm}^2(f) + L_{s\pm}^2(f)} \tag{79}$$

together with

$$\left.\begin{aligned} L_{c\pm}(f) &\triangleq \int_0^T r(t)\sqrt{2}\cos\left[2\pi(f \pm f_{sc})t\right] dt \\ L_{s\pm}(f) &\triangleq \int_0^T r(t)\sqrt{2}\sin\left[2\pi(f \pm f_{sc})t\right] dt \end{aligned}\right\} \tag{80}$$

Discretizing the integration interval results in the approximate decision rule

$$\sum_{i=0}^{G-1} I_0\left(\frac{2\sqrt{P}}{N_0}L_+(f_i)\right) I_0\left(\frac{2\sqrt{P}}{N_0}L_-(f_i)\right) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{81}$$

where the spectral envelopes required in Eq. (81) are defined analogously to Eqs. (79) and (80), with the continuous random variable $f$ replaced by the discrete random variable $f_i; i = 0, 1, \cdots, G-1$. As was the case for the single-tone result in Section V.A, the performance (ROC) of the decision rule in Eq. (81) cannot be obtained analytically.

### B. The MLR Test

Without going into great detail, it is straightforward to show (using the results of Section IV.B) that the MLR test analogous to Eq. (58) becomes

$$\max_f \exp\left(-\frac{PT}{N_0}\right) \exp\left(\frac{2\sqrt{P}}{N_0}L_+(f)\right) \exp\left(\frac{2\sqrt{P}}{N_0}L_-(f)\right) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{82}$$

or, equivalently,

$$\max_f (L_-(f) + L_+(f)) \underset{H_0}{\overset{H_1}{\gtrless}} \sqrt{\gamma} \tag{83}$$

which has the discretized version

$$\max_i (L_-(f_i) + L_+(f)) \underset{H_0}{\overset{H_1}{\gtrless}} \sqrt{\gamma} \tag{84}$$

Unfortunately, the performance of the receiver that implements the decision rule of Eq. (84) also cannot be obtained analytically.

## VIII. Numerical Results

Since the performance of none of the ALR optimum decision rules can be evaluated analytically and since the same is true for some of the MLR decision rules, a computer simulation of these metrics has been developed to numerically evaluate such performance. The results of such simulations are described as follows. Figure 8 is a sample illustration of the ROC for the case of a single tone with unknown phase and frequency (as described in Section V) and a detection SNR $d^2 = 2PT/N_0 = 6$ dB. Both ALR and MLR cases are illustrated, corresponding, respectively, to the decision rules of Eqs. (54) and (60). Also, both the best- and worst-case input frequency scenarios are considered, corresponding, respectively, to the cases where the actual input frequency is indeed equal to one of the $G$ frequencies used to approximately implement the decision rule and the case where the actual input frequency falls midway between any two of these $G$ frequencies. Clearly, the actual system performance corresponding to an input frequency arbitrarily chosen in the uncertainty band will lie between these two performance bounds. We observe from the results in Fig. 8 that the difference between best- and worst-case performance is relatively small, as well as is the difference between the ALR (optimum) and MLR (suboptimum) decision rules. There is a significant difference, however, between the performance for $G = 10$ and $G = 100$, indicating the sensitivity of the performance degradation to a factor of 10 increase in frequency uncertainty. Also, comparing Fig. 8 with the analogous curve in Fig. 2, corresponding to the case of unknown phase but *known* frequency, we again see a rather significant degradation in performance when the frequency is unknown even by only a factor of 10 relative to the observation bandwidth (reciprocal of the observation time, $T$), i.e., $G = 10$.

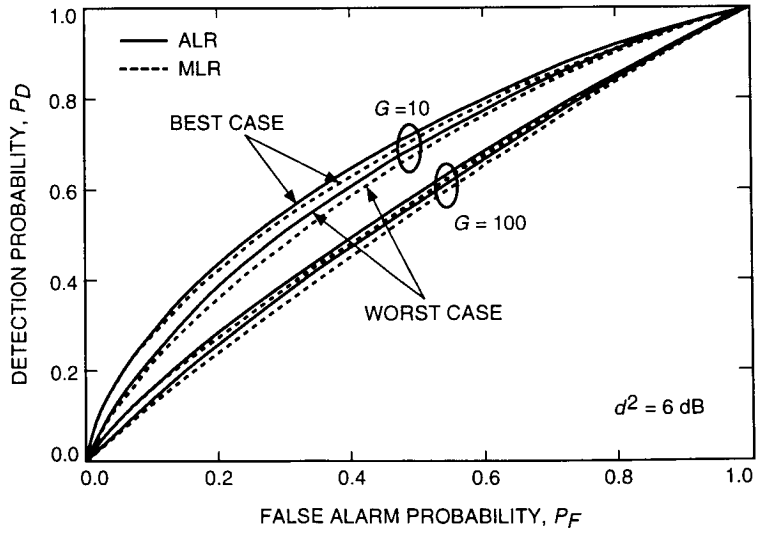


**Fig. 8. ROC: frequency and phase unknown (single-tone) simulation results.**

As verification of the MLR simulation results, we present in Fig. 9 the analogous analytical results obtained from Eqs. (69) and (75). Recall that in arriving at Eq. (75) the assumption was made that the energy spillover effect of the signal into the other frequency slots is dominated by the two adjacent ones. Thus, ignoring edge effects, it was not necessary to average over all possible worst-case (midway) input frequency positions. In the computer simulation, this assumption was not invoked, as the input frequency was allowed to occur midway between *any* two adjacent frequencies. Despite this analysis

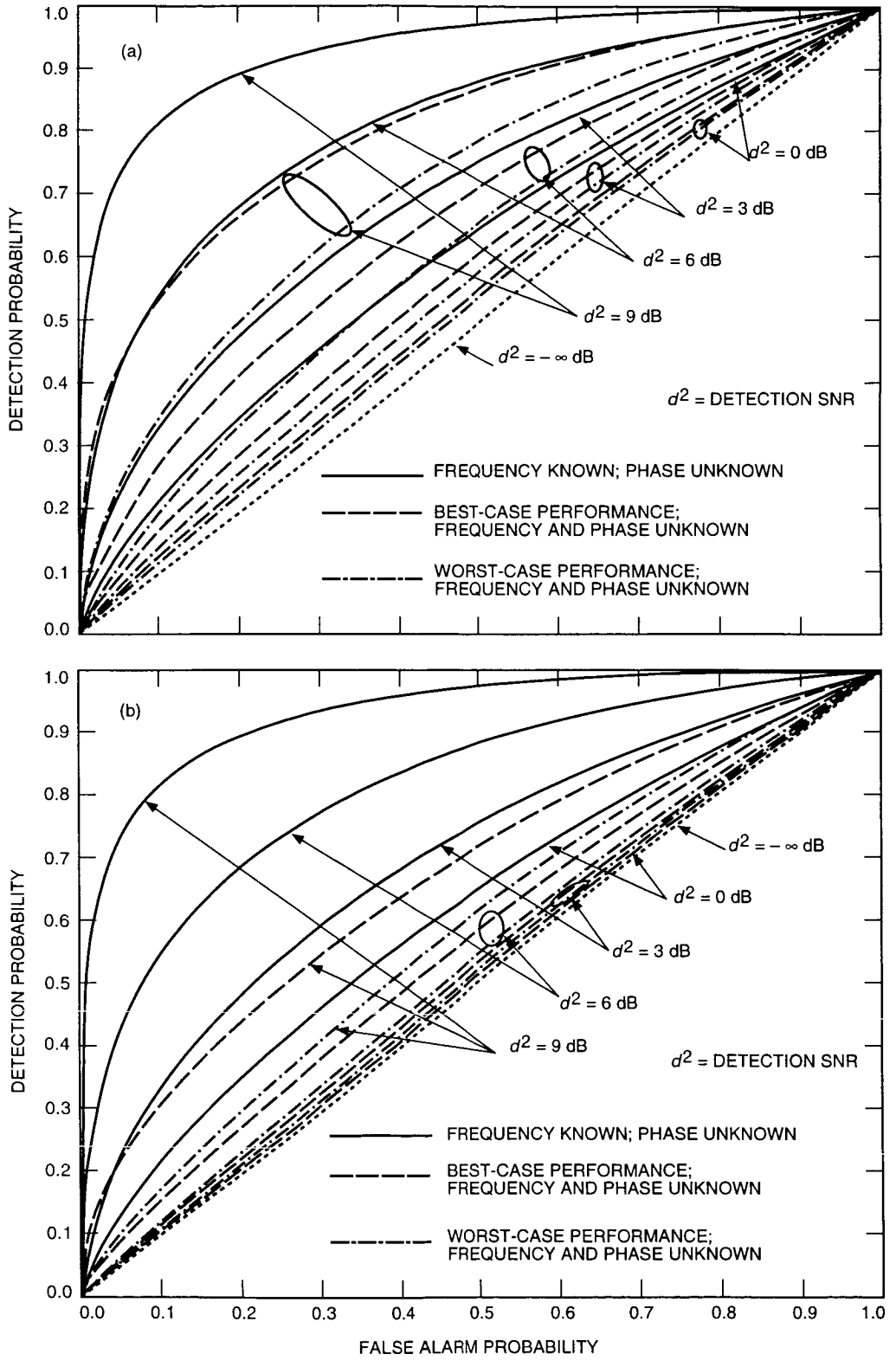


**Fig. 9. ROC single tone, analytical MLR results: (a) $G = 10$ and (b) $G = 100$.**

approximation, however, comparison of the results in Figs. 8 and 9 reveals excellent agreement between analysis and simulation, i.e., the assumption of only adjacent signal energy spillover used to arrive at Eq. (75) has been justified. Also indicated in Fig. 8 is the analytical result corresponding to known phase and frequency (recall that this result is the same for both MLR and ALR) that allows a more direct assessment of the performance degradation due to lack of perfect frequency knowledge.

Since the curves in Fig. 8 are drawn for a fixed value of detection SNR $d^2 = 2PT/N_0$, then assuming that $P/N_0$ is specified, this implies that the observation interval, $T$, is also held constant. Thus, changing the value of $G = BT$ from 10 to 100 directly translates into a change by a factor of 10 in the frequency uncertainty region $B$, which accounts for the observed degradation in performance. Another interpretation of the numerical data can be obtained by again holding $P/N_0$ fixed but observing the effect on system performance of increasing $T$ for a fixed frequency uncertainty region $B$. This necessitates plotting the ROC with both $d^2$ and $G$ increasing linearly with $T$. Such a plot for the ALR decision rule with best-case input frequency is illustrated in Fig. 10, where the ROC is plotted for values of $G = 10, 20, 40$, and 80 ($T$ increasing by a factor of 2) and corresponding values $d^2 = 6, 9, 12, 15$ dB. To directly see the dependence of MLR system performance on detection SNR, Fig. 11 illustrates the behavior of detection probability, $P_D$, versus detection SNR, $d^2$, for a fixed false alarm probability, $P_F = 10^{-2}$, and values of $G = 10$ and 100. These curves are obtained from numerical evaluation of the analytical results in Section VI. Since along any curve $G$ is held fixed, one can interpret these results as keeping the frequency uncertainty band, $B$, and observation time, $T$, fixed and observing the change in performance as $P/N_0$ is varied.

The penalty associated with detecting a pair of subcarrier tones (each at half the total transmitted power) as opposed to a single carrier tone (at full transmitted power) is illustrated by the numerical results in Fig. 12. Here we plot the ROC for both the single- and double-tone cases for the ALR decision rule with best-case input frequency and a detection SNR equal to 6 dB. The results for the single-tone case are taken directly from Fig. 8. We observe a significant performance penalty associated with using a double-tone detection scheme. Figure 13 illustrates for the double-tone detection scheme results analogous to Fig. 10 for the single-tone detection scheme. Here again, by comparing the two figures, we observe a significant penalty associated with using a pair of equal half-power subcarrier tones rather than a single tone at full power.

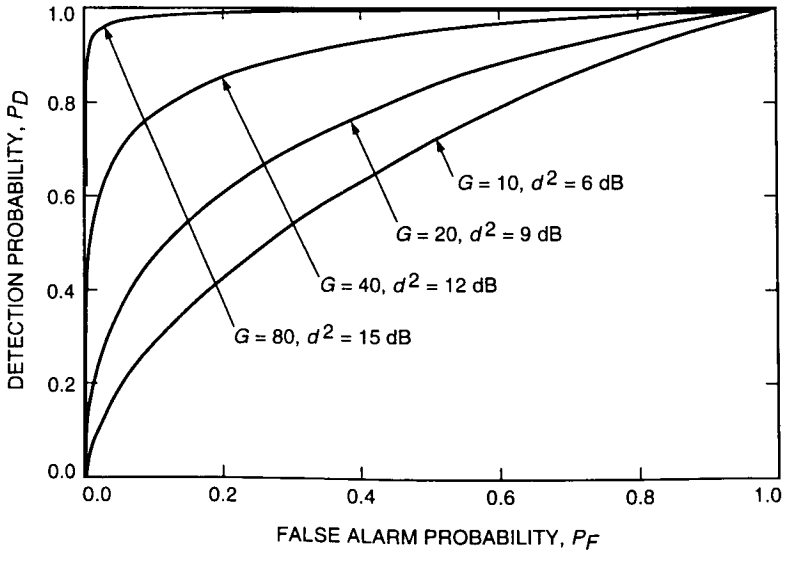


**Fig. 10. ROC simulation results: frequency and phase unknown (single tone), ALR , best-case input frequency.**

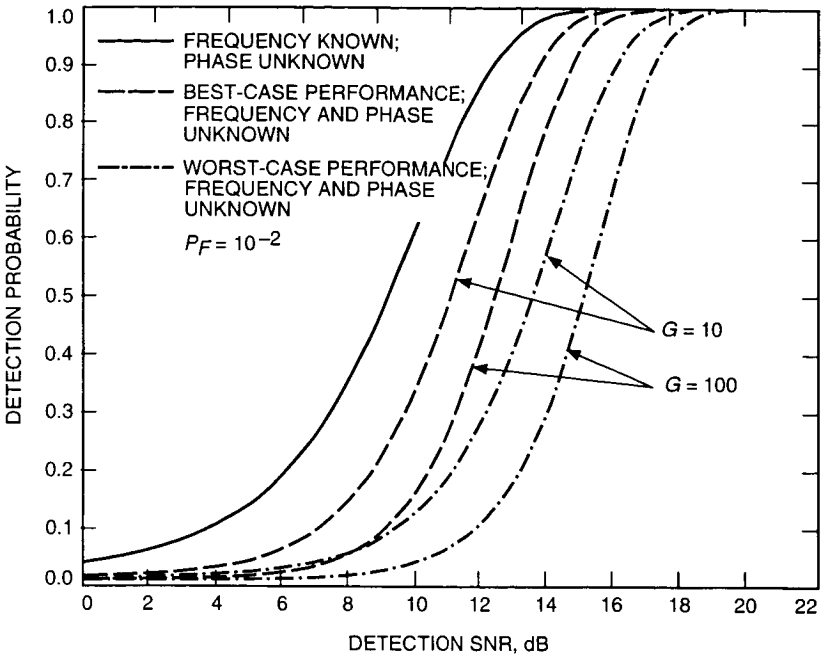


**Fig. 11. Detection probability versus detection SNR analytical MLR results (single tone ).**

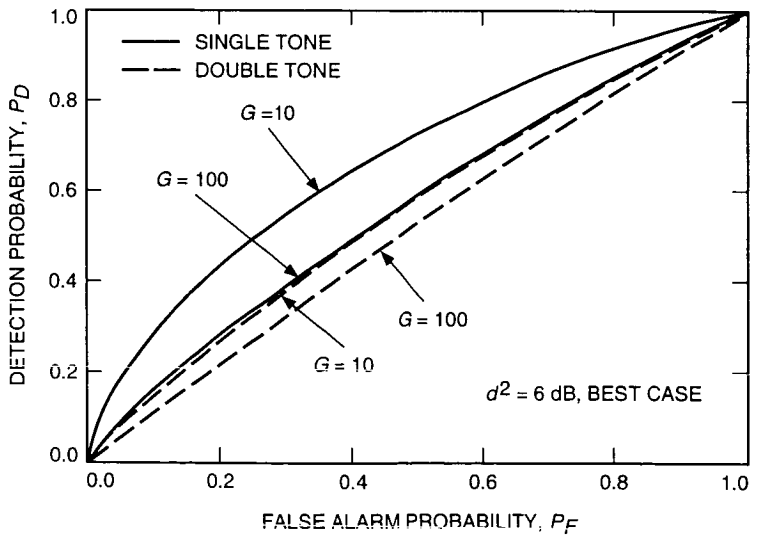


**Fig. 12. ROC: frequency and phase unknown (single/double tone), ALR, best-case input frequency.**

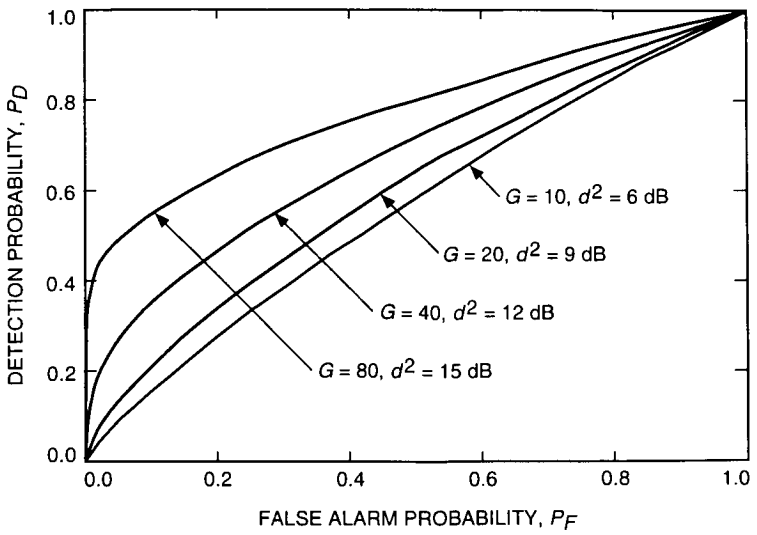


**Fig. 13. ROC simulation results: frequency and phase unknown (double tone), ALR, best-case input frequency.**

## Acknowledgment

The authors would like to thank Mr. Van Snyder of the Supercomputing and Computational Mathematics Support Group for his advice on performing the numerical integrations required in many of the analytical results.

## References

[1] H. L. VanTrees, *Detection, Estimation, and Modulation Theory, Part 1*, New York: John Wiley & Sons, Inc., 1968.

[2] M. K. Simon, S. M. Hinedi, and W. C. Lindsey, *Digital Communication Techniques: Signal Design and Detection*, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1994.

[3] A. Polydoros and C. L. Weber, "Optimal Detection Considerations for Low Probability of Intercept," *MILCOM '82 Conference Proceedings*, Boston, Massachusetts, pp. 2.1-1–2.1-5, October 1992.

[4] A. Polydoros and K. T. Woo, "Wideband Spectral Detection of Unknown Frequency Signals," presented at the International Symposium on Information Theory, Brighton, England, June 1985.

[5] C.-D. Chung and A. Polydoros, *Multi-Hop FH/LPI Detection, Part II: Spectral Techniques*, Technical Report CSI-87-09-01, University of Southern California, Los Angeles, California, September 1987.

[6] A. Polydoros and C. L. Nikias, "Detection of Unknown-Frequency Sinusoids in Noise: Spectral Versus Correlation Detection Rules," *IEEE ASSP*, vol. 35, no. 6, pp. 897–900, June 1987.

[7] A. A. G. Requicha, "Direct Computation of Distribution Function From Characteristic Functions Using the Fast Fourier Transform," *Proceedings of the IEEE*, vol. 58, no. 7, pp. 1154–1155, July 1970.

# Appendix

## On the Independence of the Sum of Difference of Two Uniformly Distributed Random Variables Modulo 2π

Consider two independent random phases $\theta_A$ and $\theta_B$ that are each uniformly distributed in the semiclosed interval $[-\pi, \pi)$. Define the sum and difference of these two random variables by

$$\left.\begin{aligned} \theta'_+ &\triangleq \theta_A + \theta_B \\ \theta'_- &\triangleq \theta_A - \theta_B \end{aligned}\right\} \tag{A-1}$$

and the modulo $2\pi$ versions of these random variables by

$$\left.\begin{aligned} \theta_+ &\triangleq (\theta'_+)_{\text{mod } 2\pi} = (\theta_A + \theta_B)_{\text{mod } 2\pi} \\ \theta_- &\triangleq (\theta'_-)_{\text{mod } 2\pi} = (\theta_A - \theta_B)_{\text{mod } 2\pi} \end{aligned}\right\} \tag{A-2}$$

The probability density functions (pdf's) of $\theta'_+$ and $\theta'_-$ are triangular in the semiclosed interval $[-2\pi, 2\pi)$, i.e., they are the convolutions of two uniform pdf's, whereas the pdf's of their modulo $2\pi$ reduced versions, $\theta_+$ and $\theta_-$, are once again uniformly distributed in $[-\pi, \pi)$ (see Fig. A-1). We would now like to show that $\theta_+$ and $\theta_-$ are indeed *independent* random variables. To do this, we shall show that the conditional pdf $p_{\theta_-}(\theta_-|\theta_+)$ satisfies $p_{\theta_-}(\theta_-|\theta_+) = p_{\theta_-}(\theta_-)$, i.e., it is a uniform distribution in $[-\pi, \pi)$. Similarly, it can be shown that $p_{\theta_+}(\theta_+|\theta_-) = p_{\theta_+}(\theta_+)$.

Let $\theta_+$ be any positive value in its region of definition, i.e., $0 \leq \theta_+ \leq \pi$. Then, $\theta_A$ and $\theta_B$ are related as follows:

$$\theta_B = \begin{cases} -\theta_A + \theta_+ - 2\pi, & -\pi < \theta_A \leq -\pi + \theta_+ \\ -\theta_A + \theta_+, & -\pi + \theta_+ \leq \theta_A \leq \pi \end{cases} \tag{A-3}$$

From Eq. (A-1), we find that

$$\theta'_- = \begin{cases} 2\theta_A - \theta_+ + 2\pi, & -\pi < \theta_A \leq -\pi + \theta_+ \\ 2\theta_A - \theta_+, & -\pi + \theta_+ \leq \theta_A \leq \pi \end{cases} \tag{A-4}$$

Thus, from Eq. (A-4) and the fact that $\theta_A$ is uniform in the interval $[-\pi, \pi)$, the conditional pdf $p_{\theta'_-}(\theta'_-|\theta_+)$ appears as in Fig. A-2(a). Reducing $\theta'_+$ modulo $2\pi$ produces the conditional pdf $p_{\theta_-}(\theta_-|\theta_+)$ as illustrated in Fig. A-2(b), i.e., a uniform distribution in the interval $[-\pi, \pi)$ Q.E.D.

