

# The Telecommunications and Data Acquisition Progress Report 42-98

April-June 1989

E. C. Posner  
Editor

August 15, 1989

**NASA**  
National Aeronautics and  
Space Administration  
  
Jet Propulsion Laboratory  
California Institute of Technology  
Pasadena, California

(NASA-CR-185093) THE TELECOMMUNICATIONS AND  
DATA ACQUISITION REPORT Progress Report,  
Apr. - Jun. 1989 (JPL) 234 p CSDL 17B

N90-12786  
--THRU--  
N90-12804  
Unclas  
0239810

G3/32

547626  
P236

# The Telecommunications and Data Acquisition Progress Report 42-98

April-June 1989

E. C. Posner  
Editor

August 15, 1989

**NASA**

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.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

# Preface

This quarterly publication provides archival reports on developments in programs managed by JPL's Office of Telecommunications and Data Acquisition (TDA). In space communications, radio navigation, radio science, and ground-based radio and radar astronomy, it reports on activities of the Deep Space Network (DSN) and its associated Ground Communications Facility (GCF) in planning, in supporting research and technology, in implementation, and in operations. Also included is TDA-funded activity at JPL on data and information systems and reimbursable DSN work performed for other space agencies through NASA. The preceding work is all performed for NASA's Office of Space Operations (OSO). The TDA Office also performs work funded by two other NASA program offices through and with the cooperation of the Office of Space Operations. These are the Orbital Debris Radar Program (with the Office of Space Station) and 21st Century Communication Studies (with the Office of Exploration).

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

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

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

## OSO Tasks:

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

## OSO Cooperative Tasks:

- Orbital Debris Radar Program
- 21st Century Communication Studies



**OSSA Tasks:**

**Search for Extraterrestrial Intelligence**

**Goldstone Solar System Radar**

**Radio Astronomy**

**Discretionary Funded Tasks**

# OSSA TASKS

## Search for Extraterrestrial Intelligence

**The NASA SETI Sky Survey: Recent Developments** ..... 218<sup>5/8</sup>  
M. J. Klein, S. Gulkis, E. T. Olsen, N. A. Renzetti  
NASA Code 199-50-62-07-02

**Referees** ..... 227

<b>Performance of the Image Statistics Decoder in Conjunction With the Goldstone-VLA Array</b> .....	138
H. C. Wang and G. H. Pitt III	<i>510</i>
NASA Code 310-30-71-83-04	

## DSN Systems Implementation CAPABILITIES FOR EXISTING PROJECTS

<b>A Closed-Cycle Refrigerator for Cooling Maser Amplifiers Below 4 Kelvin</b> .....	141
M. Britcliffe	<i>511</i>
NASA Code 314-30-42-01-34	

### NEW INITIATIVES

<b>Wideband Phase-Locked Angular Modulator</b> .....	150
L. Nguyen	<i>512</i>
NASA Code 314-30-56-57-33	

<b>Performance of the Split-Symbol Moments SNR Estimator in the Presence of Inter-Symbol Interference</b> .....	157
B. Shah and S. Hinedi	<i>513</i>
NASA Code 314-30-56-57-33	

### NETWORK UPGRADE AND SUSTAINING

<b>DSN 70-Meter Antenna Microwave Optics Design and Performance Improvements Part II: Comparison With Measurements</b> .....	174
D. A. Bathker and S. D. Slobin	<i>514</i>
NASA Code 314-30-56-44-01	

<b>Stable Kalman Filters for Processing Clock Measurement Data</b> .....	190
P. A. Clements, B. P. Gibbs, and J. S. Vandergraft	<i>515</i>
NASA Code 314-30-56-57-17	

## DSN Operations TDA PROGRAM MANAGEMENT AND ANALYSIS

<b>A Simple Model for DSS-14 Outage Times</b> .....	202
H. C. Rumsey, R. Stevens, and E. C. Posner	<i>516</i>
NASA Code 055-40-01-00-96	

## Flight-Ground Advanced Engineering

<b>Ka-Band MMIC Beam Steered Transmitter Array</b> .....	207
D. L. Rascoe, A. L. Riley, J. Huang, V. Lubecke, and L. Duffy	<i>517</i>
NASA Code 315-91-50-02-02	

# Contents

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

<b>An Analytic Development of Orbit Determination for a Distant, Planetary Orbiter</b> .....	1	51
R. K. Russell and S. W. Thurman NASA Code 310-10-63-50-00		
<b>A Statistical Study of Radio-Source Structure Effects on Astrometric Very Long Baseline Interferometry Observations</b> .....	26	52
J. S. Ulvestad NASA Code 310-10-60-60-00		
<b>An Apparatus for the Electrodynamic Containment of Charged Macroparticles</b> .....	57	53
A. Williams, R. Melbourne, L. Maleki, G. Janik, and J. Prestage NASA Code 310-10-62-15-00		