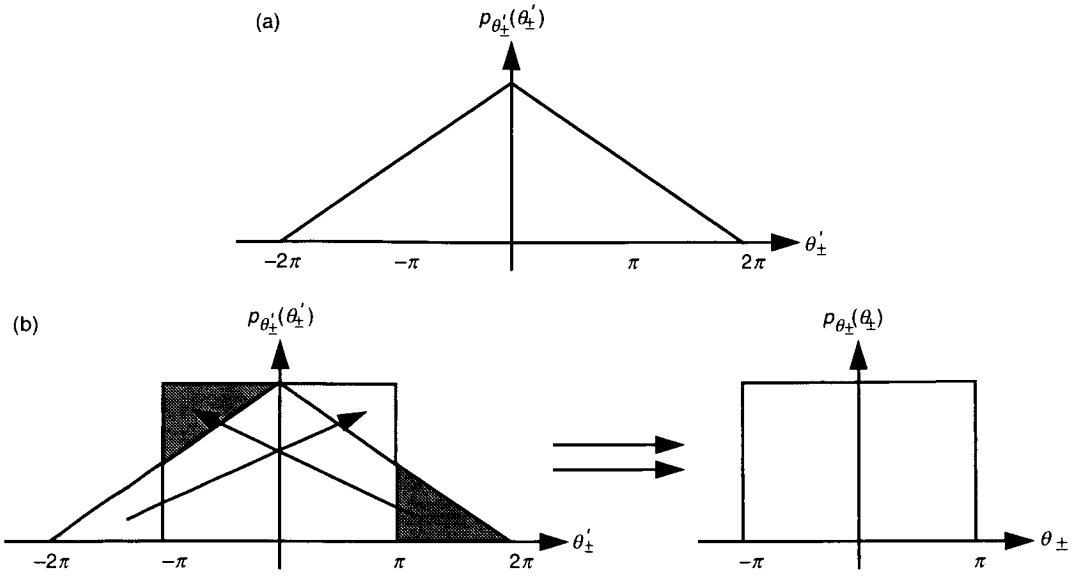


**Fig. A-1. The PDF of the sum and difference of (a) two uniformly distributed random variables and (b) two uniformly distributed random variables reduced modulo $2\pi$.**

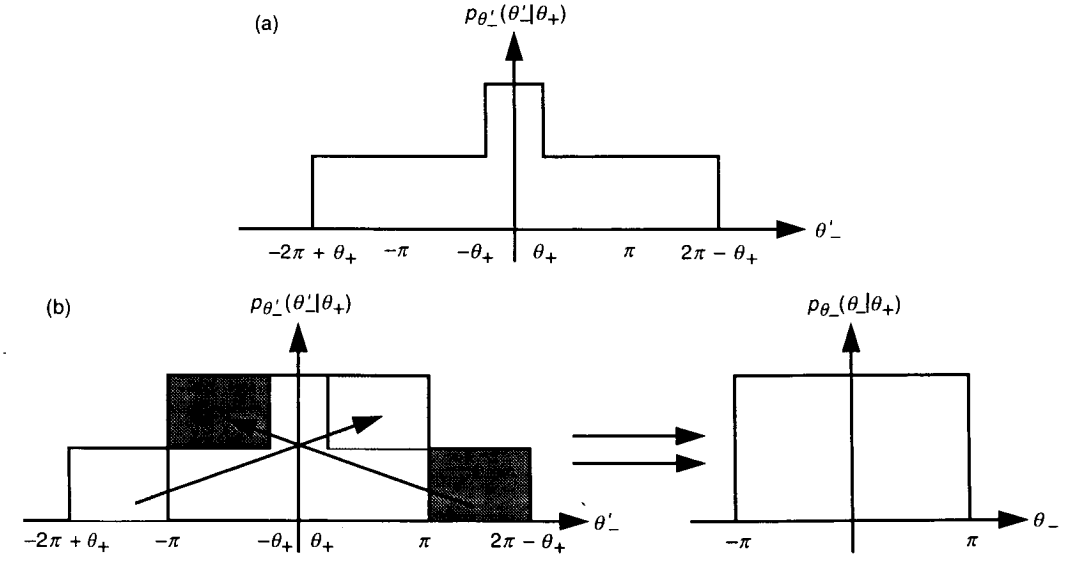


**Fig. A-2. Conditional PDF of the sum and difference of (a) two uniformly distributed random variables and (b) two uniformly distributed random variables reduced modulo $2\pi$.**

TDA Progress Report 42-123 November 15, 1995

# On the Design of Turbo Codes

D. Divsalar and F. Pollara
Communications Systems and Research Section

*In this article, we design new turbo codes that can achieve near-Shannon-limit performance. The design criterion for random interleavers is based on maximizing the effective free distance of the turbo code, i.e., the minimum output weight of codewords due to weight-2 input sequences. An upper bound on the effective free distance of a turbo code is derived. This upper bound can be achieved if the feedback connection of convolutional codes uses primitive polynomials. We review multiple turbo codes (parallel concatenation of q convolutional codes), which increase the so-called "interleaving gain" as q and the interleaver size increase, and a suitable decoder structure derived from an approximation to the maximum a posteriori probability decision rule. We develop new rate 1/3, 2/3, 3/4, and 4/5 constituent codes to be used in the turbo encoder structure. These codes, for from 2 to 32 states, are designed by using primitive polynomials. The resulting turbo codes have rates* b/n, b=*1, 2, 3, 4, and* n=*2, 3, 4, 5, 6 and include random interleavers for better asymptotic performance. These codes are suitable for deep-space communications with low throughput and for near-Earth communications where high throughput is desirable. The performance of these codes is within 1 dB of the Shannon limit at a bit-error rate of $10^{-6}$ for throughputs from 1/15 up to 4 bits/s/Hz.*

## I. Introduction

Coding theorists have traditionally attacked the problem of designing good codes by developing codes with a lot of structure, which lends itself to feasible decoders, although coding theory suggests that codes chosen "at random" should perform well if their block sizes are large enough. The challenge to find practical decoders for "almost" random, large codes has not been seriously considered until recently. Perhaps the most exciting and potentially important development in coding theory in recent years has been the dramatic announcement of "turbo codes" by Berrou et al. in 1993 [7]. The announced performance of these codes was so good that the initial reaction of the coding establishment was deep skepticism, but recently researchers around the world have been able to reproduce those results [15,19,8]. The introduction of turbo codes has opened a whole new way of looking at the problem of constructing good codes [5] and decoding them with low complexity [7,2].

Turbo codes achieve near-Shannon-limit error correction performance with relatively simple component codes and large interleavers. A required $E_b/N_0$ of 0.7 dB was reported for a bit-error rate (BER) of $10^{-5}$ for a rate 1/2 turbo code [7]. Multiple turbo codes (parallel concatenation of $q > 2$ convolutional codes) and a suitable decoder structure derived from an approximation to the maximum a posteriori (MAP) probability decision rule were reported in [9]. In [9], we explained for the first time the turbo decoding

scheme for multiple codes and its relation to the optimum bit decision rule, and we found rate 1/4 turbo codes whose performance is within 0.8 dB of Shannon's limit at BER=$10^{-5}$.

In this article, we (1) design the best component codes for turbo codes of various rates by maximizing the "effective free distance of the turbo code," i.e., the minimum output weight of codewords due to weight-2 input sequences; (2) describe a suitable trellis termination rule for $b/n$ codes; (3) design low throughput turbo codes for power-limited channels (deep-space communications); and (4) design high-throughput turbo trellis-coded modulation for bandwidth-limited channels (near-Earth communications).

## II. Parallel Concatenation of Convolutional Codes

The codes considered in this article consist of the parallel concatenation of multiple ($q \geq 2$) convolutional codes with random interleavers (permutations) at the input of each encoder. This extends the original results on turbo codes reported in [7], which considered turbo codes formed from just two constituent codes and an overall rate of 1/2.

Figure 1 provides an example of parallel concatenation of three convolutional codes. The encoder contains three recursive binary convolutional encoders with $m_1$, $m_2$, and $m_3$ memory cells, respectively. In general, the three component encoders may be different and may even have different rates. The first component encoder operates directly (or through $\pi_1$) on the information bit sequence $\mathbf{u} = (u_1, \cdots, u_N)$ of length $N$, producing the two output sequences $\mathbf{x}_0$ and $\mathbf{x}_1$. The second component encoder operates on a reordered sequence of information bits, $\mathbf{u}_2$, produced by a permuter (interleaver), $\pi_2$, of length $N$, and outputs the sequence $\mathbf{x}_2$. Similarly, subsequent component encoders operate on a reordered sequence of information bits. The interleaver is a pseudorandom block scrambler defined by a permutation of $N$ elements without repetitions: A complete block is read into the the interleaver and read out in a specified (fixed) random order. The same interleaver is used repeatedly for all subsequent blocks.

Figure 1 shows an example where a rate $r = 1/n = 1/4$ code is generated by three component codes with memory $m_1 = m_2 = m_3 = m = 2$, producing the outputs $\mathbf{x}_0 = \mathbf{u}$, $\mathbf{x}_1 = \mathbf{u} \cdot g_1/g_0$, $\mathbf{x}_2 = \mathbf{u}_2 \cdot g_1/g_0$, and $\mathbf{x}_3 = \mathbf{u}_3 \cdot g_1/g_0$ (here $\pi_1$ is assumed to be an identity, i.e., no permutation), where the generator polynomials $g_0$ and $g_1$ have octal representation $(7)_{octal}$ and $(5)_{octal}$, respectively. Note that various code rates can be obtained by proper puncturing of $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, and even $\mathbf{x}_0$ (for an example, see Section V).

We use the encoder in Fig. 1 to generate an $(n(N+m), N)$ block code, where the $m$ tail bits of code 2 and code 3 are not transmitted. Since the component encoders are recursive, it is not sufficient to set the last $m$ information bits to zero in order to drive the encoder to the all-zero state, i.e., to *terminate* the trellis. The termination (tail) sequence depends on the state of each component encoder after $N$ bits, which makes it impossible to terminate all component encoders with $m$ predetermined tail bits. This issue, which had not been resolved in the original turbo code implementation, can be dealt with by applying a simple method described in [8] that is valid for any number of component codes. A more complicated method is described in [18].

A design for constituent convolutional codes, which are not necessarily optimum convolutional codes, was originally reported in [5] for rate $1/n$ codes. In this article, we extend those results to rate $b/n$ codes. It was suggested (without proof) in [2] that good random codes are obtained if $g_a$ is a primitive polynomial. This suggestion, used in [5] to obtain "good" rate 1/2 constituent codes, will be used in this article to obtain "good" rate 1/3, 2/3, 3/4, and 4/5 constituent codes. By "good" codes we mean codes with a maximum effective free distance $d_{ef}$, those codes that maximize the minimum output weight for weight-2 input sequences, as discussed in [9], [13], and [5] (because this weight tends to dominate the performance characteristics over the region of interest).

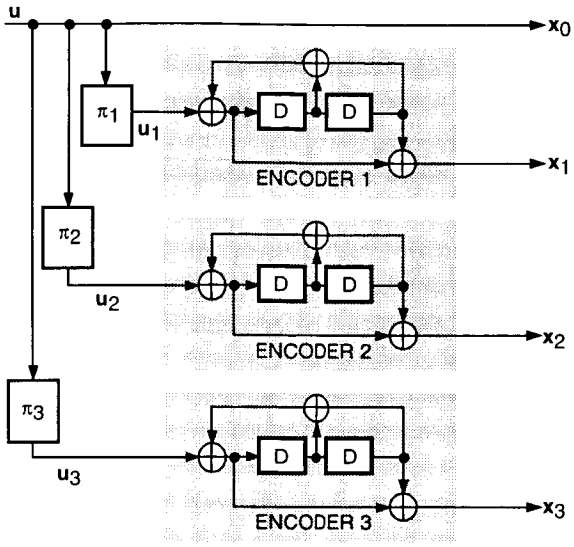


**Fig. 1. Example of encoder with three codes.**

## III. Design of Constituent Encoders

As discussed in the previous section, maximizing the weight of output codewords corresponding to weight-2 data sequences gives the best BER performance for a moderate bit signal-to-noise ratio (SNR) as the random interleaver size $N$ gets large. In this region, the dominant term in the expression for bit error probability of a turbo code with $q$ constituent encoders is

$$P_b \approx \frac{\beta}{N^{q-1}} Q\left(\sqrt{2r\frac{E_b}{N_0}\left(\sum_{j=1}^{q} d_{j,2}^p + 2\right)}\right)$$

where $d_{j,2}^p$ is the minimum parity-weight (weight due to parity checks only) of the codewords at the output of the $j$th constituent code due to weight-2 data sequences, and $\beta$ is a constant independent of $N$. Define $d_{j,2} = d_{j,2}^p + 2$ as the minimum output weight including parity and information bits, if the $j$th constituent code transmits the information (systematic) bits. Usually one constituent code transmits the information bits ($j = 1$), and the information bits of others are punctured. Define $d_{ef} = \sum_{j=1}^{q} d_{j,2}^p + 2$ as the effective free distance of the turbo code and $1/N^{q-1}$ as the "interleaver's gain." We have the following bound on $d_2^p$ for any constituent code.

**Theorem 1**. For any $r = b/(b+1)$ recursive systematic convolutional encoder with generator matrix

$$G = \begin{bmatrix} & \frac{h_1(D)}{h_0(D)} \\ & \frac{h_2(D)}{h_0(D)} \\ \mathbf{I}_{b\times b} & \cdot \\ & \cdot \\ & \frac{h_b(D)}{h_0(D)} \end{bmatrix}$$

where $I_{b\times b}$ is a $b \times b$ identity matrix, $\deg[h_i(D)] \leq m$, $h_i(D) \neq h_0(D)$, $i = 1, 2, \cdots, b$, and $h_0(D)$ is a primitive polynomial of degree $m$, the following upper bound holds:

$$d_2^p \leq \lfloor \frac{2^{m-1}}{b} \rfloor + 2$$

**Proof.** In the state diagram of any recursive systematic convolutional encoder with generator matrix $G$, there exist at least two nonoverlapping loops corresponding to all-zero input sequences. If $h_0(D)$ is a primitive polynomial, there are two loops: one corresponding to zero-input, zero-output sequences with branch length one, and the other corresponding to zero-input but nonzero-output sequences with branch length $2^m - 1$, which is the period of maximal length (ML) linear feedback shift registers (LFSRs) [14] with degree $m$. The parity codeword weight of this loop is $2^{m-1}$, due to the balance property [14] of ML sequences. This weight depends only on the degree of the primitive polynomial and is independent of $h_i(D)$, due to the invariance to initial conditions of ML LFSR sequences. In general, the output of the encoder is a linear function of its input and current state. So, for any output we may consider, provided it depends on at least one component of the state and it is not $h_0(D)$, the weight of a zero-input loop is $2^{m-1}$, by the shift-and-add property of ML LFSRs.

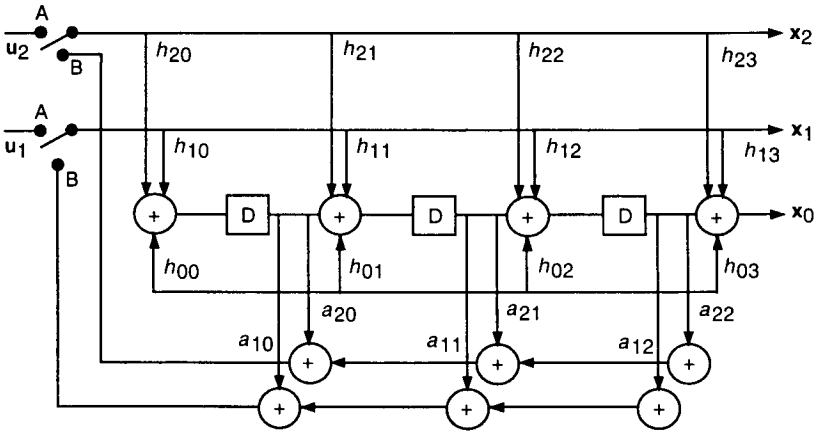


**Fig. 2. Canonical representation of a rate $(b + 1)/b$ encoder ($b = 2$, $m = 3$).**

Consider the canonical representation of a rate $(b + 1)/b$ encoder [20] as shown in Fig. 2 when the switch is in position A. Let $S^k(D)$ be the state of the encoder at time $k$ with coefficients $S_0^k, S_1^k, \cdots, S_{m-1}^k$, where the output of the encoder at time $k$ is

$$X = S_{m-1}^{k-1} + \sum_{i=1}^{b} u_i^k h_{i,m} \tag{1}$$

The state transition for input $u_1^k, \cdots, u_b^k$ at time $k$ is given by

$$S^k(D) = \left[ \sum_{i=1}^{b} u_i^k h_i(D) + DS^{k-1}(D) \right] \mod h_0(D) \tag{2}$$

From the all-zero state, we can enter the zero-input loop with nonzero input symbols $u_1, \cdots, u_b$ at state

$$S^1(D) = \sum_{i=1}^{b} u_i h_i(D) \bmod h_0(D) \tag{3}$$

From the same nonzero input symbol, we leave exactly at state $S^{2^m-1}(D)$ back to the all-zero state, where $S^{2^m-1}(D)$ satisfies

$$S^1(D) = DS^{2^m-1}(D) \bmod h_0(D) \tag{4}$$

i.e., $S^{2^m-1}(D)$ is the "predecessor" to state $S^1(D)$ in the zero-input loop. If the most significant bit of the predecessor state is zero, i.e., $S_{m-1}^{2^m-1} = 0$, then the branch output for the transition from $S^{2^m-1}(D)$ to $S^1(D)$ is zero for a zero-input symbol. Now consider any weight-1 input symbol, i.e., $u_j = 1$ for $j = i$ and $u_j = 0$ for $j \neq i$, $j = 1, 2, \cdots, b$. The question is: What are the conditions on the coefficients $h_i(D)$ such that, if we enter with a weight-1 input symbol into the zero-input loop at state $S^1(D)$, the most significant bit of the "predecessor" state $S^{2^m-1}(D)$ is zero. Using Eqs. (3) and (4), we can establish that

$$h_{i0} + h_{i,m} = 0 \tag{5}$$

Obviously, when we enter the zero-input loop from the all-zero state and when we leave this loop to go back to the all-zero state, we would like the parity output to be equal to 1. From Eqs. (1) and (5), we require

$$\left.\begin{aligned} h_{i0} &= 1 \\ h_{i,m} &= 1 \end{aligned}\right\} \tag{6}$$

With this condition, we can enter the zero-input loop with a weight-1 symbol at state $S^1(D)$ and then leave this loop from state $S^{2^m-1}(D)$ back to the all-zero state, for the same weight-1 input. The parity weight of the codeword corresponding to weight-2 data sequences is then $2^{m-1} + 2$, where the first term is the weight of the zero-input loop and the second term is due to the parity bit appearing when entering and leaving the loop. If $b = 1$, the proof is complete, and the condition to achieve the upper bound is given by Eq. (6). For $b = 2$, we may enter the zero-input loop with $\mathbf{u} = 10$ at state $S^1(D)$ and leave the loop to the zero state with $\mathbf{u} = 01$ at some state $S^j(D)$. If we can choose $S^j(D)$ such that the output weight of the zero-input loop from $S^1(D)$ to $S^j(D)$ is exactly $2^{m-1}/2$, then the output weight of the zero-input loop from $S^{j+1}(D)$ to $S^{2^m-1}(D)$ is exactly $2^{m-1}/2$, and the minimum weight of codewords corresponding to some weight-2 data sequences is

$$\frac{2^{m-1}}{2} + 2$$

In general, for any $b$, if we extend the procedure for $b = 2$, the minimum weight of the codewords corresponding to weight-2 data sequences is

$$\lfloor \frac{2^{m-1}}{b} \rfloor + 2 \tag{7}$$

where $\lfloor x \rfloor$ is the largest integer less than or equal to $x$. Clearly, this is the best achievable weight for the minimum-weight codeword corresponding to weight-2 data sequences. This upper bound can be achieved

if the maximum run length of 1's ($m$) in the zero-input loop does not exceed $\lfloor 2^{m-1}/b \rfloor$. If $m > \lfloor 2^{m-1}/b \rfloor$, then the minimum weight of the codewords corresponding to weight-2 data sequences will be strictly less than $\lfloor 2^{m-1}/b \rfloor + 2$.

The run property of ML LFSRs [14] can help us in designing codes achieving this upper bound. Consider only runs of 1's with length $l$ for $0 < l < m-1$; then there are $2^{m-2-l}$ runs of length $l$, no runs of length $m-1$, and only one run of length $m$. ❐

**Corollary 1.** For any $r = b/n$ recursive systematic convolutional code with $b$ inputs, $b$ systematic outputs, and $n-b$ parity output bits using a primitive feedback generator, we have

$$d_2^p \leq \lfloor \frac{(n-b)2^{m-1}}{b} \rfloor + 2(n-b) \tag{8}$$

**Proof.** The total output weight of a zero-input loop due to parity bits is $(n-b)2^{M-1}$. In this zero-input loop, the largest minimum weight (due to parity bits) for entering and leaving the loop with any weight-1 input symbol is $[(n-b)2^{M-1}]/b$. The output weight due to parity bits for entering and leaving the zero-input loop (both into and from the all-zero state) is $2(n-b)$. ❐

There is an advantage to using $b > 1$, since the bound in Eq. (8) for rate $b/bn$ codes is larger than the bound for rate $1/n$ codes. Examples of codes are found that meet the upper bound for $b/bn$ codes.

### A. Best Rate *b/b* + 1 Constituent Codes

We obtained the best rate 2/3 codes as shown in Table 1, where $d_2 = d_2^p + 2$. The minimum-weight codewords corresponding to weight-3 data sequences are denoted by $d_3$, $d_{min}$ is the minimum distance of the code, and $k = m+1$ in all the tables. By "best" we mean only codes with a large $d_2$ for a given $m$ that result in a maximum effective free distance. We obtained the best rate 3/4 codes as shown in Table 2 and the best rate 4/5 codes as shown in Table 3.

**Table 1. Best rate 2/3 constituent codes.**

| $k$ | Code generator | | | $d_2$ | $d_3$ | $d_{min}$ |
|---|---|---|---|---|---|---|
| 3 | $h_0 = 7$ | $h_1 = 3$ | $h_2 = 5$ | 4 | 3 | 3 |
| 4 | $h_0 = 13$ | $h_1 = 15$ | $h_2 = 17$ | 5 | 4 | 4 |
| 5 | $h_0 = 23$ | $h_1 = 35$ | $h_2 = 27$ | 8 | 5 | 5 |
| | $h_0 = 23$ | $h_1 = 35$ | $h_2 = 33$ | 8 | 5 | 5 |
| 6 | $h_0 = 45$ | $h_1 = 43$ | $h_2 = 61$ | 12 | 6 | 6 |

**Table 2. Best rate 3/4 constituent codes.**

| $k$ | Code generator | | | | $d_2$ | $d_3$ | $d_{min}$ |
|---|---|---|---|---|---|---|---|
| 3 | $h_0 = 7$ | $h_1 = 5$ | $h_2 = 3$ | $h_3 = 1$ | 3 | 3 | 3 |
| | $h_0 = 7$ | $h_1 = 5$ | $h_2 = 3$ | $h_3 = 4$ | 3 | 3 | 3 |
| | $h_0 = 7$ | $h_1 = 5$ | $h_2 = 3$ | $h_3 = 2$ | 3 | 3 | 3 |
| 4 | $h_0 = 13$ | $h_1 = 15$ | $h_2 = 17$ | $h_3 = 11$ | 4 | 4 | 4 |
| 5 | $h_0 = 23$ | $h_1 = 35$ | $h_2 = 33$ | $h_3 = 25$ | 5 | 4 | 4 |
| | $h_0 = 23$ | $h_1 = 35$ | $h_2 = 27$ | $h_3 = 31$ | 5 | 4 | 4 |
| | $h_0 = 23$ | $h_1 = 35$ | $h_2 = 37$ | $h_3 = 21$ | 5 | 4 | 4 |
| | $h_0 = 23$ | $h_1 = 27$ | $h_2 = 37$ | $h_3 = 21$ | 5 | 4 | 4 |

**Table 3. Best rate 4/5 constituent codes.**

| $k$ | Code generator | | | | | $d_2$ | $d_3$ | $d_{min}$ |
|---|---|---|---|---|---|---|---|---|
| 4 | $h_0 = 13$ | $h_1 = 15$ | $h_2 = 17$ | $h_3 = 11$ | $h_4 = 7$ | 4 | 3 | 3 |
| | $h_0 = 13$ | $h_1 = 15$ | $h_2 = 17$ | $h_3 = 11$ | $h_4 = 5$ | 4 | 3 | 3 |
| 5 | $h_0 = 23$ | $h_1 = 35$ | $h_2 = 33$ | $h_3 = 37$ | $h_4 = 31$ | 5 | 4 | 4 |
| | $h_0 = 23$ | $h_1 = 35$ | $h_2 = 27$ | $h_3 = 37$ | $h_4 = 31$ | 5 | 4 | 4 |
| | $h_0 = 23$ | $h_1 = 35$ | $h_2 = 21$ | $h_3 = 37$ | $h_4 = 31$ | 5 | 4 | 4 |

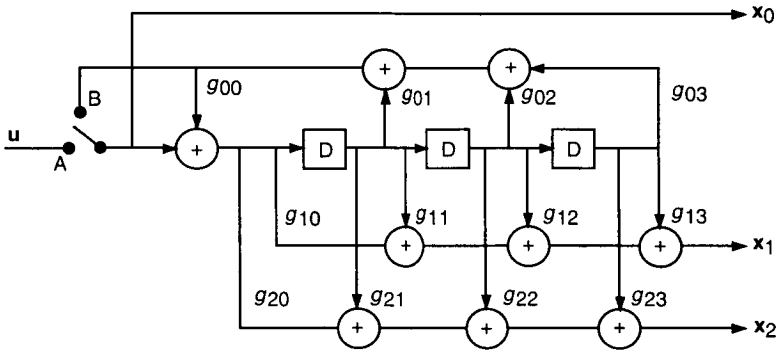


**Fig. 3. Rate 1/*n* code.**

## B. Trellis Termination for *b/n* Codes

Trellis termination is performed (for $b = 2$, as an example) by setting the switch shown in Fig. 2 in position B. The tap coefficients $a_{i0}, \cdots, a_{i,m-1}$ for $i = 1, 2, \cdots, b$ can be obtained by repeated use of Eq. (2) and by solving the resulting equations. The trellis can be terminated in state zero with at least $m/b$ and at most $m$ clock cycles. When Fig. 3 is extended to multiple input bits ($b$ parallel feedback shift registers), a switch should be used for each input bit.

## C. Best Punctured Rate 1/2 Constituent Codes

A rate 2/3 constituent code can be derived by puncturing the parity bit of a rate 1/2 recursive systematic convolutional code using, for example, a pattern $P = [10]$. A puncturing pattern $P$ has zeros where parity bits are removed.

Consider a rate 1/2 recursive systematic convolutional code $(1, g_1(D)/(g_0(D))$. For an input $u(D)$, the parity output can be obtained as

$$x(D) = \frac{u(D)g_1(D)}{g_0(D)} \tag{9}$$

We would like to puncture the output $x(D)$ using, for example, the puncturing pattern $P[10]$ (decimation by 2) and obtain the generator polynomials $h_0(D)$, $h_1(D)$, and $h_2(D)$ for the equivalent rate 2/3 code:

$$G = \begin{bmatrix} 1 & 0 & \dfrac{h_1(D)}{h_0(D)} \\ 0 & 1 & \dfrac{h_2(D)}{h_0(D)} \end{bmatrix}$$

We note that any polynomial $f(D) = \sum a_i D^i$, $a_i \in GF(2)$, can be written as

$$f(D) = f_1(D^2) + Df_2(D^2) \tag{10}$$

where $f_1(D^2)$ corresponds to the even power terms of $f(D)$, and $Df_2(D^2)$ corresponds to the odd power terms of $f(D)$. Now, if we use this approach and apply it to the $u(D)$, $g_1(D)$, and $g_0(D)$, then we can rewrite Eq. (9) as

$$x_1\left(D^2\right) + Dx_2\left(D^2\right) = \frac{\left(u_1\left(D^2\right) + Du_2\left(D^2\right)\right)\left(g_{11}\left(D^2\right) + Dg_{12}\left(D^2\right)\right)}{g_{01}\left(D^2\right) + Dg_{02}\left(D^2\right)} \tag{11}$$

where $x_1(D)$ and $x_2(D)$ correspond to the punctured output $x(D)$ using puncturing patterns $P[10]$ and $P[01]$, respectively. If we multiply both sides of Eq. (11) by $(g_{01}(D^2) + Dg_{02}(D^2))$ and equate the even and the odd power terms, we obtain two equations in two unknowns, namely $x_1(D)$ and $x_2(D)$. For example, solving for $x_1(D)$, we obtain

$$x_1(D) = u_1(D)\frac{h_1(D)}{h_0(D)} + u_2(D)\frac{h_2(D)}{h_0(D)} \tag{12}$$

where $h_0(D) = g_0(D)$ and

$$\left.\begin{aligned} h_1(D) &= g_{11}(D)g_{01}(D) + Dg_{12}(D)g_{02}(D) \\ h_2(D) &= Dg_{12}(D)g_{01}(D) + Dg_{11}(D)g_{02}(D) \end{aligned}\right\} \tag{13}$$

From the second equation in Eq. (13), it is clear that $h_{2,0} = 0$. A similar method can be used to show that for P[01] we get $h_{1,m} = 0$. These imply that the condition of Eq. (6) will be violated. Thus, we have the following theorem.

**Theorem 2**. If the parity puncturing pattern is $P = [10]$ or $P = [01]$, then it is impossible to achieve the upper bound on $d_2 = d_2^p + 2$ for rate 2/3 codes derived by puncturing rate 1/2 codes.

The best rate 1/2 constituent codes with puncturing pattern $P = [10]$ that achieve the largest $d_2$ are given in Table 4.

**Table 4. Best rate 1/2 punctured constituent codes.**

| $k$ | Code generator | | $d_2$ | $d_3$ | $d_{min}$ |
|---|---|---|---|---|---|
| 3 | $g_0 = 7$ | $g_1 = 5$ | 4 | 3 | 3 |
| 4 | $g_0 = 13$ | $g_1 = 15$ | 5 | 4 | 4 |
| 5 | $g_0 = 23$ | $g_1 = 37$ | 7 | 4 | 4 |
| | $g_0 = 23$ | $g_1 = 31$ | 7 | 4 | 4 |
| | $g_0 = 23$ | $g_1 = 33$ | 6 | 5 | 5 |
| | $g_0 = 23$ | $g_1 = 35$ | 6 | 4 | 4 |
| | $g_0 = 23$ | $g_1 = 27$ | 6 | 4 | 4 |

### D. Best Rate 1/*n* Constituent Codes

For rate $1/n$ codes, the upper bound in Eq. (7) for $b = 1$ reduces to

$$d_2^p \leq (n-1)(2^{m-1}+2)$$

This upper bound was originally derived in [5], where the best rate 1/2 constituent codes meeting the bound were obtained. Here we present a simple proof based on our previous general result on rate $b/n$ codes. Then we obtain the best rate 1/3 and 1/4 codes.

**Theorem 3**. For rate $1/n$ recursive systematic convolutional codes with primitive feedback, we have

$$d_2^p \leq (n-1)(2^{m-1}+2)$$

**Proof**. Consider a rate $1/n$ code, shown in Fig. 3. In this figure, $g_0(D)$ is assumed to be a primitive polynomial. As discussed above, the output weight of the zero-input loop for parity bits is $2^{m-1}$ independent of the choice of $g_i(D)$, $i = 1, 2, \cdots, n-1$, provided that $g_i(D) \neq 0$ and that $g_i(D) \neq g_0(D)$, by the shift-and-add and balance properties of ML LFSRs. If $S(D)$ represents the state polynomial, then we can enter the zero-input loop only at state $S^1(D) = 1$ and leave the loop to the all-zero state at state $S^{2^m-1}(D) = D^{m-1}$. The $i$th parity output on the transition $S^{2^m-1}(D) \rightarrow S^1(D)$ with a zero input bit is

$$x_i = g_{i0} + g_{i,m}$$

If $g_{i0} = 1$ and $g_{i,m} = 1$ for $i = 1, \cdots, n-1$, the output weight of the encoder for that transition is zero. The output weight due to the parity bits when entering and leaving the zero-input loop is $(n-1)$ for each case. In addition, the output weight of the zero-input loop will be $(n-1)2^{m-1}$ for $(n-1)$ parity bits. Thus, we established the upper bound on $d_2^p$ for rate 1/n codes. ❐

We obtained the best rate 1/3 and 1/4 codes without parity repetition, as shown in Tables 5 and 6, where $d_2 = d_2^p + 2$ represents the minimum output weight given by weight-2 data sequences. The best rate 1/2 constituent codes are given by $g_0$ and $g_1$ in Table 5, as was also reported in [5].

**Table 5. Best rate 1/3 constituent codes.**

| $k$ | Code generator | | | $d_2$ | $d_3$ | $d_{min}$ |
|---|---|---|---|---|---|---|
| 2 | $g_0 = 3$ | $g_1 = 2$ | $g_2 = 1$ | 4 | $\infty$ | 4 |
| 3 | $g_0 = 7$ | $g_1 = 5$ | $g_2 = 3$ | 8 | 7 | 7 |
| 4 | $g_0 = 13$ | $g_1 = 17$ | $g_2 = 15$ | 14 | 10 | 10 |
| 5 | $g_0 = 23$ | $g_1 = 33$ | $g_2 = 37$ | 22 | 12 | 10 |
| | $g_0 = 23$ | $g_1 = 25$ | $g_2 = 37$ | 22 | 11 | 11 |

**Table 6. Best rate 1/4 constituent codes.**

| $k$ | Code generator | | | | $d_2$ | $d_3$ | $d_{min}$ |
|---|---|---|---|---|---|---|---|
| 4 | $g_0 = 13$ | $g_1 = 17$ | $g_2 = 15$ | $g_3 = 11$ | 20 | 12 | 12 |
| 5 | $g_0 = 23$ | $g_1 = 35$ | $g_2 = 27$ | $g_3 = 37$ | 32 | 16 | 14 |
| | $g_0 = 23$ | $g_1 = 33$ | $g_2 = 27$ | $g_3 = 37$ | 32 | 16 | 14 |
| | $g_0 = 23$ | $g_1 = 35$ | $g_2 = 33$ | $g_3 = 37$ | 32 | 16 | 14 |
| | $g_0 = 23$ | $g_1 = 33$ | $g_2 = 37$ | $g_3 = 25$ | 32 | 15 | 15 |

## E. Recursive Systematic Convolutional Codes With a Nonprimitive Feedback Polynomial

So far, we assumed that the feedback polynomial for recursive systematic convolutional code is a primitive polynomial. We could ask whether it is possible to exceed the upper bound given in Theorem 1 and Corollary 1 by using a nonprimitive polynomial. The answer is negative, thanks to a new theorem by Solomon W. Golomb (Appendix).

**Theorem 4.**[1] For any rate $1/n$ linear recursive systematic convolutional code generated by a nonprimitive feedback polynomial, the upper bound in Theorem 3 cannot be achieved, i.e.,

$$d_2^p < (n-1)(2^{m-1}+2)$$

**Proof.** Using the results of Golomb (see the Appendix) for a nonprimitive feedback polynomial, there are more than two cycles (zero-input loops) in LFSR. The "zero cycle" has weight zero, and the weights of other cycles are nonzero. Thus, the weight of each cycle due to the results of the Appendix is strictly less than $(n-1)2^{m-1}$. If we enter from the all-zero state with input weight-1 to one of the cycles of the shift register, then we have to leave the same cycle to the all-zero state with input weight-1, as discussed in Theorem 1. Thus, $d_2^p < (n-1)(2^{m-1}+2)$. ❐

**Theorem 5.** For any rate $b/b+1$ linear recursive systematic convolutional code generated by a nonprimitive feedback polynomial, the upper bound in Theorem 1 cannot be excceded, i.e.,

$$d_2^p \leq \lfloor \frac{2^{m-1}}{b} \rfloor + 2$$

**Proof.** Again using the results of the Appendix, there is a "zero cycle" with weight zero and at least two cycles with nonzero weights, say $q$ cycles with weights $w_1, w_2, \cdots, w_q$. The sum of the weights of all cycles is exactly $2^{m-1}$, i.e., $\sum w_i = 2^{m-1}$. For a $b/b+1$ code, we have $b$ weight-1 symbols. Suppose that with $b_i$ of these weight-1 symbols we enter from the all-zero state to the $i$th cycle with weight $w_i$; then we have to leave the same cycle to the all-zero state with the same $b_i$ symbols for $i = 1, 2, \cdots, q$, such that $\sum b_i = b$. Based on the discussion in the proof of Theorem 1, the largest achievable minimum output weight of codewords corresponding to weight-2 sequences is $\min(w_1/b_1, w_2/b_2, \cdots, w_q/b_q) + 2$. But it is easy to show that $\min(w_1/b_1, w_2/b_2, \cdots, w_q/b_q) \leq (\sum w_i / \sum b_i) = 2^{m-1}/b$. ❐

[1] The proofs of Theorems 4 and 5 are based on a result by S. W. Golomb (see the Appendix), University of Southern California, Los Angeles, California, 1995. Theorem 4 and Corollary 2 were proved for more general cases when the code is generated by multiple LFSRs by R. J. McEliece, Communications Systems and Research Section, Jet Propulsion Laboratory, Pasadena, California, and California Institute of Technology, Pasadena, California, 1995, using a state-space approach.

**Corollary 2.** For any rate $b/n$ linear recursive systematic convolutional code generated by a non-primitive feedback polynomial, the upper bound in Corollary 1 cannot be exceeded.

**Proof.** The proof is similar to the Proof of Theorem 5, but now $\sum w_i = (n-b)2^{m-1}$. ❐

## IV. Turbo Decoding for Multiple Codes

In [9] we described a new turbo decoding scheme for $q$ codes based on approximating the optimum bit decision rule. The scheme is based on solving a set of nonlinear equations given by ($q = 3$ is used to illustrate the concept)

$$\left.\begin{aligned}
\tilde{L}_{0k} &= 2\rho \mathbf{y}_{0k} \\
\tilde{L}_{1k} &= \log \frac{\sum_{\mathbf{u}:u_k=1} P(\mathbf{y}_1|\mathbf{u}) \prod_{j\neq k} e^{u_j(\tilde{L}_{0j}+\tilde{L}_{2j}+\tilde{L}_{3j})}}{\sum_{\mathbf{u}:u_k=0} P(\mathbf{y}_1|\mathbf{u}) \prod_{j\neq k} e^{u_j(\tilde{L}_{0j}+\tilde{L}_{2j}+\tilde{L}_{3j})}} \\
\tilde{L}_{2k} &= \log \frac{\sum_{\mathbf{u}:u_k=1} P(\mathbf{y}_2|\mathbf{u}) \prod_{j\neq k} e^{u_j(\tilde{L}_{0j}+\tilde{L}_{1j}+\tilde{L}_{3j})}}{\sum_{\mathbf{u}:u_k=0} P(\mathbf{y}_2|\mathbf{u}) \prod_{j\neq k} e^{u_j(\tilde{L}_{0j}+\tilde{L}_{1j}+\tilde{L}_{3j})}} \\
\tilde{L}_{3k} &= \log \frac{\sum_{\mathbf{u}:u_k=1} P(\mathbf{y}_3|\mathbf{u}) \prod_{j\neq k} e^{u_j(\tilde{L}_{0j}+\tilde{L}_{1j}+\tilde{L}_{2j})}}{\sum_{\mathbf{u}:u_k=0} P(\mathbf{y}_3|\mathbf{u}) \prod_{j\neq k} e^{u_j(\tilde{L}_{0j}+\tilde{L}_{1j}+\tilde{L}_{2j})}}
\end{aligned}\right\} \tag{14}$$

for $k = 1, 2, \cdots, N$. In Eq. (14), $\tilde{L}_{ik}$ represents extrinsic information and $\mathbf{y}_i$, $i = 0, 1, 2, 3$ are the received observation vectors corresponding to $\mathbf{x}_i$, $i = 0, 1, 2, 3$ (see Fig. 1), where $\rho = \sqrt{2rE_b/N_0}$, if we assume the channel noise samples have unit variance per dimension. The final decision is then based on

$$L_k = \tilde{L}_{0k} + \tilde{L}_{1k} + \tilde{L}_{2k} + \tilde{L}_{3k} \tag{15}$$

which is passed through a hard limiter with a zero threshold.

The above set of nonlinear equations is derived from the optimum bit decision rule, i.e.,

$$L_k = \log \frac{\sum_{\mathbf{u}:u_k=1} P(\mathbf{y}_0|\mathbf{u})P(\mathbf{y}_1|\mathbf{u})P(\mathbf{y}_2|\mathbf{u})P(\mathbf{y}_3|\mathbf{u})}{\sum_{\mathbf{u}:u_k=0} P(\mathbf{y}_0|\mathbf{u})P(\mathbf{y}_1|\mathbf{u})P(\mathbf{y}_2|\mathbf{u})P(\mathbf{y}_3|\mathbf{u})} \tag{16}$$

using the following approximation:

$$P(\mathbf{u}|\mathbf{y}_i) \approx \prod_{k=1}^{N} \frac{e^{u_k \tilde{L}_{ik}}}{1 + e^{\tilde{L}_{ik}}} \tag{17}$$

Note that, in general, $P(\mathbf{u}|\mathbf{y}_i)$ is not separable. The smaller the Kullback cross entropy [3,17] between right and left distributions in Eq. (17), the better is the approximation and, consequently, the closer is turbo decoding to the optimum bit decision.

We attempted to solve the nonlinear equations in Eq. (14) for $\tilde{\mathbf{L}}_1$, $\tilde{\mathbf{L}}_2$, and $\tilde{\mathbf{L}}_3$ by using the iterative procedure

$$\tilde{L}_{1k}^{(m+1)} = \alpha_1^{(m)} \log \frac{\sum_{\mathbf{u}:u_k=1} P(\mathbf{y}_1|\mathbf{u}) \prod_{j\neq k} e^{u_j(\tilde{L}_{0j}+\tilde{L}_{2j}^{(m)}+\tilde{L}_{3j}^{(m)})}}{\sum_{\mathbf{u}:u_k=0} P(\mathbf{y}_1|\mathbf{u}) \prod_{j\neq k} e^{u_j(\tilde{L}_{0j}+\tilde{L}_{2j}^{(m)}+\tilde{L}_{3j}^{(m)})}} \tag{18}$$

for $k = 1, 2, \cdots, N$, iterating on $m$. Similar recursions hold for $\tilde{L}_{2k}^{(m)}$ and $\tilde{L}_{3k}^{(m)}$. The gain $\alpha_1^{(m)}$ should be equal to one, but we noticed experimentally that better convergence can be obtained by optimizing this gain for each iteration, starting from a value less than 1 and increasing toward 1 with the iterations, as is often done in simulated annealing methods. We start the recursion with the initial condition[2] $\tilde{\mathbf{L}}_1^{(0)} = \tilde{\mathbf{L}}_2^{(0)} = \tilde{\mathbf{L}}_3^{(0)} = \tilde{\mathbf{L}}_0$. For the computation of Eq. (18), we use a modified MAP algorithm[3] with permuters (direct and inverse) where needed, as shown in Fig. 4. The MAP algorithm [1] always starts and ends at the all-zero state since we always terminate the trellis as described in [8]. We assumed $\pi_1 = I$ (identity); however, any $\pi_1$ can be used. The overall decoder is composed of block decoders connected as in Fig. 4, which can be implemented as a pipeline or by feedback. In [10] and [11], we proposed an alternative version of the above decoder that is more appropriate for use in turbo trellis-coded modulation, i.e., set $\tilde{L}_0 = 0$ and consider $\mathbf{y}_0$ as part of $\mathbf{y}_1$. If the systematic bits are distributed among encoders, we use the same distribution for $\mathbf{y}_0$ among the MAP decoders.

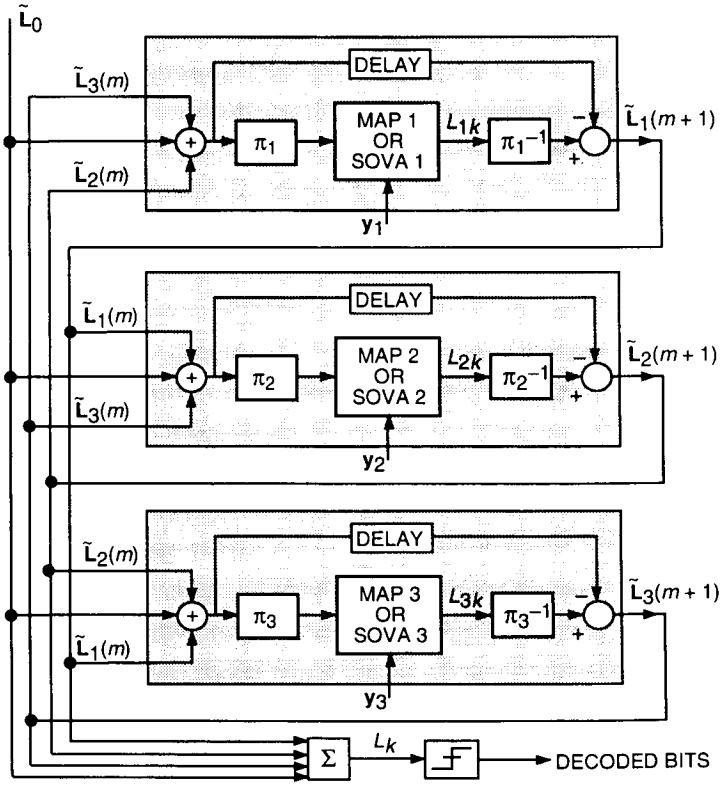


**Fig. 4. Multiple turbo decoder structure.**

[2] Note that the components of the $\tilde{\mathbf{L}}_i$'s corresponding to the tail bits, i.e., $\tilde{L}_{ik}$ for $k = N+1, \cdots, N+M_i$, are set to zero for all iterations.

[3] The modified MAP algorithm is described in S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Soft-Output Decoding Algorithms in Iterative Decoding of Parallel Concatenated Convolutional Codes," submitted to *ICC '96*.

At this point, further approximation for turbo decoding is possible if one term corresponding to a sequence $\mathbf{u}$ dominates other terms in the summation in the numerator and denominator of Eq. (18). Then the summations in Eq. (18) can be replaced by "maximum" operations with the same indices, i.e., replacing $\sum_{\mathbf{u}:u_k=i}$ with $\max_{\mathbf{u}:u_k=i}$ for $i = 0, 1$. A similar approximation can be used for $\tilde{L}_{2k}$ and $\tilde{L}_{3k}$ in Eq. (14). This suboptimum decoder then corresponds to a turbo decoder that uses soft output Viterbi (SOVA)-type decoders rather than MAP decoders. Further approximations, i.e., replacing $\sum$ with max, can also be used in the MAP algorithm.[4]

### A. Decoding Multiple Input Convolutional Codes

If the rate $b/n$ constituent code is not equivalent to a punctured rate $1/n'$ code or if turbo trellis-coded modulation is used, we can first use the symbol MAP algorithm[5] to compute the log-likelihood ratio of a symbol $\mathbf{u} = u_1, u_2, \cdots, u_b$ given the observation $\mathbf{y}$ as

$$\lambda(\mathbf{u}) = \log \frac{P(\mathbf{u}|\mathbf{y})}{P(\mathbf{0}|\mathbf{y})}$$

where $\mathbf{0}$ corresponds to the all-zero symbol. Then we obtain the log-likelihood ratios of the $j$th bit within the symbol by

$$L(u_j) = \log \frac{\sum_{\mathbf{u}:u_j=1} e^{\lambda(\mathbf{u})}}{\sum_{\mathbf{u}:u_j=0} e^{\lambda(\mathbf{u})}}$$

In this way, the turbo decoder operates on bits, and bit, rather than symbol, interleaving is used.

## V. Performance and Simulation Results

The BER performance of these codes was evaluated by using transfer function bounds [4,6,12]. In [12], it was shown that transfer function bounds are very useful for SNRs above the cutoff rate threshold and that they cannot accurately predict performance in the region between cutoff rate and capacity. In this region, the performance was computed by simulation.

Figure 5 shows the performance of turbo codes with $m$ iterations and an interleaver size of $N = 16,384$. The following codes are used as examples:

(1) Rate 1/2 Turbo Codes.

Code A: Two 16-state, rate 2/3 constituent codes are used to construct a rate 1/2 turbo code as shown in Fig. 6. The (worst-case) minimum codeword weights, $d_i$, corresponding to a weight-$i$ input sequence for this code are $d_{ef}$=14, $d_3$=7, $d_4$=8, $d_5$=5=$d_{min}$, and $d_6$=6.

[4] Ibid.

[5] Ibid.

Code B: A rate 1/2 turbo code also was constructed by using a differential encoder and a 32-state, rate 1/2 code, as shown in Fig. 7. This is an example where the systematic bits for both encoders are not transmitted. The (worst-case) minimum codeword weights, $d_i$, corresponding to a weight-$i$ input sequence for this code are $d_{ef}$=19, $d_4$=6=$d_{min}$, $d_6$=9, $d_8$=8, and $d_{10}$=11. The output weights for odd $i$ are large.

(2) Rate 1/3 Turbo Code.

Code C: Two 16-state, rate 1/2 constituent codes are used to construct a rate 1/3 turbo code as shown in Fig. 8. The (worst-case) minimum codeword weights, $d_i$, corresponding to a weight-$i$ input sequence for this code are $d_{ef}$=22, $d_3$ = 11, $d_4$=12, $d_5$ = 9 = $d_{min}$, $d_6$=14, and $d_7$=15.

(3) Rate 1/4 Turbo Code.

Code D: Two 16-state, rate 1/2 and rate 1/3 constituent codes are used to construct a rate 1/4 turbo code, as shown in Fig. 9, with $d_{ef}$ = 32, $d_3$ = 15 = $d_{min}$, $d_4$ = 16, $d_5$ = 17, $d_6$ = 16, and $d_7$ = 19.

(4) Rate 1/15 Turbo Code.

Code E: Two 16-state, rate 1/8 constituent codes are used to construct a rate 1/15 turbo code, $(1, g_1/g_0, g_2/g_0, g_3/g_0, g_4/g_0, g_5/g_0, g_6/g_0, g_7/g_0)$ and $(g_1/g_0, g_2/g_0, g_3/g_0, g_4/g_0, g_5/g_0, g_6/g_0, g_7/g_0)$, with $g_0 = (23)_{octal}$, $g_1 = (21)_{octal}$, $g_2 = (25)_{octal}$, $g_3 = (27)_{octal}$, $g_4 = (31)_{octal}$, $g_5 = (33)_{octal}$, $g_6 = (35)_{octal}$, and $g_7 = (37)_{octal}$. The (worst-case) minimum codeword weights, $d_i$, corresponding to a weight $i$ input sequence for this code are $d_{ef}$=142, $d_3$=39=$d_{min}$, $d_4$=48, $d_5$=45, $d_6$=50, and $d_7$=63.

The simulation performance of other codes reported in this article is still in progress.

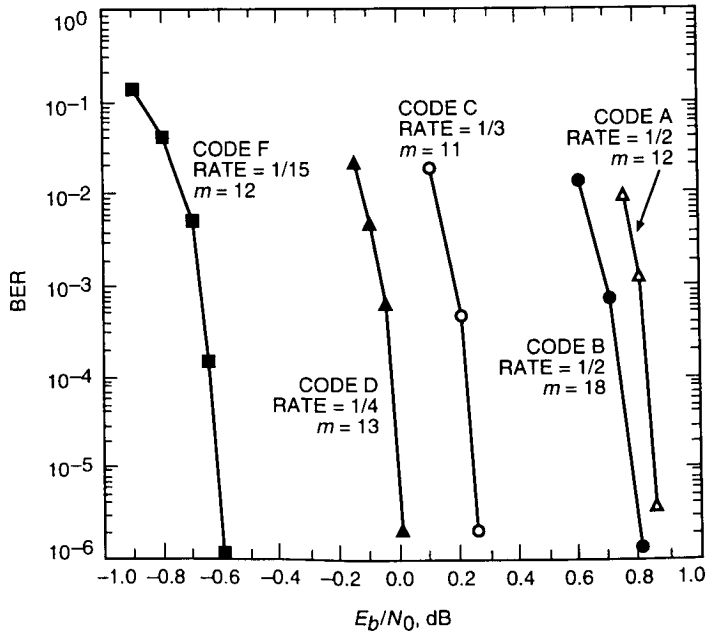


**Fig. 5. Performance of turbo codes.**

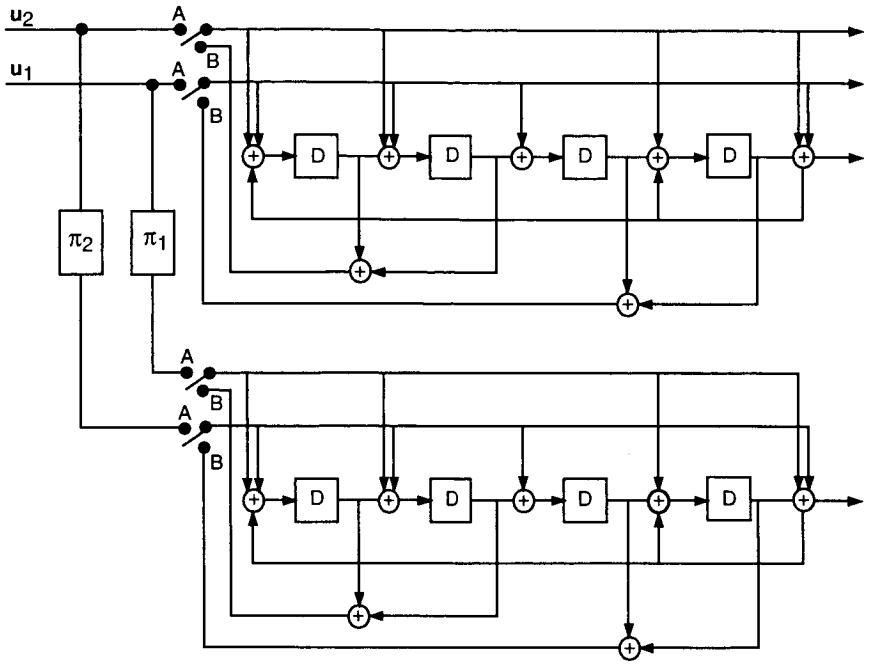


**Fig. 6. Rate 1/2 turbo code constructed from two codes ($h_0 = 23$, $h_1 = 35$, $h_2 = 33$).**

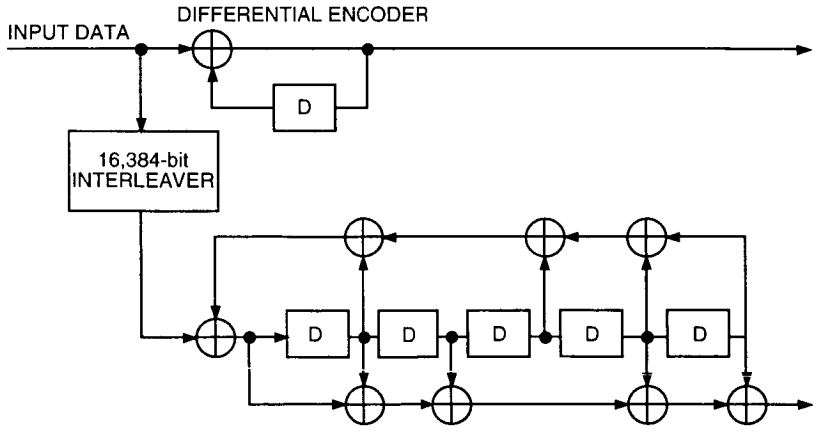


**Fig. 7. Rate 1/2 turbo code constructed from a differential encoder and code ($g_0 = 67$, $g_1 = 73$).**

## VI. Turbo Trellis-Coded Modulation

A pragmatic approach for turbo codes with multilevel modulation was proposed in [16]. Here we propose a different approach that outperforms the results in [16] when M-ary quadrature amplitude modulation (M-QAM) or M-ary phase shift keying (MPSK) modulation is used. A straightforward method for the use of turbo codes for multilevel modulation is first to select a rate $b/(b+1)$ constituent code, where the outputs are mapped to a $2^{b+1}$-level modulation based on Ungerboeck's set partitioning method [21] (i.e., we can use Ungerboeck's codes with feedback). If MPSK modulation is used, for every $b$ bits at the input of the turbo encoder, we transmit two consecutive $2^{b+1}$ phase-shift keying (PSK) signals, one per each encoder output. This results in a throughput of $b/2$ bits/s/Hz. If M-QAM modulation is

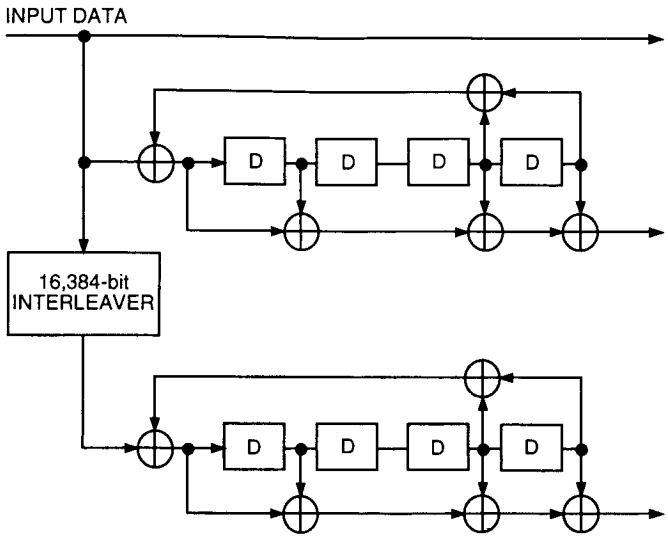


**Fig. 8. Rate 1/3 turbo code constructed from two identical codes ($g_0 = 23$, $g_1 = 33$).**

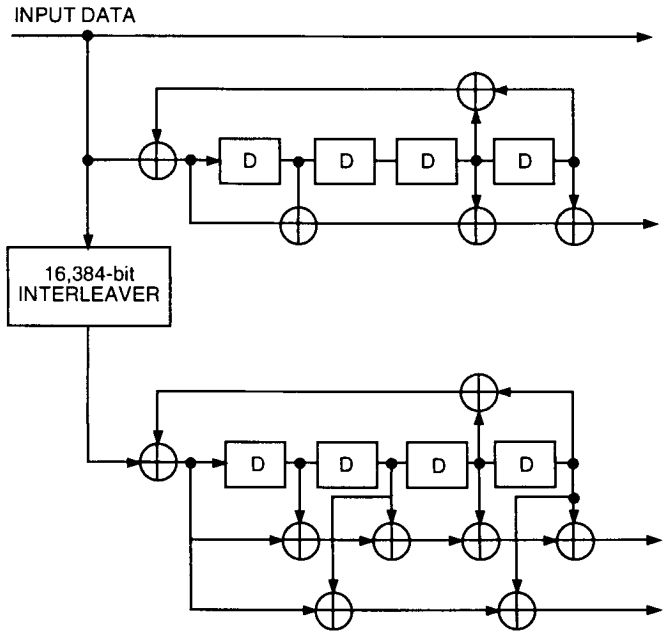


**Fig. 9. Rate 1/4 turbo code constructed from two codes ($g_0 = 23$, $g_1 = 33$) and ($g_0 = 23$, $g_1 = 37$, $g_2 = 25$).**

used, we map the $b + 1$ outputs of the first component code to the $2^{b+1}$ in-phase levels (I-channel) of a $2^{2b+2}$-QAM signal set and the $b + 1$ outputs of the second component code to the $2^{b+1}$ quadrature levels (Q-channel). The throughput of this system is $b$ bits/s/Hz.

First, we note that these methods require more levels of modulation than conventional trellis-coded modulation (TCM), which is not desirable in practice. Second, the input information sequences are used twice in the output modulation symbols, which also is not desirable. An obvious remedy is to puncture the output symbols of each trellis code and select the puncturing pattern such that the output symbols of the turbo code contain the input information only once. If the output symbols of the first encoder are

punctured, for example as $101010\cdots$, the puncturing pattern of the second encoder must be nonuniform to guarantee that all information symbols are used, and it depends on the particular choice of interleaver. Now, for example, for $2^{b+1}$ PSK, a throughput $b$ can be achieved. This method has two drawbacks: It complicates the encoder and decoder, and the reliability of punctured symbols may not be fully estimated at the decoder. A better remedy, for rate $b/(b+1)$ ($b$ even) codes, is discussed in the next section.

### A. A New Method to Construct Turbo TCM

For a $q = 2$ turbo code with rate $b/(b+1)$ constituent encoders, select the $b/2$ systematic outputs and puncture the rest of the systematic outputs, but keep the parity bit of the $b/(b+1)$ code (note that the rate $b/(b+1)$ code may have been obtained already by puncturing a rate 1/2 code). Then do the same to the second constituent code, but select only those systematic bits that were punctured in the first encoder. This method requires at least two interleavers: The first interleaver permutes the bits selected by the first encoder and the second interleaver those punctured by the first encoder. For MPSK (or M-QAM), we can use $2^{1+b/2}$ PSK symbols (or $2^{1+b/2}$ QAM symbols) per encoder and achieve throughput $b/2$. For M-QAM, we can also use $2^{1+b/2}$ levels in the I-channel and $2^{1+b/2}$ levels in the Q-channel and achieve a throughput of $b$ bits/s/Hz. These methods are equivalent to a multidimensional trellis-coded modulation scheme (in this case, two multilevel symbols per branch) that uses $2^{b/2} \times 2^{1+b/2}$ symbols per branch, where the first symbol in the branch (which depends only on uncoded information) is punctured. Now, with these methods, the reliability of the punctured symbols can be fully estimated at the decoder. Obviously, the constituent codes for a given modulation should be redesigned based on the Euclidean distance. In this article, we give an example for $b = 2$ with 16-QAM modulation where, for simplicity, we can use the 2/3 codes in Table 1 with Gray code mapping. Note that this may result in suboptimum constituent codes for multilevel modulation. The turbo encoder with 16 QAM and two clock-cycle trellis termination is shown in Fig. 10. The BER performance of this code with the turbo decoding structure for two codes discussed in Section IV is given in Fig. 11. For permutations $\pi_1$ and $\pi_2$, we used S-random permutations [9] with $S = 40$ and $S = 32$, with a block size of 16,384 bits. For 8 PSK, we used two 16-state, rate 4/5 codes given in Section V to achieve throughput 2. The parallel concatenated trellis codes with 8 PSK and two clock-cycle trellis termination is shown in Fig. 12. The BER performance of this code is given in Fig. 13. For 64 QAM, we used two 16-state, rate 4/5 codes given in Section V to achieve throughput 4. The parallel concatenated trellis codes with 64 QAM and two clock-cycle trellis termination is shown in Fig. 14. The BER performance of this code is given in Fig. 15. For permutations $\pi_1$, $\pi_2$, $\pi_3$, and $\pi_4$ in Figs. 10, 12, and 14, we used random permutations, each with a block size of 4096 bits. As was discussed above, there is no need to use four permutations; two permutations suffice, and they may even result in a better performance. Extension of the described method for construction of turbo TCM based on Euclidean distance is straightforward.[6]

## VII. Conclusions

In this article, we have shown that powerful turbo codes can be obtained if multiple constituent codes are used. We reviewed an iterative decoding method for multiple turbo codes by approximating the optimum bit decision rule. We obtained an upper bound on the effective free Euclidean distance of $b/n$ codes. We found the best rate 2/3, 3/4, 4/5, and 1/3 constituent codes that can be used in the design of multiple turbo codes. We proposed new schemes that can be used for power- and bandwidth-efficient turbo trellis-coded modulation.

[6] This is discussed in S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Parallel Concatenated Trellis Coded Modulation," submitted to *ICC '96*.

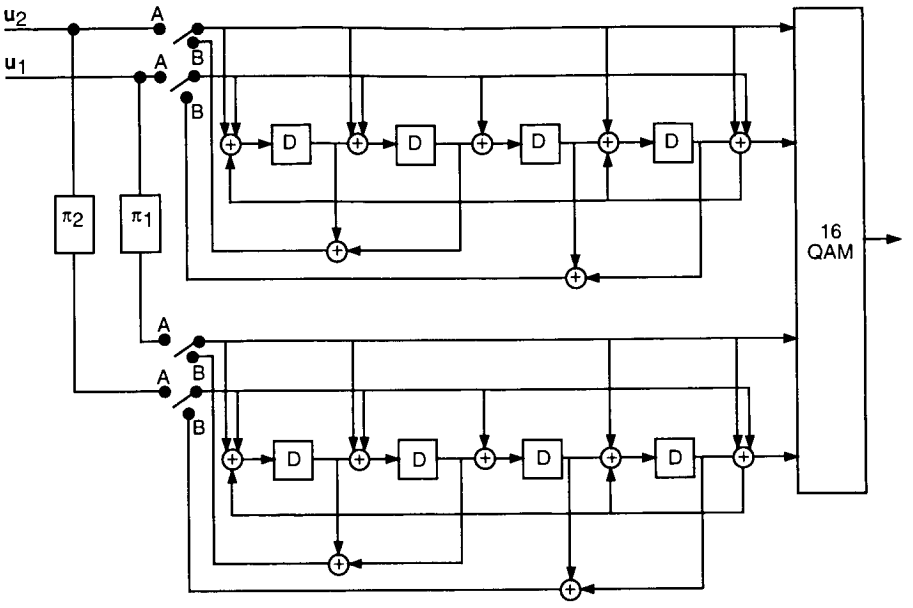


Fig. 10. Turbo trellis-coded modulation, 16 QAM, 2 bits/s/Hz.

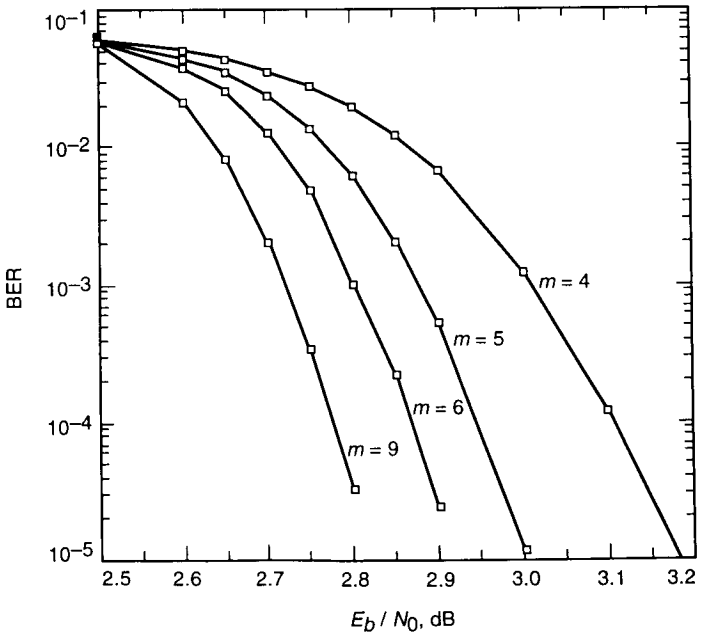


Fig. 11. BER performance of turbo trellis-coded modulation, 16 QAM, 2 bits/s/Hz.

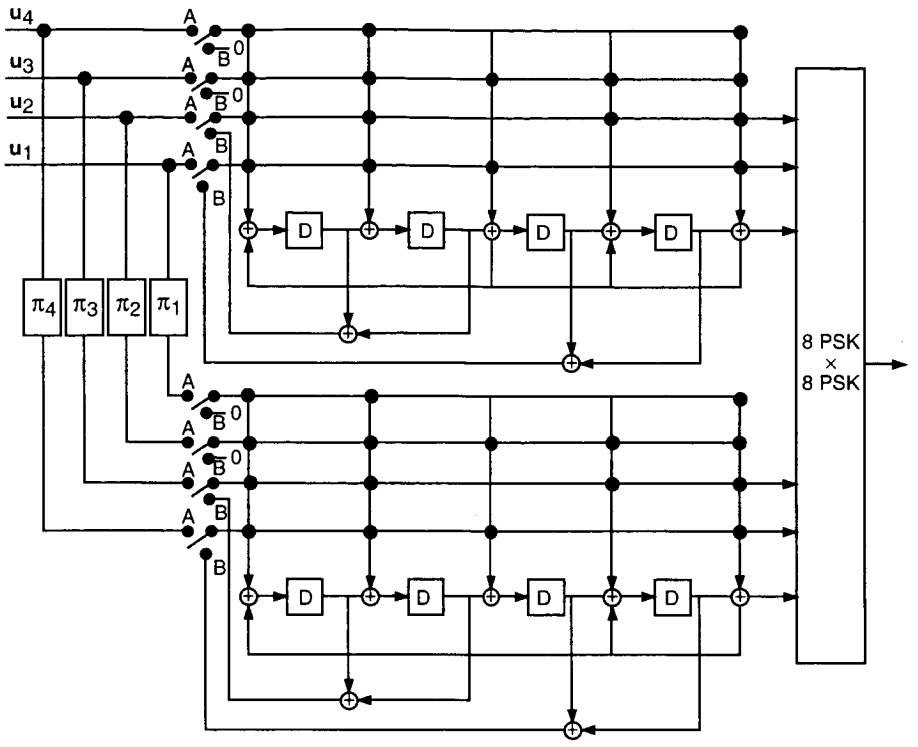


Fig. 12. Parallel concatenated trellis-coded modulation, 8 PSK, 2 bits/s/Hz.

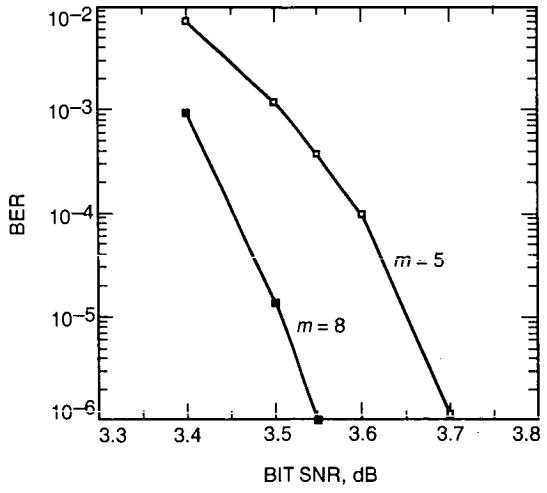


Fig. 13. BER performance of parallel concatenated trellis-coded modulation, 8 PSK, 2 bits/s/Hx.

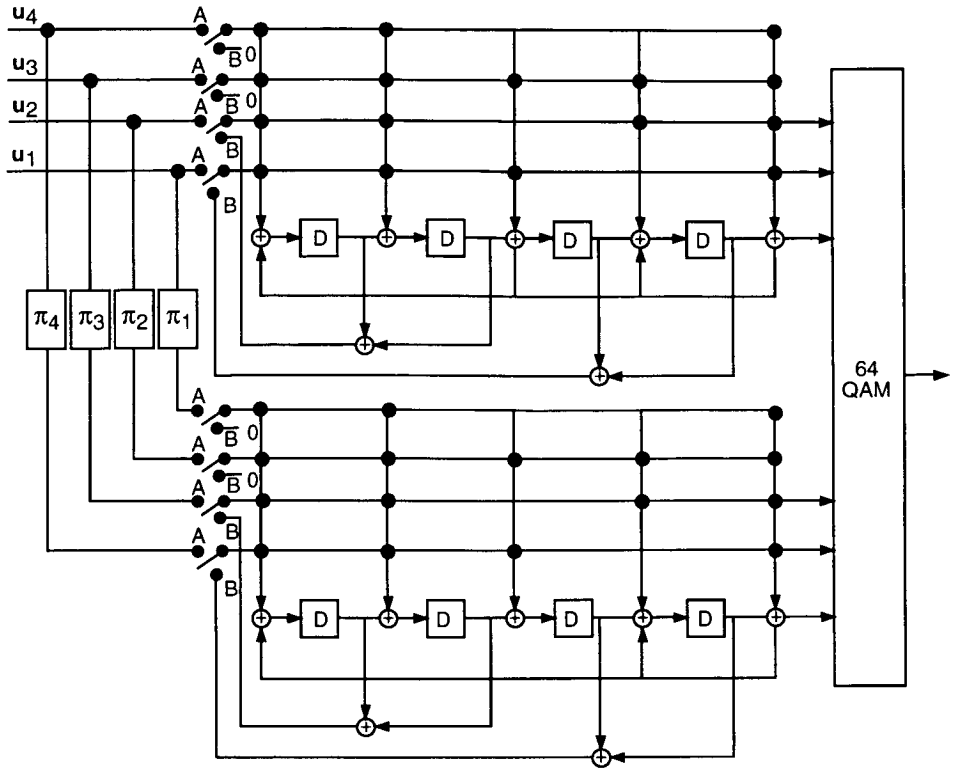


**Fig. 14. Parallel concatenated trellis-coded modulation, 64 QAM, 4 bits/s/Hz.**

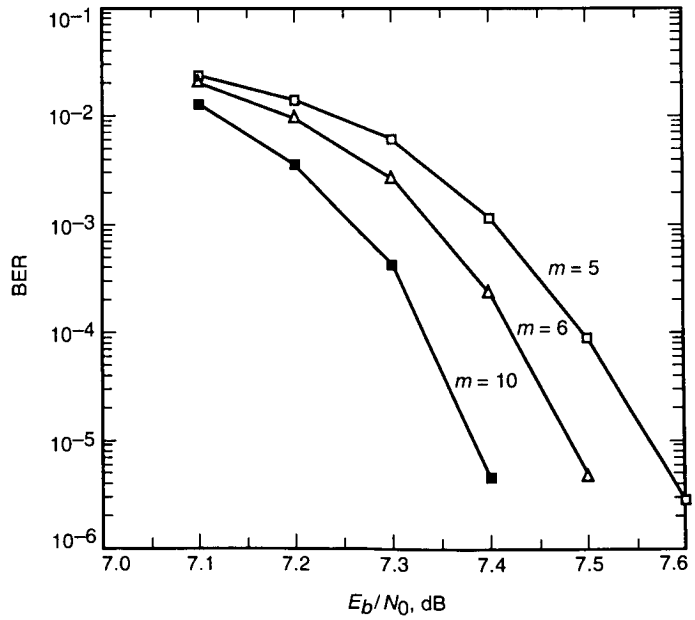


**Fig. 15. BER performance of parallel concatenated trellis-coded modulation, 64 QAM, 4 bits/s/Hz.**

## Acknowledgments

The authors are grateful to S. Dolinar and R. J. McEliece for their helpful comments throughout this article, to S. Benedetto and G. Montorsi for their helpful comments on the turbo trellis-coded modulation section, and special thanks to S. W. Golomb for his contribution, as reported in the Appendix.

## References

[1] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, 1974.

[2] G. Battail, C. Berrou, and A. Glavieux, "Pseudo-Random Recursive Convolutional Coding for Near-Capacity Performance," *Comm. Theory Mini-Conference, GLOBECOM '93*, Houston, Texas, December 1993.

[3] G. Battail and R. Sfez, "Suboptimum Decoding Using the Kullback Principle," *Lecture Notes in Computer Science*, vol. 313, pp. 93–101, 1988.

[4] S. Benedetto, "Unveiling Turbo Codes," IEEE Communication Theory Workshop, Santa Cruz, California, April 23–26, 1995.

[5] S. Benedetto and G. Montorsi, "Design of Parallel Concatenated Convolutional Codes," to be published in *IEEE Transactions on Communications*, 1996.

[6] S. Benedetto and G. Montorsi, "Performance Evaluation of Turbo-Codes," *Electronics Letters*, vol. 31, no. 3, pp. 163–165, February 2, 1995.

[7] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding: Turbo Codes," *Proc. 1993 IEEE International Conference on Communications*, Geneva, Switzerland, pp. 1064–1070, May 1993.

[8] D. Divsalar and F. Pollara, "Turbo Codes for Deep-Space Communications," *The Telecommunications and Data Acquisition Progress Report 42-120, October–December 1994*, Jet Propulsion Laboratory, Pasadena, California, pp. 29–39, February 15, 1995, URL http://edms-www.jpl.nasa.gov/tda/progress_report/42-120/120D.pdf.

[9] D. Divsalar and F. Pollara, "Multiple Turbo Codes for Deep-Space Communications," *The Telecommunications and Data Acquisition Progress Report 42-121, January–March 1995*, Jet Propulsion Laboratory, Pasadena, California, pp. 66–77, May 15, 1995, URL http://edms-www.jpl.nasa.gov/tda/progress_report/42-121/121T.pdf.

[10] D. Divsalar and F. Pollara, "Turbo Codes for PCS Applications," *Proceedings of IEEE ICC'95*, Seattle, Washington, pp. 54–59, June 1995.

[11] D. Divsalar and F. Pollara, "Turbo Codes for Deep-Space Communications," IEEE Communication Theory Workshop, Santa Cruz, California, April 23–26, 1995.

[12] D. Divsalar, S. Dolinar, R. J. McEliece,and F. Pollara, "Transfer Function Bounds on the Performance of Turbo Codes," *MILCOM 95*, San Diego, California, November 5–8, 1995.

[13] S. Dolinar and D. Divsalar, "Weight Distributions for Turbo Codes Using Random and Nonrandom Permutations," *The Telecommunications and Data Acquisition Progress Report 42-122, April–June 1995*, Jet Propulsion Laboratory, Pasadena, California, pp. 56–65, August 15, 1995, URL http://edms-www.jpl.nasa.gov/tda/progress_report/42-122/122B.pdf.

[14] S. W. Golomb, *Shift Register Sequences*, Revised Edition, Laguna Beach, California: Aegean Park Press, 1982.

[15] J. Hagenauer and P. Robertson, "Iterative (Turbo) Decoding of Systematic Convolutional Codes With the MAP and SOVA Algorithms," *Proc. of the ITG Conference on Source and Channel Coding*, Frankfurt, Germany, pp. 1–9, October 1994.

[16] S. LeGoff, A. Glavieux, and C. Berrou, "Turbo Codes and High Spectral Efficiency Modulation," *Proceedings of IEEE ICC'94*, New Orleans, Louisiana, pp. 645–651, May 1–5, 1994.

[17] M. Moher, "Decoding Via Cross-Entropy Minimization," *Proceedings GLOBECOM '93*, Houston, Texas, pp. 809–813, December 1993.

[18] A. S. Barbulescu and S. S. Pietrobon, "Terminating the Trellis of Turbo-Codes in the Same State," *Electronics Letters*, vol. 31, no. 1, pp. 22–23, January 1995.

[19] P. Robertson, "Illuminating the Structure of Code and Decoder of Parallel Concatenated Recursive Systematic (Turbo) Codes," *Proceedings GLOBECOM '94*, San Francisco, California, pp. 1298–1303, December 1994.

[20] G. D. Forney, Jr., "Convolutional Codes I: Algebraic Structure," *IEEE Transactions on Information Theory*, vol. IT-16, pp. 720–738, November 1970.

[21] G. Ungerboeck, "Channel Coding With Multi-Level Phase Signals," *IEEE Transactions on Information Theory*, vol. IT-28, pp. 55–67, January 1982.

# Appendix

# A Bound on the Weights of Shift Register Cycles[1]

## I. Introduction

A maximum-length linear shift register sequence—a pseudonoise (PN)-sequence or a maximal length (m)-sequence—of degree $m$ has period $p = 2^m - 1$, with $2^{m-1}$ ones and $2^{m-1} - 1$ zeroes in each period. Thus, the weight of a PN cycle is $2^{m-1}$. From a linear shift register whose characteristic polynomial is reducible, or irreducible but not primitive, in addition to the "zero-cycle" of period 1, there are several other possible cycles, depending on the initial state of the register, and each of these cycles has a period less than $2^m - 1$.

The question is whether it is possible for any cycle, from any linear shift register of degree $m$, to have a weight greater than $2^{m-1}$. We shall show that the answer is "no" and that this result does not depend on the shift register being linear.

### II. The Main Result

Let $S$ be any feedback shift register of length $m$, linear or not. We need not even specify that the shift register produce "pure" cycles, without branches. We will use only the fact that each state of the shift register has a unique successor state. For any given initial state, we define the length $L$ of the string starting from that state to be the number of states, counting from the initial state, prior to the second appearance of any state in the string. (In the case of branchless cycles, this is the length of the cycle with the given initial state.)

The string itself is this succession of states of length $L$. The corresponding string sequence is the sequence of 0's and 1's appearing in the right-most position of the register (or any other specific position of the register that has been agreed upon) as the string goes through its succession of $L$ states.

**Theorem 1.** From a feedback shift register $S$ of length $m$, the maximum number of 1's that can appear in any string sequence is $2^{m-1}$.

**Proof.** There are $2^m$ possible states of the shift register $S$ altogether. In any fixed position of the shift register, $2^{m-1}$ of these states have a 0 and $2^{m-1}$ states have a 1. In a string of length $L$, all $L$ of the states are distinct, and in any given position of the register, neither 0 nor 1 can occur more than $2^{m-1}$ times. In particular, the weight of a string sequence from a register of length $m$ cannot exceed $2^{m-1}$. ❐

**Corollary 1.** No cycle from a feedback shift register of length $m$ can have weight exceeding $2^{m-1}$.

[1] S. W. Golomb, personal communication, University of Southern California, Los Angeles, California, 1995.

57-32
6350 p. 18

N96-16691

TDA Progress Report 42-123 November 15, 1995

# The Trellis Complexity of Convolutional Codes

R. J. McEliece
Communications Systems and Research Section
and
California Institute of Technology
Pasadena, California

W. Lin[1]

*It has long been known that convolutional codes have a natural, regular trellis structure that facilitates the implementation of Viterbi's algorithm [30,10]. It has gradually become apparent that linear block codes also have a natural, though not in general a regular, "minimal" trellis structure, which allows them to be decoded with a Viterbi-like algorithm [2,31,22,11,27,14,12,16,24,25,8,15]. In both cases, the complexity of the Viterbi decoding algorithm can be accurately estimated by the number of trellis edges per encoded bit. It would, therefore, appear that we are in a good position to make a fair comparison of the Viterbi decoding complexity of block and convolutional codes. Unfortunately, however, this comparison is somewhat muddled by the fact that some convolutional codes, the punctured convolutional codes [4], are known to have trellis representations that are significantly less complex than the conventional trellis. In other words, the conventional trellis representation for a convolutional code may not be the minimal trellis representation. Thus, ironically, at present we seem to know more about the minimal trellis representation for block than for convolutional codes. In this article, we provide a remedy, by developing a theory of minimal trellises for convolutional codes. (A similar theory has recently been given by Sidorenko and Zyablov [29].) This allows us to make a direct performance-complexity comparison for block and convolutional codes. A by-product of our work is an algorithm for choosing, from among all generator matrices for a given convolutional code, what we call a trellis-minimal generator matrix, from which the minimal trellis for the code can be directly constructed. Another by-product is that, in the new theory, punctured convolutional codes no longer appear as a special class, but simply as high-rate convolutional codes whose trellis complexity is unexpectedly small.*

## I. Introduction

We begin with the standard definition of a convolutional code [9,26], always assuming that the underlying field is $F = GF(2)$. An $(n, k)$ convolutional code $\mathcal{C}$ is a $k$-dimensional subspace of $F(D)^n$, where $F(D)$ is the field of rational functions in the indeterminate $D$ over the field $F$. The memory, or degree,

[1] Graduate student at the California Institute of Technology, Pasadena, California.

of $\mathcal{C}$, is the smallest integer $m$ such that $\mathcal{C}$ has an encoder requiring only $m$ delay units. An $(n, k)$ convolutional code with memory $m$ is said to be a $(n, k, m)$ convolutional code. The free distance of $\mathcal{C}$ is the minimum Hamming weight of any codeword in $\mathcal{C}$. An $(n, k, m)$ convolutional code with free distance $d$ is said to be an $(n, k, m, d)$ code.

A minimal generator matrix $G(D)$ for an $(n, k, m)$ convolutional code $\mathcal{C}$ is a $k \times n$ matrix with polynomial entries, whose row space is $\mathcal{C}$, such that the direct-form realization of an encoder for $\mathcal{C}$ based on $G(D)$ uses exactly $m$ delay elements [9,26]. From a minimal generator matrix $G(D)$, or rather from a physical encoder built using $G(D)$ as a blueprint, it is possible to construct a conventional trellis representation for $\mathcal{C}$. This trellis is, in principle, infinite, but it has a very regular structure, consisting (after a short initial transient) of repeated copies of what we shall call the "trellis module" associated with $G(D)$. The trellis module consists of $2^m$ initial states and $2^m$ final states, with each initial state being connected by a directed edge to exactly $2^k$ final states. Thus, the trellis module has $2^{k+m}$ edges. Each edge is labeled with an $n$-bit binary vector, namely, the output produced by the encoder in response to the given state transition. Thus, each edge has length (measured in edge labels) $n$, and so the total edge length of the conventional trellis module is $n2^{k+m}$. Since each trellis module represents the encoder's response to $k$ input bits, we are led to define the conventional trellis complexity of the trellis module as

$$\frac{n}{k} \cdot 2^{m+k} \quad \text{edge labels per encoded bit} \tag{1}$$

or edges per bit, for short. If the code $\mathcal{C}$ is decoded using Viterbi's maximum-likelihood algorithm on the trellis [30,10], the work factor involved in updating the metrics and survivors at each trellis module is proportional to the edge length of the trellis module, so that the trellis complexity as defined in Eq. (1) is a measure of the effort *per decoded bit* required by Viterbi's algorithm. (For a more detailed discussion of the complexity of Viterbi's algorithm on a trellis, see [25, Section 2].)

For example, consider the $(3, 2, 2)$ convolutional code with minimal generator matrix given by

$$G_1(D) = \begin{pmatrix} 1+D & 1+D & 1 \\ D & 0 & 1+D \end{pmatrix} \tag{2}$$

This code has the largest possible free distance, viz., $d_{\text{free}} = 3$, for any $(3, 2, 2)$ code. A "direct-form" encoder based on the generator matrix $G_1(D)$ is shown in Fig. 1. If the input pair is $(u_1, u_2)$ and the state of the encoder is $(s, t)$, then the output $(x_1, x_2, x_3)$ is given by

$$\left.\begin{aligned} x_1 &= u_1 + s + t \\ x_2 &= u_1 + s \\ x_3 &= u_1 + u_2 + t \end{aligned}\right\} \tag{3}$$

and the "next state" is just the input pair $(u_1, u_2)$. The conventional trellis module for the code with minimal generator matrix $G_1(D)$ given in Eq. (2) is shown in Fig. 2. The three-bit edge label on the edge from $(s, t)$ to $(u_1, u_2)$ is the triple $(x_1, x_2, x_3)$ given in Eq. (3). The total edge length is 48, so that the conventional trellis complexity corresponding to the matrix $G_1(D)$ is $48/2 = 24$ edges per bit, as predicted by Eq. (1).

But we can do substantially better than this, if we use the fact that this particular code is a punctured convolutional code. We now briefly review the theory of punctured convolutional codes to see how simplified trellises result.

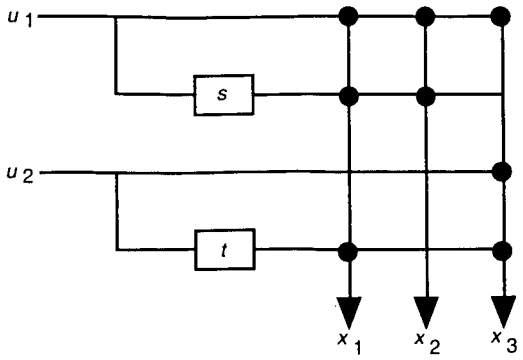


**Fig. 1. A direct-form encoder based on the generator matrix $G_1(D)$ in Eq. (2). The input is $(u_1, u_2)$, the output is $(x_1, x_2, x_3)$, and the state of the encoder is $(s, t)$. (The boxes labeled $s$ and $t$ are unit delay elements.)**

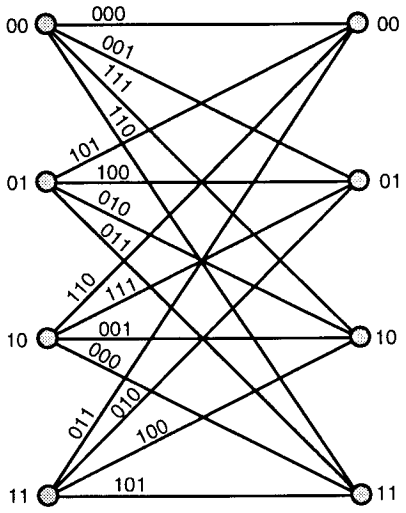


**Fig. 2. The conventional trellis module for the code with minimal generator matrix $G_1(D)$ given in Eq. (2).**

If we begin with a parent $(N, 1, m)$ convolutional code, and block it to depth $k$, i.e., group the input bit stream into blocks of $k$ bits each, the result is an $(Nk, k, m)$ convolutional code. If we now delete, or puncture, all but $n$ bits from each $Nk$-bit output block, the result is an $(n, k, m)$ convolutional code.[2] This punctured code can be represented by a trellis whose trellis module is built from $k$ copies of the trellis modules from the parent $(N, 1, m)$ code, each of which has only $2^{m+1}$ edges, so that the total number of edge labels on the trellis module is $n \cdot 2^{m+1}$, which means that the trellis complexity of an $(n, k, m)$ punctured code is

$$\frac{n}{k} \cdot 2^{m+1} \quad \text{edges per bit} \tag{4}$$

which is a factor of $2^{k-1}$ smaller than the complexity of the conventional trellis given in Eq. (1). For $k = 1$, this is no improvement, but for larger values of $k$, the decoding complexity reduction afforded

[2] In fact, the memory of the punctured code may be less than $m$, but for most interesting punctured codes, no memory reduction will take place.

by puncturing becomes increasingly significant. And while the class of punctured convolutional codes is considerably smaller than the class of unrestricted convolutional codes, nevertheless many punctured convolutional codes with good performance properties are known [4,13,3,7], and punctured convolutional codes, especially high-rate ones, are often preferred in practice.

For example, consider the $(2, 1, 2, 5)$ convolutional code defined by the minimal generator matrix

$$G_2(D) = (\, 1 + D + D^2 \quad 1 + D^2 \,) \tag{5}$$

The conventional trellis module for this code is shown in Fig. 3. If we block this code into blocks of size $k = 2$, we obtain a $(4, 2, 2)$ convolutional code, still with $d_{\text{free}} = 5$, for which the conventional trellis module is two copies of the trellis module shown in Fig. 3; see Fig. 4.

Now we can do the puncturing. Take the $(4, 2, 2)$ code, as represented by the trellis module in Fig. 4, and delete the second output bit on each of the edges in the second part of the module. The result is shown in Fig. 5. This structure can be thought of as the trellis module for a $(3, 2, 2)$ code; the corresponding $d_{\text{free}}$ turns out to be 3. According to Eq. (1), the conventional trellis complexity of a $(3, 2, 2)$ code is $3/2 \cdot 2^4 = 24$ edges per bit. But if we use instead the punctured trellis corresponding to the $k = 2$ blocked version of the parent $(2, 1, 2)$ code, we find from Eq. (4), or Fig. 5, that the trellis complexity is instead only $3/2 \cdot 2^3 = 12$ edges per bit. In fact, it can be shown that this punctured $(3, 2, 2)$ code is the same as the conventional code with generator matrix $G_1(D)$ given in Eq. (2). (Indeed, this example is taken almost verbatim from [4], where it was used to illustrate the way puncturing can reduce decoding complexity.)

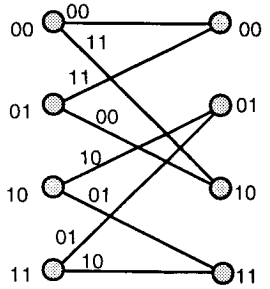


**Fig. 3. The trellis module for the (2,1,2) code with generator matrix $G_2(D)=(1 + D + D^2 \quad 1 + D^2)$; total edge length is 16, so the trellis complexity is 16 edges per bit.**

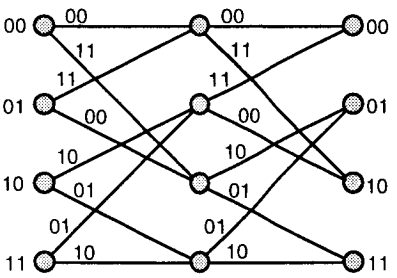


**Fig. 4. The trellis module for the (4,2,2) code obtained from the code of Fig. 3 by blocking the inputs in blocks of size 2; total edge length is 32, so the trellis complexity is 32/2 = 16 edges per bit, the same as for the original code.**

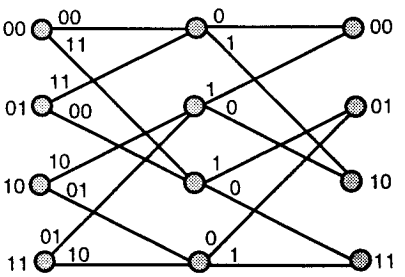


**Fig. 5. The trellis module for the (3,2,2) punctured code obtained from the code of Fig. 4 by deleting every fourth bit; total edge length is 16 + 8 = 24, so the trellis complexity is 24/2 = 12 edges per bit.**

It seems mysterious that an ordinary-looking generator matrix like $G_1(D)$ produces a code whose trellis complexity can be significantly reduced (if one knows that it is, in fact, a punctured code), whereas for an almost identical code, say one defined by the generator matrix

$$\begin{pmatrix} 1+D & D & 1+D \\ D & 1 & 1 \end{pmatrix}$$

no such reduction is apparently possible. In Section II, we will resolve this mystery by developing a simple algorithm for constructing the minimum possible trellis for any convolutional code. Our technique will always find a simplified trellis for a punctured code, with complexity at least as small as that given by Eq. (4), even if we are not told in advance that the code can be obtained by puncturing. But more

important, it will often result in considerable simplification of the trellis representation of a convolutional code that is not a punctured code. We will illustrate this with worked examples in Sections II and III and numerical tables in the Appendix.

## II. Construction of Minimal Trellises

If $G(D)$ is a minimal generator matrix for an $(n, k, m)$ convolutional code $\mathcal{C}$, then we can write $G(D)$ in the form

$$G(D) = G_0 + G_1 D + \cdots + G_L D^L \tag{6}$$

where $G_0, \cdots, G_L$ are $k \times n$ scalar matrices (i.e., matrices whose entries are from $GF(2)$), and $L$ is the maximum degree of any entry of $G(D)$. If we concatenate the $L+1$ matrices $G_0, \cdots, G_L$, we obtain a $k \times (L+1)n$ scalar matrix, which we denote by $\widetilde{G}$:

$$\widetilde{G} = (\, G_0 \quad G_1 \quad \cdots \quad G_L \,) \tag{7}$$

It is well known [23, Chapter 9] that the matrix $\widetilde{G}$ and its shifts can be used to build a scalar generator matrix $G_{\text{scalar}}$ for the code $\mathcal{C}$ (for simplicity of notation, we illustrate the case $L = 2$):

$$G_{\text{scalar}} = \begin{bmatrix} G_0 & G_1 & G_2 & & & \\ & G_0 & G_1 & G_2 & & \\ & & G_0 & G_1 & G_2 & \\ & & & G_0 & G_1 & G_2 \\ & & & & \ddots & \end{bmatrix} \tag{8}$$

The matrix in Eq. (8) is, except for the fact that it continues forever, the generator matrix for a binary block code (with a very regular structure), and so the techniques that have been developed for finding minimal trellises for block codes are useful for constructing trellis representations for convolutional codes. Here we apply the techniques developed in [25, Section 7], which show how to construct a trellis directly from any generator matrix for a given block code, and the minimal trellis if the generator is in minimal span form, to construct a trellis for $\mathcal{C}$ based on the infinite scalar generator matrix $G_{\text{scalar}}$.

The trellis module for the trellis associated with $G_{\text{scalar}}$ corresponds to the $(L+1)k \times n$ matrix module,

$$\widehat{G} = \begin{pmatrix} G_L \\ G_{L-1} \\ \vdots \\ G_0 \end{pmatrix} \tag{9}$$

which repeatedly appears as a vertical "slice" in $G_{\text{scalar}}$. Using the techniques in [25, Section 7], it is easy to show that the number of edges in this trellis module is

$$\text{edge count} = \sum_{j=1}^{n} 2^{a_j} \tag{10}$$

where $a_j$ is the number of active entries in the $j$th column of the matrix module $\widehat{G}$. (An element is called active if it belongs to the active span of one of the rows of $\widetilde{G}$. We will elaborate on this below.) Our object, then, is to find a generator matrix for which the edge count in the corresponding trellis module is as small as possible.

To clarify these ideas, we consider the $(3, 2, 1)$ code with (minimal) generator matrix

$$G_3(D) = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1+D & 1+D \end{pmatrix} \tag{11}$$

According to Eq. (1), the conventional trellis complexity for this code is 12 edges per bit. However, we can do better. The scalar matrix $\widetilde{G_3}$ corresponding to $G_3(D)$ is [cf. Eq. (7)]

$$\widetilde{G_3} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{1} & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} \end{pmatrix} \tag{12}$$

In Eq. (12), we have shown the active elements of each row, i.e., the entries from the first nonzero entry to the last nonzero entry, in boldface. The span length of (i.e., the number of active entries in) the first row is, therefore, three; and the span length of the second row is six. The matrix module corresponding to $\widetilde{G_3}$ is [cf. Eq. (9)]

$$\widehat{G_3} = \begin{pmatrix} 0 & 0 & 0 \\ \mathbf{0} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} \end{pmatrix}$$

Thus, $a_1 = 3$, $a_2 = 3$, and $a_3 = 3$, which by Eq. (10) means that the corresponding trellis module has $2^3 + 2^3 + 2^3 = 24$ edges. Since each trellis module represents two encoded bits, the resulting trellis complexity is $24/2 = 12$ edges per bit. Since we have already noted that the conventional trellis complexity for this code is also 12 edges per bit, the trellis corresponding to $G_3(D)$ is not better than (in fact, it is isomorphic to) the conventional trellis. To do better, we need to find a generator matrix for the code for which $\sum_i 2^{a_i}$ is less than 24. Using the results of [25, Section 6], it is possible to show that minimizing $\sum_i 2^{a_i}$ is equivalent to minimizing $\sum_i a_i$, i.e., the total span length of the corresponding $\widetilde{G}$, and so we shall look for generator matrices for which the span of $\widetilde{G}$ is reduced.

Note that if we add the first row of $G_3(D)$ to the second row, the resulting generator matrix, which is still minimal, is

$$G_3'(D) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1+D & D \end{pmatrix}$$

The scalar matrix $\widetilde{G_3'}$ corresponding to $G_3'(D)$ is [cf. Eq. (7)]

$$\widetilde{G_3'} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} \end{pmatrix} \tag{13}$$

The span length of the first row of $G_3'(D)$ is three, and the span length of the second row is five, and so the total span length is eight, one less than that of $G_3(D)$. The matrix module corresponding to $\widetilde{G_3'}$ is [cf. Eq. (9)]

$$\widehat{G'_3} = \begin{pmatrix} 0 & 0 & 0 \\ \mathbf{0} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} & \mathbf{1} \\ 0 & 1 & 0 \end{pmatrix}$$

Here $a_1 = 2, a_2 = 3$, and $a_3 = 3$, and so by Eq. (10) the corresponding trellis module has $2^2 + 2^3 + 2^3 = 20$ edges, so that the resulting trellis complexity is $20/2 = 10$ edges per bit. The trellis module itself, constructed using the technique described in [25, Section 7] is shown in Fig. 6.

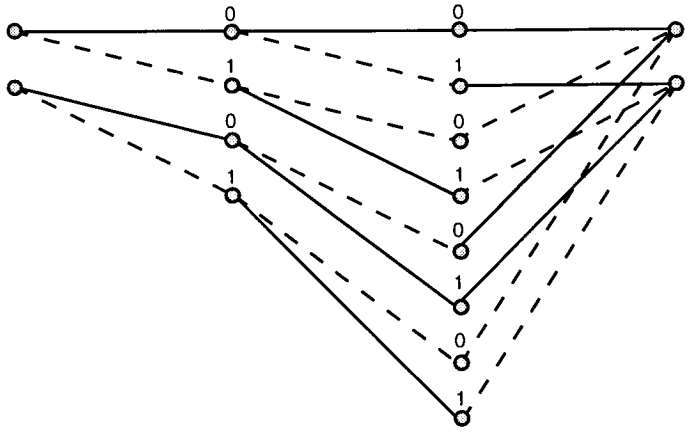


**Fig. 6. The trellis module for the (3,2,1) code with generator matrix $G_3'(D)$. (Solid edges represent "0" code bits, and dashed edges represent "1" code bits. The labels on the vertices correspond to the information bits.)**

But we can do still better. If we multiply the first row of $G'_3(D)$ by $D$ and add it to the second row, the resulting generator matrix, which is still minimal, is

$$G''_3(D) = \begin{pmatrix} 1 & 0 & 1 \\ D & 1+D & 0 \end{pmatrix}$$

The scalar matrix $\widetilde{G''_3}$ corresponding to $G''_3(D)$ is [cf. Eq. (7)]

$$\widetilde{G''_3} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} & 0 \end{pmatrix} \tag{14}$$

The span length of $G''_3(D)$ is seven, one less than that of $G'_3(D)$. The matrix module corresponding to $\widetilde{G''_3}$ is [cf. Eq. (9)]

$$\widehat{G''_3} = \begin{pmatrix} 0 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{0} & \mathbf{1} \\ 0 & 1 & \mathbf{0} \end{pmatrix}$$

Here $a_1 = 2, a_2 = 3$, and $a_3 = 2$, and so by Eq. (10), the corresponding trellis module has $2^2 + 2^3 + 2^2 = 16$ edges, so that the resulting trellis complexity is $16/2 = 8$ edges per bit. The trellis module itself, again constructed using the techniques described in [25, Section 7] is shown in Fig. 7.

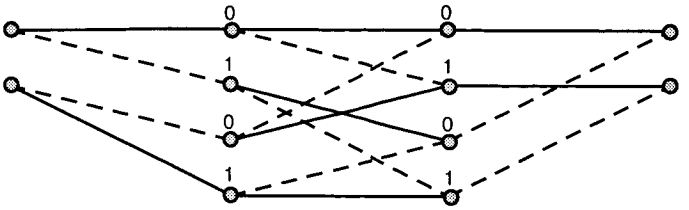


**Fig. 7. The trellis module for the (3,2,1) code with generator matrix $G_3^{\prime\prime}(D)$. This is the minimal trellis module for this code.**

Furthermore, it is easy to see that there is no generator matrix for this code with span length less than seven, so that the trellis module shown in Fig. 7 yields the minimal trellis for the code. Alternatively, we examine the scalar generator matrix for the code corresponding to $\widetilde{G_3''}$ [cf. Eq. (8)]:

$$G_{\text{scalar}} = \begin{bmatrix} \underline{1} & 0 & \overline{1} & 0 & 0 & 0 & & & & & & & \\ 0 & \underline{1} & 0 & 1 & \overline{1} & 0 & & & & & & & \\ & & & \underline{1} & 0 & \overline{1} & 0 & 0 & 0 & & & & \\ & & & 0 & \underline{1} & 0 & 1 & \overline{1} & 0 & & & & \\ & & & & & & 0 & \underline{1} & 0 & \overline{1} & 0 & 0 & 0 \\ & & & & & & 0 & \underline{1} & 0 & 1 & \overline{1} & 0 & \\ & & & & & & & & & \ddots & & & \end{bmatrix} \tag{15}$$

In Eq. (15), we see that $G_{\text{scalar}}$ has the property that no column contains more than one underlined entry, the leftmost nonzero entry in its row (L), or more than one overlined entry, the rightmost nonzero entry in its row (R). Thus, $G_{\text{scalar}}$ has the LR property, and so, *if it were a finite matrix*, it would produce the minimal trellis for the code [25, Sec. 6]. To circumvent the problem that $G_{\text{scalar}}$ is infinite, we can define the $M$th truncation of the code $\mathcal{C}$, denoted by $\mathcal{C}^{[M]}$, as the $((M+L)n, Mk)$ block code obtained by taking only the first $Mk$ rows of $G_{\text{scalar}}$, i.e., the code with $Mk \times (M+L)n$ generator matrix

$$G_{\text{scalar}}^{[M]} = \begin{bmatrix} \underline{1} & 0 & \overline{1} & 0 & 0 & 0 & & & & & & \\ 0 & \underline{1} & 0 & 1 & \overline{1} & 0 & & & & & & \\ & & & \underline{1} & 0 & \overline{1} & 0 & 0 & 0 & & & \\ & & & 0 & \underline{1} & 0 & 1 & \overline{1} & 0 & & & \\ & & & & & & & & & \ddots & & \\ & & & & & & \underline{1} & 0 & \overline{1} & 0 & 0 & 0 \\ & & & & & & 0 & \underline{1} & 0 & 1 & \overline{1} & 0 \end{bmatrix} \tag{16}$$

Plainly, if $G_{\text{scalar}}$ has the LR property, so does $G_{\text{scalar}}^{[M]}$, for all $M \geq 1$. Thus, it follows from the standard theory of trellises for block codes that the matrix $G_{\text{scalar}}^{[M]}$ produces the minimal trellis for $\mathcal{C}^{[M]}$, for all $M$, and so we can safely call the infinite trellis, built from trellis modules corresponding to $\widehat{G}$, *the* minimal trellis for the code. (Note that, in this example, the ratio of the conventional trellis complexity to the minimal trellis complexity is $12/8 = 3/2$. If this code were punctured, then according to Eqs. (1) and (4), the ratio would be at least 2. Thus, we conclude that the code with generator matrix $G_3(D)$ as given in Eq. (11) is not a punctured code, which shows that the theory of minimal trellises for convolutional codes goes beyond merely "explaining" punctured codes.)

The preceding argument, though it was presented in terms of a specific example, is entirely general. It shows that a basic generator matrix $G(D)$ produces a minimal trellis if and only if $G(D)$ has the property that the span length of the corresponding $\widetilde{G}$ cannot be reduced by an operation of the form

$$g_i(D) \leftarrow g_i(D) + D^{\ell} g_j(D)$$

where $g_i(D)$ is the $i$th row of $G(D)$ and $\ell$ is an integer in the range $0 \leq \ell \leq L$. We shall call a generator matrix with this property a trellis-minimal generator matrix for $\mathcal{C}$. A trellis-minimal generator matrix must be minimal, but the converse need not be true, as the example of this section shows. Furthermore, it can be shown that the set of trellis-minimal generator matrices for a given code $\mathcal{C}$ coincides with the set of generator matrices for which the span length of the corresponding $\widetilde{G}$ is a minimum. In the next section, we will give two more examples of minimal trellises.

## III. Two More Examples

Our first example is for the code whose generator matrix is given in Eq. (2). The corresponding decomposition [cf. Eq. (6)] is

$$G_1(D) = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} D$$

The scalar matrix $\widetilde{G}$ is, thus,

$$\widetilde{G}_1 = \begin{pmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & 0 \\ 0 & 0 & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} \end{pmatrix}$$

and the matrix module $\widehat{G}$ from Eq. (9) is then

$$\widehat{G}_1 = \begin{pmatrix} \mathbf{1} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} \\ 0 & 0 & \mathbf{1} \end{pmatrix} \tag{17}$$

Since there are three active entries in each column of $\widehat{G}$, it follows from Eq. (10) that the edge count for the trellis module is $2^3 + 2^3 + 2^3 = 24$, so that the trellis complexity for this trellis module is $24/2 = 12$ edges per bit, the same as given by Eq. (4) for the punctured trellis. To actually construct the trellis module, we can use the techniques of [25, Section 7], and the result is shown in Fig. 8. Finally, we note that the $G_{\text{scalar}}$ corresponding to the matrix $\widehat{G}_1$ of Eq. (17) is [cf. Eq. (8)]

$$\begin{bmatrix} \underline{1} & 1 & 1 & 1 & \overline{1} & 0 & & & & & & \\ 0 & 0 & \underline{1} & 1 & 0 & \overline{1} & & & & & & \\ & & & \underline{1} & 1 & 1 & 1 & \overline{1} & 0 & & & \\ & & & 0 & 0 & \underline{1} & 1 & 0 & \overline{1} & & & \\ & & & & & & \underline{1} & 1 & 1 & 1 & \overline{1} & 0 \\ & & & & & & 0 & 0 & \underline{1} & 1 & 0 & \overline{1} \\ & & & & & & & \ddots & & & & \end{bmatrix}$$

which has the LR property, and so $G_1(D)$ is trellis-minimal. (This code is the first code listed in Table 2 in the Appendix.)

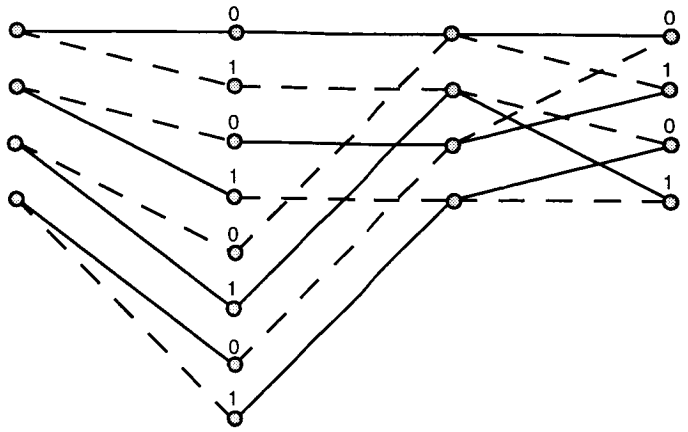


**Fig. 8. The trellis module for the (3,2,2) code with generator matrix $G_1(D)$. This module is isomorphic to the one in Fig. 5.**

As our second example, we consider a partial-unit-memory code, taken from [20,1]. It is an $(8, 4, 3)$ code with $d_{\text{free}} = 8$ and with minimal generator matrix (as taken from [1])

$$G(D) = \begin{pmatrix} 11111111 \\ 11101000 \\ 10110100 \\ 10011010 \end{pmatrix} + \begin{pmatrix} 00000000 \\ 11011000 \\ 10101100 \\ 10010110 \end{pmatrix} D \tag{18}$$

The conventional trellis complexity for this code is, by Eq. (1), $8/4 \cdot 2^7 = 256$ edges per bit. We can reduce this number to 120, as follows. First, we concatenate the two matrices in Eq. (18), obtaining the following $4 \times 16$ scalar matrix $\widetilde{G}$:

$$\widetilde{G} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Next, using the techniques developed in [25, Section 6], we perform a series of elementary row operations on $\widetilde{G}$, transforming it to the minimal span, or trellis oriented form, $\widetilde{G}'$:

$$\widetilde{G}' = \begin{pmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & 0 \end{pmatrix} \tag{19}$$

The matrix module $\widehat{G}$ defined in Eq. (9) is, thus,

$$\widehat{G} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ 0 & 0 & 0 & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ 0 & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ 0 & 0 & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} \end{pmatrix}$$

and so by Eq. (10) the total edge length of the trellis module is $2^4 + 2^5 + 2^6 + 2^7 + 2^7 + 2^6 + 2^5 + 2^4 = 480$. Since each trellis module represents four encoded bits, it follows that the trellis complexity is $480/4 = 120$ edges per bit, compared to the conventional trellis complexity, cited above, of 256 edges per bit.

The matrix $G_{\text{scalar}}$ corresponding to the matrix $\widetilde{G}'$ in Eq. (19) is easily seen to have the LR property, and so the generator matrix [cf. Eq. (19)]

$$G'(D) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \end{pmatrix} D$$

is trellis-minimal. However, the trellis complexity can be reduced still further, if we allow column permutations of the original generator matrix $G(D)$ in Eq. (18). Indeed, by computer search, we have found that one minimal complexity column permutation for this particular code is the permutation (01243567), which results in the generator matrix [cf. Eq. (18)]

$$G(D) = \begin{pmatrix} 11111111 \\ 11110000 \\ 10101100 \\ 10011010 \end{pmatrix} + \begin{pmatrix} 00000000 \\ 11011000 \\ 10110100 \\ 10001110 \end{pmatrix} D \tag{20}$$

Then, after putting the minimal generator matrix of Eq. (20) into trellis-minimal form, it becomes

$$G(D) = \begin{pmatrix} 11111111 \\ 00001111 \\ 01111111 \\ 00111111 \end{pmatrix} + \begin{pmatrix} 00000000 \\ 11111000 \\ 11111100 \\ 11111110 \end{pmatrix} D \tag{21}$$

The trellis complexity of the generator matrix in Eq. (21) turns out to be 104 edges per encoded bit. (This code is the seventh code listed in Table 6 in the Appendix.) The minimal trellis complexity of unit memory and partial unit memory convolutional codes has also been studied in [6] and [32].

## IV. LTC Versus ACG

In this section, we will attempt to compare the trellis complexity of a number of codes to their performance. To do this, we define the logarithmic trellis complexity (LTC) of a code, block or convolutional, as the base-2 logarithm of the minimal trellis complexity (edges per encoded bit) and the asymptotic coding gain (ACG) as the code's rate times its minimum (or free) distance. An empirical study, based on existing tables of convolutional codes (e.g., the tables in [19,28,20,5,7]), reveals the interesting fact that LTC / ACG lies between 1.5 and 2.0 for most "good" convolutional codes. For example, for the $(3, 2, 2, 3)$

code discussed in Section III, the ratio is 1.79, and for the $(8, 4, 3, 8)$ code, it is 1.68. By comparison, for the "NASA standard" $(2, 1, 6, 10)$ convolutional code, for which, as for all $(n, 1, m)$ convolutional codes, the minimal trellis complexity is given by the formula of Eq. (1), the ratio is 1.60. In the Appendix, we list the (ACG, LTC) pairs for a large number of convolutional codes and a few block codes. In Fig. 9, we show a scatter plot of these pairs. It is interesting to note how close most of these pairs are to the line of slope 2. This experimental fact may be related to a recent theorem of Lafourcade and Vardy [18], which implies that for any sequence of block codes with a fixed rate $R > 0$ and fixed value of $d/n > 0$, as $n \to \infty$,

$$\liminf_{n\to\infty} \frac{\text{LTC}}{\text{ACG}} \geq 2 \tag{22}$$

In any case, we have been able to show that for all codes, the ratio LTC / ACG must be strictly greater than 1. (This result is similar to Theorem 3 in [17].)

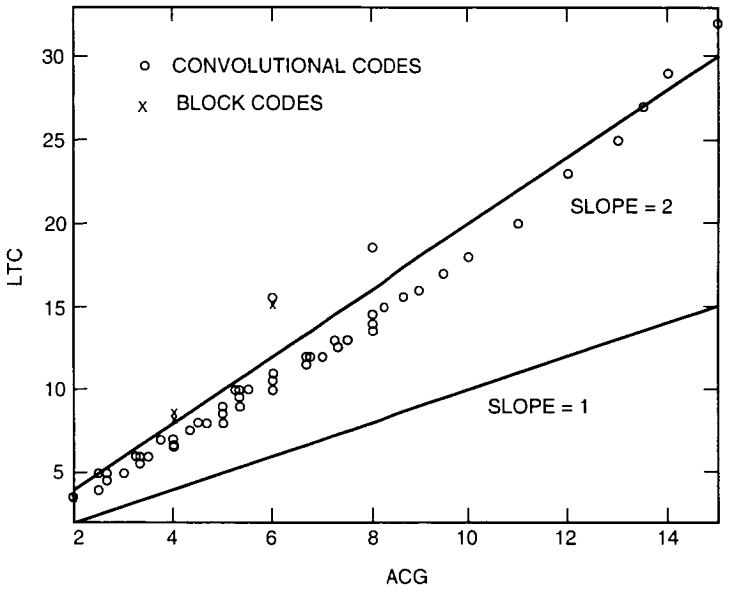


**Fig. 9. A scatter plot of the pairs (ACG, LTC) for the codes listed in the Appendix.**

## V. Conclusion and Open Problems

In this article, we have shown that every convolutional code has a unique minimal trellis representation, which is in many cases considerably simpler than the conventional trellis for the code. We have also presented a simple technique for actually constructing the minimal trellis for any convolutional code, and we have numerically computed the trellis complexity for many convolutional codes. In principle, the theory of minimal trellises for convolutional codes can be deduced from the general Forney–Trott theory [12], but we believe the observation that the Viterbi decoding complexity of many convolutional codes, including many nonpunctured codes, can thereby be reduced systematically is new, as are the details of the algorithms for producing the minimal trellises.

We close with a list of research problems that suggest themselves.

(1) A given convolutional code will, in general, have many different minimal generator matrices [21], but as we saw in Section II, not all minimal generator matrices are trellis minimal. What can be said about the class of trellis minimal generator matrices?

(2) A theoretical explanation of the experimental observation that most of the codes shown in Fig. 9 lie near the line of slope 2 would be welcome.

(3) The design and implementation of Viterbi's algorithm on conventional trellises is well understood. Since the techniques described here lead to greatly reduced trellis complexity, it would be worthwhile to make a careful study of how best to implement Viterbi's algorithm on minimal trellises.

(4) From our current viewpoint, punctured convolutional codes are just codes whose trellis module has fewer edges than would normally be expected. Indeed, it is easy to prove that the minimal trellis complexity of any punctured convolutional code is at least as small as the punctured trellis complexity given in Eq. (4). This is because in the scalar matrix $\widetilde{G}$ for a punctured code, certain entries are guaranteed to be zero. For example, for a $(4,3,3)$ punctured code, the matrix $\widetilde{G}$ has the template structure

$$\widetilde{G} = \begin{pmatrix} x & x & x & x & x & x & 0 & 0 \\ 0 & 0 & x & x & x & x & x & 0 \\ 0 & 0 & 0 & x & x & x & x & x \end{pmatrix}$$

where the $x$'s can be arbitrary (actually, there are restrictions on the $x$'s that depend in detail on how the code is constructed), but the eight zero positions must be respected. Any $(4,3,3)$ convolutional code with such a template structure will have trellis complexity at most $4/3 \cdot 2^4 = 211/3$. An obvious question is whether other low complexity templates support good convolutional codes.

(5) In our computer-aided search for the "best" column permutation of the $(8,4,3,8)$ code, we found that each of the $8! = 40,326$ possible column permutations had minimal trellis complexity of either 120 or 104. This strongly suggests an equivalence among permutations that, if understood theoretically, could make it much simpler to find the best column permutation.

Finally, we remark that when the bulk of this article was written, we were not aware of the important earlier work of Sidorenko and Zyablov [29], which deals explicitly with the minimal trellis for a convolutional code, and we wish to acknowledge their priority. Their work, like ours, develops the theory of minimal trellises for convolutional codes from the corresponding theory for block codes. However, their trellis construction is based on the parity-check matrix of the code rather than the generator matrix, and their emphasis is quite different. One advantage of the Sidorenko–Zyablov approach is that it leads to the following upper bound on the number of nodes at depth $i$ in the minimal trellis for a $(n,k,m)$ convolutional code [29, Theorem 1]:

$$N_i \leq 2^{m+\min(k,n-k)}$$

It is not easy to derive this bound using our methods. On the other hand, the present article contains a number of things not present in [29], among them being

(1) The observation that the minimal trellis for a punctured convolutional code is at least as simple as the punctured trellis.

(2) The concept of a trellis-minimal generator matrix for a convolutional code, and an algorithm for computing one.

(3) The ACG versus LTC comparison for block and convolutional codes.

## References


[1] K. Abdel-Ghaffar, R. J. McEliece, and G. Solomon, "Some Partial Unit Memory Convolutional Codes," *The Telecommunications and Data Acquisition Progress Report 42-107, July–September 1991*, Jet Propulsion Laboratory, Pasadena, California, pp. 57–72, November 15, 1991. Also see *Proc. 1991 International Symposium on Information Theory*, Budapest, p. 196, June 1991.

[2] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, March 1974.

[3] G. Bégin and D. Haccoun, "High-Rate Punctured Convolutional Codes: Structure Properties and Construction Technique," *IEEE Trans. Comm.*, vol. COM-37, pp. 1381–1385, December 1989.

[4] J. B. Cain, G. C. Clark, and J. M. Geist, "Punctured Convolutional Codes of Rate $(n-1)/n$ and Simplified Maximum Likelihood Decoding," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 97–100, January 1979.

[5] D. G. Daut, J. W. Modestino, and L. D. Wismer, "New Short Constraint Length Convolutional Code Construction for Selected Rational Rates," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 794–800, September 1982.

[6] U. Dettmar and U. Sorger, "On Maximum Likelihood Decoding of Unit Memory Codes," *Proc. 6th Swedish–Russian International Workshop on Information Theory*, pp. 184–188, August 1993.

[7] A. Dholakia, *Introduction to Convolutional Codes with Applications*, Boston: Kluwer Academic Publishers, 1994.

[8] S. Dolinar, L. Ekroot, A. Kiely, R. McEliece, and W. Lin, "The Permutation Trellis Complexity of Linear Block Codes," *Proc. 32nd Annual Allerton Conference on Communication, Control, and Computing*, Allerton Park, Illinois, pp. 60–74, September 1994.

[9] G. D. Forney Jr., "Convolutional Codes I: Algebraic Structure," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 268–278, November 1970.

[10] G. D. Forney, Jr., "The Viterbi Algorithm," *Proc. IEEE*, vol. 61, pp. 268–276, March 1973.

[11] G. D. Forney, Jr., "Coset Codes—Part II: Binary Lattices and Related Codes," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 1152–1187, September 1988.

[12] G. D. Forney, Jr., and M. D. Trott, "The Dynamics of Group Codes: State Spaces, Trellis Diagrams, and Canonical Encoders," *IEEE Trans. Inform. Theory*, vol. IT-39, pp. 1491–1513, September 1993.

[13] D. Haccoun and G. Bégin, "High-Rate Punctured Convolutional Codes for Viterbi and Sequential Decoding," *IEEE Trans. Comm.*, vol. COM-37, pp. 1113–1125, November 1989.

[14] B. Honary, G. Markarian, and M. Darnell, "Trellis Decoding for Block Codes," *Proc. 3rd IEE Int. Symp. Comm. Theory Appl.*, Ambleside, United Kingdom, pp. 79–93, July 1993.

[15] A. Kiely, S. Dolinar, R. McEliece, L. Ekroot, and W. Lin, "Trellis Decoding Complexity of Linear Block Codes," to appear in *IEEE Trans. Inform. Theory*, vol. IT-42, November 1996.

[16] F. R. Kschischang and V. Sorokine, "On the Trellis Structure of Block Codes," *IEEE Trans. Inform. Theory*, vol. IT-41, November 1995, in press.

[17] A. Lafourcade and A. Vardy, "Asymptotically Good Codes Have Infinite Trellis Complexity," *IEEE Trans. Inform. Theory*, vol. IT-41, pp. 555–559, March 1995.

[18] A. Lafourcade and A. Vardy, "Lower Bounds on Trellis Complexity of Block Codes," *IEEE Trans. Inform. Theory*, vol. IT-41, November 1995, in press.

[19] K. Larsen, "Short Convolutional Codes With Maximal Free Distance for Rates 1/2, 1/3, and 1/4," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 371–372, May 1973.

[20] G. S. Lauer, "Some Optimal Partial-Unit-Memory Codes," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 540–547, March 1979.

[21] K. Lumbard and R. J. McEliece, "Counting Minimal Generator Matrices," *Proc. 1994 IEEE Inter. Symp. Inform. Theory*, Trondheim, Norway, p. 18, June 1994.

[22] J. L. Massey, "Foundations and Methods of Channel Coding," *Proc. Int. Conf. Inform. Theory and Systems*, NTG-Fachberichte, vol. 65, pp. 148–157, 1978.

[23] R. J. McEliece, *The Theory of Information and Coding,* Reading, Massachusetts: Addison-Wesley, 1977.

[24] R. J. McEliece, "The Viterbi Decoding Complexity of Linear Block Codes," *Proc. 1994 IEEE Inter. Symp. Inform. Theory*, Trondheim, Norway, p. 341, June 1994.

[25] R. J. McEliece, "On the BCJR Trellis," to appear in *IEEE Trans. Inform. Theory.*, vol. IT-42, 1996.

[26] R. J. McEliece, "The Algebraic Theory of Convolutional Codes," to appear as a chapter in the *Handbook of Coding Theory*, edited by R. A. Brualdi, W. C. Huffman, and V. Pless, Amsterdam: Elsinore Science Publishers, 1996.

[27] D. J. Muder, "Minimal Trellises for Block Codes," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 1049–1053, September 1988.

[28] E. Paaske, "Short Binary Convolutional Codes With Maximal Free Distance," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 683–688, September 1974.

[29] V. Sidorenko and V. Zyablov, "Decoding of Convolutional Codes Using a Syndrome Trellis," *IEEE Trans. Inform. Theory*, vol. IT-40, pp. 1663–1666, September 1994.

[30] A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 260–269, April 1967.

[31] J. K. Wolf, "Efficient Maximum Likelihood Decoding of Linear Block Codes," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 76–80, January 1978.

[32] V. Zyablov and V. Sidorenko, "Soft Decision Maximum Likelihood Decoding of Partial Unit Memory Codes," *Problems of Information Transmission*, vol. 28, no. 1, pp. 18–22, July 1992.

# Appendix

# Tables of LTC Versus ACG

In this appendix, we list the ACG and the LTC for a large number of "good" convolutional codes and a few block codes. A scatter plot of these (ACG, LTC) pairs appears as Fig. 9 in Section IV.

**Table 1. Best (2,1,*m*) codes.**[a]

| Code | LTC | ACG | LTC–ACG ratio |
|---|---|---|---|
| (2,1,2,5) | 4 | 2.5 | 1.60 |
| (2,1,3,6) | 5 | 3 | 1.67 |
| (2,1,4,7) | 6 | 3.5 | 1.71 |
| (2,1,5,8) | 7 | 4 | 1.75 |
| (2,1,6,10) | 8 | 5 | 1.60 |
| (2,1,8,12) | 10 | 6 | 1.67 |
| (2,1,10,14) | 12 | 7 | 1.71 |
| (2,1,11,15) | 13 | 7.5 | 1.73 |
| (2,1,12,16) | 14 | 8 | 1.75 |
| (2,1,14,18) | 16 | 9 | 1.78 |
| (2,1,15,19) | 17 | 9.5 | 1.79 |
| (2,1,16,20) | 18 | 10 | 1.80 |
| (2,1,18,22) | 20 | 11 | 1.82 |
| (2,1,21,24) | 23 | 12 | 1.92 |
| (2,1,23,26) | 25 | 13 | 1.92 |
| (2,1,25,27) | 27 | 13.5 | 2.00 |
| (2,1,27,28) | 29 | 14 | 2.07 |
| (2,1,30,30) | 32 | 15 | 2.13 |

[a] From pp. 85–88 in [7].

**Table 2. Best (3,2,*m*) codes.**[a]

| Code | LTC | ACG | LTC–ACG ratio |
|---|---|---|---|
| (3,2,2,3) | 3.58 | 2.00 | 1.79 |
| (3,2,3,4) | 5.00 | 2.67 | 1.87 |
| (3,2,4,5) | 6.00 | 3.33 | 1.80 |
| (3,2,5,6) | 7.00 | 4.00 | 1.75 |
| (3,2,6,7) | 8.00 | 4.67 | 1.71 |
| (3,2,7,8) | 9.00 | 5.33 | 1.69 |
| (3,2,8,8) | 10.00 | 5.33 | 1.88 |
| (3,2,9,9) | 11.00 | 6.00 | 1.83 |
| (3,2,10,10) | 12.00 | 6.67 | 1.80 |

[a] From p. 90 in [7].

**Table 3. Best (4,3,*m*) codes.**[a]

| Code | LTC | ACG | LTC–ACG ratio |
|---|---|---|---|
| (4,3,3,4) | 5.00 | 3.00 | 1.67 |
| (4,3.5,5) | 7.00 | 3.75 | 1.87 |
| (4,3,6,6) | 8.00 | 4.50 | 1.78 |
| (4,3,8,7) | 10.00 | 5.25 | 1.90 |
| (4,3,9,8) | 11.00 | 6.00 | 1.83 |

[a] From p. 90 in [7].

**Table 4. Best (3,1,*m*) codes.**[a]

| Code | LTC | ACG | LTC–ACG ratio |
|---|---|---|---|
| (3,1,2,8) | 4.58 | 2.67 | 1.72 |
| (3,1,3,10) | 5.58 | 3.33 | 1.68 |
| (3,1,4,12) | 6.58 | 4.00 | 1.64 |
| (3,1,5,13) | 7.58 | 4.33 | 1.75 |
| (3,1,6,15) | 8.58 | 5.00 | 1.72 |
| (3,1,7,16) | 9.58 | 5.33 | 1.80 |
| (3,1,8,18) | 10.58 | 6.00 | 1.76 |
| (3,1,9,20) | 11.58 | 6.67 | 1.74 |
| (3,1,10,22) | 12.58 | 7.33 | 1.72 |
| (3,1,11,24) | 13.58 | 8.00 | 1.70 |
| (3,1,12,24) | 14.58 | 8.00 | 1.82 |
| (3,1,13,26) | 15.58 | 8.67 | 1.80 |

[a] From p. 89 in [7].

**Table 5. Best (4,1,*m*) codes.**[a]

| Code | LTC | ACG | LTC–ACG ratio |
|---|---|---|---|
| (4,1,2,10) | 5.00 | 2.50 | 2.00 |
| (4,1,3,13) | 6.00 | 3.25 | 1.85 |
| (4,1,4,16) | 7.00 | 4.00 | 1.75 |
| (4,1,5,18) | 8.00 | 4.50 | 1.78 |
| (4,1,6,20) | 9.00 | 5.00 | 1.80 |
| (4,1,7,22) | 10.00 | 5.50 | 1.82 |
| (4,1,8,24 | 11.00 | 6.00 | 1.83 |
| (4,1,9,27) | 12.00 | 6.75 | 1.78 |
| (4,1,10,29) | 13.00 | 7.25 | 1.79 |
| (4,1,11,32) | 14.00 | 8.00 | 1.75 |
| (4,1,12,33) | 15.00 | 8.25 | 1.82 |
| (4,1,13,36) | 16.00 | 9.00 | 1.78 |

[a] From p. 89 in [7].

**Table 6. Some block codes and partial unit memory convolutional codes.**

| Code | LTC | ACG | LTC–ACG ratio |
|---|---|---|---|
| [8,4,4] Self-dual code | 3.46 | 2.00 | 1.73 |
| [24,12,8] Golay code | 8.22 | 4.00 | 2.06 |
| [32,16,8] BCH[a] code | 8.64 | 4.00 | 2.16 |
| [48,24,12] Self-dual code | 15.13 | 6.00 | 2.52 |
| $[n, n-1, 2]$ Parity-check code | 2.00 | $\frac{2(n-1)}{n}$ | $\frac{n}{n-1}$ |
| $[n, 1, n]$ Repetition code | $1 + \log_2 n$ | 1 | $1 + \log_2 n$ |
| (8,4,3,8) PUM[b] code | 6.70 | 4.00 | 1.68 |
| (24,12,7,12) PUM code | 15.58 | 6.00 | 2.60 |
| (24,12,10,16) PUM code | 18.58 | 8.00 | 2.32 |

[a] Bose–Chaudhuri–Hocquenghem.

[b] Partial unit memory.

# System Noise Temperature Investigation of the DSN S-Band Polarization Diverse Systems for the Galileo S-Band Contingency Mission

J. E. Fernandez and D. L. Trowbridge
Communications Ground Systems Section

*This article describes measurements made at all three Deep Space Network 70-m S-band polarization diverse (SPD) systems to determine and eliminate the cause of the 1-K elevation in follow-up noise temperature in the listen-only mode of the SPD systems at DSS 43 and DSS 63. The system noise temperatures obtained after finding and correcting the cause of the elevated follow-up noise temperature are also reported.*

## I. Introduction

In response to the Galileo spacecraft's X-band (8.45 GHz) antenna deployment failure, an emergency effort to optimize S-band (2.3 GHz) downlink performance was conducted. As part of this effort, termed the Galileo S-band Contingency Mission, the three 70-m DSN S-band polarization diverse (SPD) systems in the listen-only mode (see Fig. 1) have been carefully evaluated. Results of this initial evaluation were that both DSS 43 and DSS 63 at the Canberra and Madrid Deep Space Communications Complexes, respectively, exhibited elevated follow-up noise temperature contributions—defined as the contribution to system operating noise temperature of all components following the first low-noise amplifier (LNA)—of 1.25 K in comparison with the predicted values of 0.35 K. The system noise temperature predictions for these systems are shown in Tables 1 and 2. During the course of the evaluation, DSS-43 personnel determined the cause of this elevated follow-up noise temperature contribution in their 70-m SPD system to be due to a nonstandard configuration. This problem was corrected, and the antenna subsequently performed within predicted performance limits.

The reason for the elevated follow-up noise temperature at DSS 63, also determined during the course of this work, was a higher than normal attenuation level in the system path behind the maser LNA. Once this attenuation was reduced by about 4 dB, the measured follow-up noise temperature at DSS 63 of 0.4 K agreed very closely with the predicted value of 0.35 K. Also documented during the course of the investigation were high and dissimilar noise figures of the S-band Block IV receivers at DSS 14, as well as differences in gain of the right-hand circular polarization (RCP) and left-hand circular polarization (LCP) channels of the very long baseline interferometry (VLBI) downconverter at DSS 63.

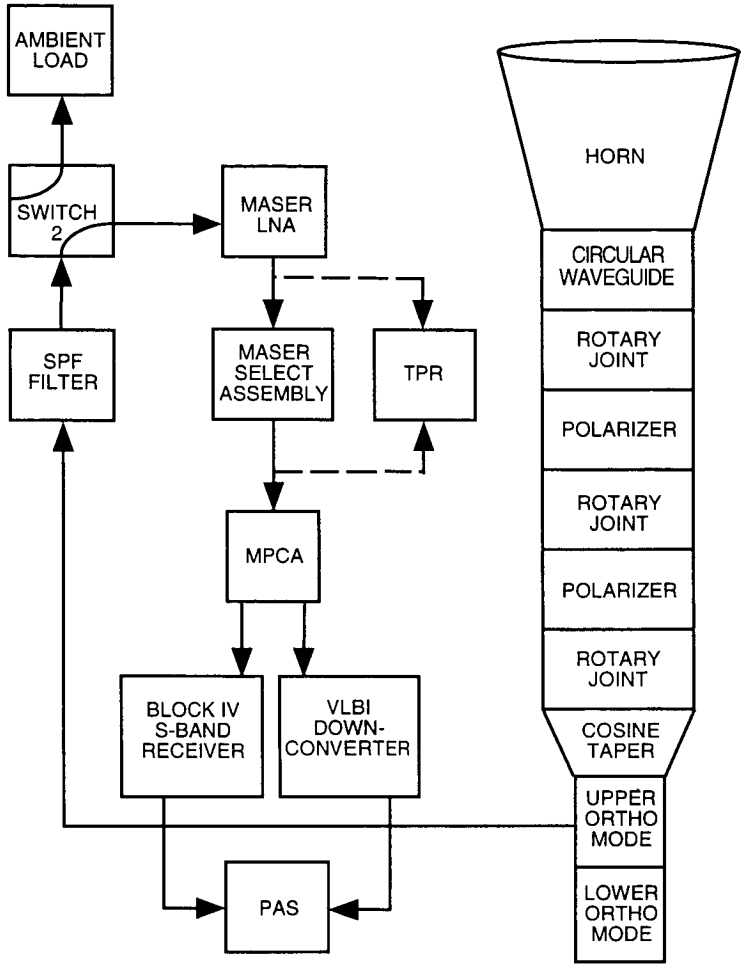


**Fig. 1. The 70-m SPD diverse front-end listen-only mode for DSS 43 and DSS 63.**

## II. Noise Measurement Technique

In the course of this investigation, two types of noise measurements made at different points along the front-end component string were needed to assess where the noise temperature problem was located. The first measurement is the total system noise temperature, $T_{op}$, while the antenna is pointed at zenith. The noise power level is measured using the 50-MHz precision attenuator assembly (PAS) or other suitable receiver. Switch 2 in Fig. 1 is switched so that the maser LNA is on the sky. Switch 2 is then switched so that the maser input is on the ambient load. The difference in power levels, $Y$, the Y-factor in dB, is then used to determine the $T_{op}$:

$$T_{op} = \frac{T_{load} + T_{rcvr}}{10^{Y/10}} \tag{1}$$

where $T_{load}$ is the ambient load temperature, K, and $T_{rcvr}$ is the maser input noise temperature, K, including the follow-up noise temperature (FNT) contribution. This approach is extremely accurate for $T_{rcvr} << T_{load}$ and requires independent knowledge of $T_{rcvr}$.

The second measurement required is the FNT. This is accomplished by measuring the difference in power level in dB, the Y-factor, by using the PAS or other suitable receiver, with the maser switched to

**Table 1. Noise budget for DSS 14.[a]**

| Component | Part number | $\Delta T$, K | $L$, dB | $T$, K | $G$, dB | $T_{in}$, K | Noise term, K |
|---|---|---|---|---|---|---|---|
| Cosmic background | — | 2.7 | — | — | — | — | 2.657 |
| Atmosphere | — | 1.8 | — | — | — | — | 1.772 |
| Antenna spillover | — | 1.44 | — | — | — | — | 1.417 |
| Antenna scatter | — | 2.3 | — | — | — | — | 2.264 |
| Main reflector | — | 0.08 | — | — | — | — | 0.079 |
| Subreflector | — | 0.07 | — | — | — | — | 0.069 |
| Main reflector gap leakage | — | 0.1 | — | — | — | — | 0.098 |
| Feedhorn | 9449420-1 | — | 0.003 | 293 | — | — | 0.199 |
| Waveguide round | 9457310-1 | — | 0.0015 | 293 | — | — | 0.100 |
| Waveguide round, 15.141 in. + rotary joint 1 in. | — | — | 0.0018 | 293 | — | — | 0.120 |
| Rotary joints (2) | 9457311-1 | — | 0.003 | 293 | — | — | 0.200 |
| Polarizers (1) | 9449405-1 | — | 0.0035 | 293 | — | — | 0.233 |
| Cosine taper | 9457389-1 | — | 0.002 | 293 | — | — | 0.133 |
| Orthomode, upper | 9457308-1 | — | 0.005 | 293 | — | — | 0.333 |
| Matched coupler, 40 dB injected = 0.029 K | 9457331-1 | — | 0.00345 | 293 | — | — | 0.230 |
| Elbow, H-plane | 9451160-2 | — | 0.0037 | 293 | — | — | 0.247 |
| S-band passband filter | 9430960 | — | 0.021 | 293 | — | — | 1.406 |
| 3-position switch | 9443100-1 | — | 0.008 | 293 | — | — | 0.538 |
| Elbow, E-plane | 9451159-2 | — | 0.0037 | 293 | — | — | 0.249 |
| Straight, 13 in. | 9459426-3 | — | 0.003 | 293 | — | — | 0.202 |
| 35-dB coupler (loss) | SR8148D | — | 0.0066 | 293 | — | — | 0.445 |
| 35-dB coupler (injected) | — | — | 35 | 293 | — | — | 0.093 |
| Maser/CCR VSWR[b] (not used) | — | 0.00 | — | — | — | — | 0.000 |
| Maser/CCR package | — | — | — | — | 45 | 2 | 2.000 |
| LP filter | — | — | 0.1 | 293 | — | — | 0.000 |
| 10-dB couplers (2) | — | — | 1 | 293 | — | — | 0.002 |
| Cable loss, 1/2 in. spiraline | — | — | 1.7 | 293 | — | — | 0.006 |
| Maser select box | — | — | 0.8 | 293 | — | — | 0.004 |
| Cable | — | — | 0.5 | 293 | — | — | 0.003 |
| Loss | — | — | 0 | 0 | — | — | 0.000 |
| Avantek amplifier | AN-2200N | — | — | — | 25 | 870 | 0.071 |
| 20-dB coupler | — | — | 20 | 293 | — | — | 0.007 |
| Cable | — | — | 1 | 293 | — | — | 0.002 |
| Downconverter $T_{in}$ | — | — | — | — | — | 8881 | 0.287 |
| Total antenna system noise temperature (referred to input of maser) | | | | | | | 15.47 |
| Follow-up noise contribution $T_f$ | | | | | | | 0.382 |

[a] SPD feedcone, low-noise path, 2295 MHz, 90-deg elevation angle, and clear weather.
[b] Closed-cycle refrigerator (CCR) voltage standing-wave ratio (VSWR).

**Table 2. Noise budget for DSS 43 and DSS 63.[a]**

| Component | Part number | $\Delta T$, K | $L$, dB | $T$, K | $G$, dB | $T_{in}$, K | Noise term, K |
|---|---|---|---|---|---|---|---|
| Cosmic background | — | 2.7 | — | — | — | — | 2.656 |
| Atmosphere | — | 1.9 | — | — | — | — | 1.869 |
| Antenna spillover | — | 1.44 | — | — | — | — | 1.416 |
| Antenna scatter | — | 2.3 | — | — | — | — | 2.262 |
| Main reflector | — | 0.08 | — | — | — | — | 0.079 |
| Subreflector | — | 0.07 | — | — | — | — | 0.069 |
| Main reflector gap leakage | — | 0.1 | — | — | — | — | 0.098 |
| Feedhorn | 9449420-1 | — | 0.003 | 293 | — | — | 0.199 |
| Waveguide round | 9457310-1 | — | 0.0015 | 293 | — | — | 0.100 |
| Rotary joints (3) | 9457311-1 | — | 0.0045 | 293 | — | — | 0.299 |
| Polarizers (2) | 9449405-1 | — | 0.007 | 293 | — | — | 0.466 |
| Cosine taper | 9457389-1 | — | 0.002 | 293 | — | — | 0.133 |
| Orthomode, upper | 9457308-1 | — | 0.005 | 293 | — | — | 0.333 |
| Matching section, upper | 9457331-1 | — | 0.003 | 293 | — | — | 0.200 |
| Elbow, H-plane | 9451160-2 | — | 0.0037 | 293 | — | — | 0.247 |
| S-band passband filter | 9430960 | — | 0.021 | 293 | — | — | 1.406 |
| 3-position switch | 9443100-1 | — | 0.008 | 293 | — | — | 0.538 |
| Elbow, E-plane | 9451159-2 | — | 0.0037 | 293 | — | — | 0.249 |
| Straight, 13 in. | 9459426-3 | — | 0.003 | 293 | — | — | 0.202 |
| 35-dB coupler (loss) | SR8148D | — | 0.0066 | 293 | — | — | 0.445 |
| 35-dB coupler (injected) | — | — | 35 | 293 | — | — | 0.093 |
| Maser/CCR[b] package | — | — | — | — | 45 | 2 | 2.000 |
| LP filter | — | — | 0.1 | 293 | — | — | 0.000 |
| 10-dB couplers (2) | — | — | 1 | 293 | — | — | 0.002 |
| Cable loss, 1/2 in. spiraline | — | — | 1.7 | 293 | — | — | 0.006 |
| Maser select box | — | — | 0.8 | 293 | — | — | 0.004 |
| Cable | — | — | 0.5 | 293 | — | — | 0.003 |
| Avantek amplifier | AN-2200N | — | — | — | 25 | 870 | 0.071 |
| 20-dB coupler | — | — | 20 | 293 | — | — | 0.007 |
| Cable | — | — | 1 | 293 | — | — | 0.002 |
| Downconverter $T_{in}$ | — | — | — | — | — | 8881 | 0.287 |
| Total antenna system noise temperature (referred to input of maser) | | | | | | | 15.74 |
| Follow-up noise contribution $T_f$ | | | | | | | 0.382 |

[a] SPD feedcone, low-noise path, 2295 MHz, 90-deg elevation angle, and clear weather.
[b] Closed-cycle refrigerator.

the ambient load while switching the maser pump source on and off. The Y-factor, $Y$, is then used to determine the FNT:

$$FNT = \frac{T_{load}}{10^{Y/10}} \tag{2}$$

where $T_{load}$ is the ambient load temperature, K, and the difference in power level between the maser pump on and off is $Y$, dB.

## III. Preliminary Investigation and Baseline Data

A noise budget was prepared for the DSS-14, DSS-43, and DSS-63 SPD systems in the listen-only mode. These noise budgets used our best estimates of microwave performance for each component in the system. Some measured data were available; other figures are theoretical. Measurements made at the stations were compared with these noise budget predictions. While the DSS-14 SPD system noise temperature agreed closely with its noise budget, those at DSS 43 and DSS 63 did not agree with predicted performance. Further FNT measurements isolated the problem at DSS 43 and DSS 63 to that part of the SPD system following the maser. This was determined after comparing the over-1-K FNT measured at both stations to the 0.4-K predicted noise. Since DSS 14 was the only station that closely agreed with predictions, and since it was the most readily available SPD system, it was carefully evaluated and used as a baseline against which to compare the other two stations.

A $T_{op}$ measurement made at DSS 14 using the PAS yielded the data shown in row 1 of Table 3. A similar $T_{op}$ measurement was made at the immediate output of the maser using the JPL total power radiometer (TPR); this yielded the data in row 2. The $T_{op}$ measured at the input to the multiport coupler assembly (MPCA) is shown in row 3. An FNT measurement made using the PAS gave the data in row 4, while an FNT measurement made using the JPL TPR is shown in row 5. Next, the Block IV receivers' noise performances were checked using a Hewlett Packard (HP) 8970B noise figure meter. The resulting noise figure (NF) and gain information obtained at 2295 MHz is displayed in row 6 for receiver 1 and row 7 for receiver 2.

The high noise figures and the difference in noise figures, 21.4 dB for receiver 1 and 17.3 dB for receiver 2, of the Block IV receivers were noted. This poor performance results in the $T_{op}$ being more than 0.5-K above what good engineering practice should provide. A further explanation of the problem and a proposed solution appear in the recommendations section.

**Table 3. Measurement data for DSS 14.**

| Data no., type | $Y$, dB | Load, deg C | $T_{rcvr}$, K | $T_{op}$, K | FNT, K | NF, dB | Gain, dB |
|---|---|---|---|---|---|---|---|
| 1, PAS | 12.88 | 18 | 2.5 | 15.1 | — | — | — |
| 2, TPR | 13.05 | 18 | 2.5 | 14.5 | — | — | — |
| 3, MPCA | 13.05 | 18 | 2.5 | 14.5 | — | — | — |
| 4, PAS | 25 | 18 | — | — | 0.92 | — | — |
| 5, TPR | 35 | 18 | — | — | 0.04 | — | — |
| 6, Receiver 1 | — | — | — | — | — | 21.4 | 33 |
| 7, Receiver 2 | — | — | — | — | — | 17.3 | 33 |

## IV. DSS-43 Measurements

Measurements made at DSS 43 at the start of this investigation gave system noise temperature values of 17.2 K. Measurements made later on in the course of the investigation by station personnel at DSS 43 using the 50-MHz PAS resulted in the data in rows 1 and 3 of Table 4. These data were reduced by station personnel and, therefore, the raw data are unavailable. Measurements made at DSS 43 using the JPL TPR yielded the data in rows 2 and 4.

Station personnel explained the difference in measurements before and after evaluation activities as follows: At DSS 43, there never was a problem with the actual system noise temperature, and the precision power monitor method of measurement was reporting the correct result. The station chose to publish a determination of the system noise temperature based on a Y-factor detector result. This result was in error because of a nonstandard configuration of the Y-factor detector assembly. The station corrected this and confirmed that results from the three different methods[1] of measuring system noise temperature at the station agreed within 0.5 K.

**Table 4. Measurement data for DSS 43.**

| Data no., type | $Y$, dB | Load, deg C | $T_{rcvr}$, K | $T_{op}$, K | FNT, K | NF, dB | Gain, dB |
|---|---|---|---|---|---|---|---|
| 1, PAS | — | — | — | 14.7 | — | — | — |
| 2, TPR | 13.06 | 18.8 | 2.5 | 14.7 | — | — | — |
| 3, PAS | — | — | — | — | 0.4 | N/A | N/A |
| 4, TPR | 39.23 | 19 | — | — | 0.039 | N/A | N/A |

## V. DSS-63 Measurements

The system operating noise temperature, $T_{op}$, of the SPD system at DSS 63 was measured using three methods: Row 1 of Table 5 shows the measurement result using the 50-MHz PAS; row 2 shows the results using the JPL TPR at the output of the maser LNA; and row 3 shows the results using the JPL TPR at the input to the multiport coupler assembly. An FNT measurement made using the PAS gave the data in row 4, and an FNT measurement made using the JPL TPR is shown in row 5. The input noise figure of the Block IV receivers was measured using an HP 8970B noise figure meter, and the resulting noise figure and gain information obtained at 2295 MHz is displayed in row 6 for receiver 1 and row 7 for receiver 2. The VLBI downconverter input noise figure and gain for the RCP and LCP channels were measured using an automated test setup developed at JPL. These data are shown in Fig. 2, which compares the noise figures of the two channels, and Fig. 3, which compares the gains.

Included in earlier measurements were data taken of waveguide (WG) components on the ground using the S-band test horn, the TPR, and a Block IV S-band maser that was brought from JPL. These tests were inconclusive due to the interaction of the WG components; that is, the WG components must be tested as a system and not independently in order to obtain accurate results.

Plotting system noise temperature and ambient load temperature as a function of time (Fig. 4) revealed that the highest noise temperatures were occurring at the warmest load temperatures (i.e., the warmest time of the day). This was determined to be due to inadequate air conditioning in the cone, which was causing elevated physical temperatures of the WG components. The air conditioning was improved by station personnel to bring the SPD cone physical temperature down so that it more closely matched DSS 14 and DSS 43; this improved the stability of the system noise temperature over time.

[1] The third method uses the open-loop VLBI/radio science downconverter output and a quality spectrum analyzer for the Y-factor measurement.

**Table 5. Measurement data for DSS 63.**

| Data no., type | $Y$, dB | Load, deg C | $T_{rcvr}$, K | $T_{op}$, K | FNT, K | NF, dB | Gain, dB |
|---|---|---|---|---|---|---|---|
| 1, PAS | 12.54 | 24 | 4.5 | 16.8 | — | — | — |
| 2, TPR | 12.8 | 27.5 | 4.5 | 16.0 | — | — | — |
| 3, MPCA | 12.5 | 27 | 4.5 | 17.1 | — | — | — |
| 4, PAS | 23.8 | 25.5 | — | — | 1.25 | — | — |
| 5, TPR | 39 | 27 | — | — | 0.04 | — | — |
| 6, Receiver 1 | — | — | — | — | — | 15.7 | 27 |
| 7, Receiver 2 | — | — | — | — | — | 15.9 | 30 |

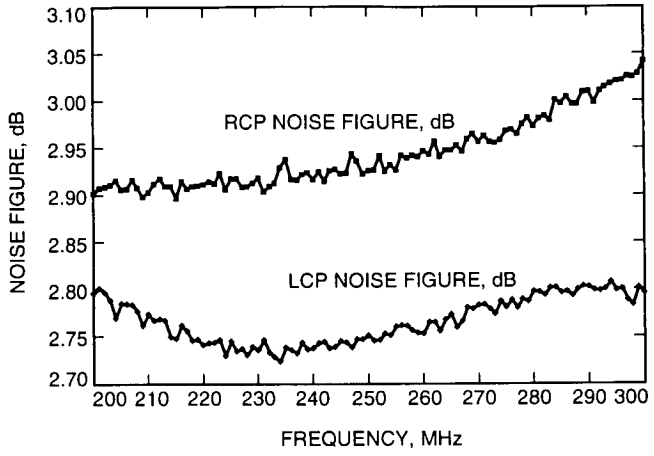


**Fig. 2. Comparison of the noise figures for the RCP and LCP channels of the VLBI downconverter at DSS 63.**

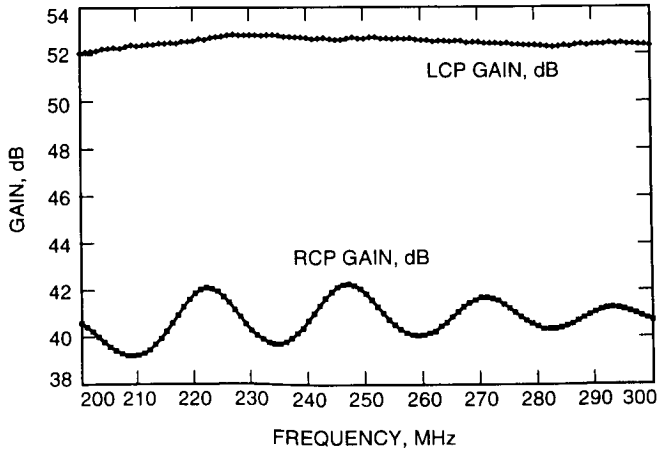


**Fig. 3. Comparison of the gain for the RCP and LCP channels of the VLBI downconverter at DSS 63.**

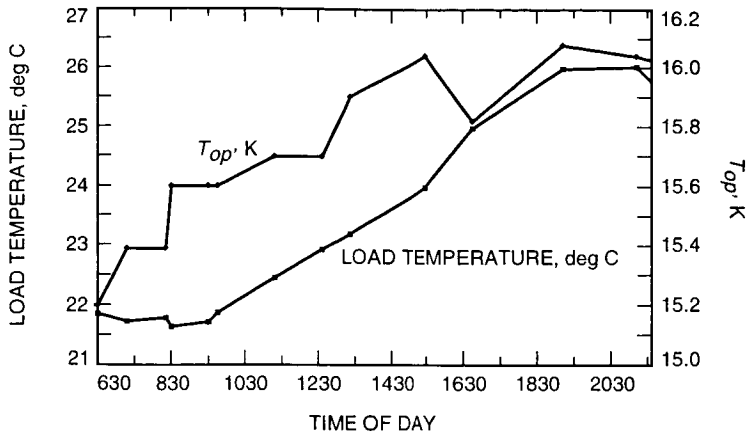


**Fig. 4. Comparison of $T_{op}$ and load temperature versus the time of day.**

## VI. Conclusions

The result of the evaluation was an overall improvement of the S-band polarization diverse systems at all stations except DSS 14. DSS 14 showed a measurable increase in system noise temperature performance; this is most likely due to an elevated noise temperature in the S-band Block IV receivers at the station. Pre-evaluation results showed a $T_{op}$ of 14.7 K, 17.2 K, and 17.6 K at DSS 14, DSS 43, and DSS 63, respectively. Postevaluation results showed a $T_{op}$ of 15.15 K, 14.7 K, and 15.5 K at DSS 14, DSS 43, and DSS 63, respectively.

This task was successful in achieving its goals. The data taken and the equipment and procedures developed will assist in future investigations of station system noise temperatures.

## VII. Recommendations

It is recommended that tolerances be established for the 70-m SPD system $T_{op}$ and FNT contributions at the stations. When an out-of-tolerance $T_{op}$ is measured, the FNT and linearity should be checked, and troubleshooting should proceed to identify the problem. This could and should include use of the HP 8970B noise figure meter with an HP 346-type noise source for the purpose of noise figure and gain measurements of station equipment behind the LNA.

The Stelzried spreadsheet for checking Y-factor linearity should be implemented at all DSN stations. The stations should also have the capability of measuring the LNA noise temperature on the ground and at the output of the LNA in the cone, using the same system the JPL Microwave Electronics Group uses—a calibrated horn (for ground tests), absorber load, and the JPL total power radiometer.

The attenuators and strip chart recorders should be replaced with a precision power meter or spectrum analyzer capable of measuring Y-factor power ratios at 50 MHz with an accuracy of ±0.01 dB.

The $T_{op}$ of the 70-m SPD systems can be reduced an additional 0.3 K by reducing the follow-up noise temperature contribution by a factor of 10, from 0.4 to <0.04 K. This can be achieved by replacing the existing S-band follow-up amplifier with a state-of-the-art amplifier, installing this postamplifier in front of any losses between the LNA and the downconverter, and replacing the downconverters at the stations with state-of-the-art downconverters having noise figures of 5 dB or less.

The RCP channel of the VLBI downconverter at DSS 63 should be investigated. It has 14-dB less gain than the LCP channel. The fact that the RCP channel has a very similar noise figure when compared with the LCP channel indicates that this problem is in the output of the downconverter.

Finally, the two S-band Block IV downconverters at DSS 14 exhibit elevated noise temperatures. This is most likely due to an elevated loss in either the preselector filter or the mixer and should be investigated and corrected.

## Acknowledgments

The authors would like to thank the DSN stations and their knowledgeable personnel, who are extremely familiar with the complete end-to-end systems they operate, for their assistance during the course of this work: Larry Bracamonte at DSS 14; John Murray and Allen Robinson at DSS 43; and Jose Luis Galvez, Jose Requilme, Juan Acena, and Candido Illescas at DSS-63. We would also like to thank those people at JPL who assisted with their knowledge about specific subsystems; these include Dennis Rowan, Mike Barnard, and Sam Petty. Finally, we wish to thank Charles Stelzried for use of the linearity spreadsheet.