## COMMUNICATIONS, SPACECRAFT-GROUND

<b>Spin-Lattice Relaxation and the Calculation of Gain, Pump Power, and Noise Temperature in Ruby</b> .....	63	54
J. R. Lyons NASA Code 310-20-66-87-02		
<b>Performance Effects of Tie-Truss Modifications for a 70-Meter Centerline Beam Waveguide Antenna</b> .....	86	55
J. J. Cucchissi NASA Code 310-20-65-86-04		
<b>A Performance Comparison Between Block Interleaved and Helically Interleaved Concatenated Coding Systems</b> .....	95	56
K.-M. Cheung and L. Swanson NASA Code 310-30-71-83-02		
<b>On the Decoder Error Probability of Linear Codes</b> .....	104	57
K.-M. Cheung NASA Code 310-30-71-83-02		

## STATION CONTROL AND SYSTEM TECHNOLOGY

<b>Automated Monitor and Control for Deep Space Network Subsystems</b> .....	110	58
P. Smyth NASA Code 310-30-70-89-02		
<b>Frame Synchronization Methods Based on Channel Symbol Measurements</b> .....	121	59
S. Dolinar and K.-M. Cheung NASA Code 310-30-71-83-02		

5471035 51-13

N90-12787

250

# An Analytic Development of Orbit Determination for a Distant, Planetary Orbiter

R. K. Russell and S. W. Thurman

Navigation Systems Section

*With the advent of the Mariner '71 Mission, NASA has been sending spacecraft to orbit various distant bodies within the solar system. At present, there is still no adequate theory describing the inherent state estimation accuracy, based on two-way, coherent range-rate data. It is the purpose of this article to lay the groundwork for a general elliptic theory, and in addition to provide an analytic solution for the special case of circular orbits.*

## I. Introduction

When one begins to analyze a possible future mission employing an orbiting spacecraft, it is desirable to know the characteristics of the position and velocity errors of that spacecraft, based on a single revolution of range-rate data. To know these errors as a function of orbit shape, size, and orientation with respect to the Earth is the central issue. If range were the data type of interest, the problem would be vastly simpler, since velocities would not be involved. Unfortunately for the analyst, however, range-rate data is central and must be treated. This particular problem has proved to be a most formidable analytic endeavor.

shifts, astronomers discovered that they could determine all the elements of the motion except one. The one element, which turned out to be completely indeterminate, is the longitude of the ascending node relative to the plane-of-the-sky, with the plane-of-the-sky being that plane normal to the line-of-sight from the observer to the binary system.

The reason for this indeterminacy is that the binary star system is effectively "at infinity," meaning that no significant parallax motion exists between the observer and the observed.

In this problem, the body about which the spacecraft is in orbit is not "at infinity," but is relatively near. This implies that there will be some parallax motion between the observer and the observed, and hence we should be able to obtain a "reasonable" solution for the longitude of the ascending node relative to the plane-of-the-sky. This parameter will typically be the principal contributor to the spacecraft state error. As such, classical elements relative to the plane-of-the-sky appear to be a sensible "eigen" system in which to solve the problem.

## II. Development of the Observable

One of the most important aspects of this orbital estimation problem is to pose it properly. In this regard, the "binary star problem" of classical astronomy can provide insight. In attempting to obtain the orbits of a distant binary star system through the use of spectroscopic Doppler

The geometry of a spacecraft in orbit about a distant body is depicted in Fig. 1. Let:

$$\mathbf{r}_p = \begin{pmatrix} x_p \\ y_p \\ z_p \end{pmatrix} \quad (1)$$

$$\dot{\mathbf{r}}_p = \begin{pmatrix} \dot{x}_p \\ \dot{y}_p \\ \dot{z}_p \end{pmatrix} \quad (2)$$

where  $x_p$ ,  $y_p$ , and  $z_p$  are the plane-of-the-sky Cartesian position coordinates of the spacecraft and  $\dot{x}_p$ ,  $\dot{y}_p$ , and  $\dot{z}_p$  are the plane-of-the-sky Cartesian velocity coordinates of the spacecraft.

The orbital coordinates are then defined as:

$$\mathbf{r}_o = \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} \quad (3)$$

$$\dot{\mathbf{r}}_o = \begin{pmatrix} \dot{\xi} \\ \dot{\eta} \\ \dot{\zeta} \end{pmatrix} \quad (4)$$

with

$$\xi = a(\cos E - e) \quad (5)$$

$$\eta = a\beta \sin E \quad (6)$$

$$\zeta = 0 \quad (7)$$

$$\dot{\xi} = - \left( \frac{an \sin E}{1 - e \cos E} \right) \quad (8)$$

$$\dot{\eta} = \left( \frac{an\beta \cos E}{1 - e \cos E} \right) \quad (9)$$

$$\dot{\zeta} = 0 \quad (10)$$

$$\beta = \sqrt{1 - e^2} \quad (11)$$

where  $E$  is the *eccentric anomaly*,  $a$  is the orbit *semi-major axis* (km),  $e$  is the orbit *eccentricity* (dimensionless), and  $n$  is the orbit *mean motion* ( $s^{-1}$ ).

The orbital coordinates can be related to the Cartesian, plane-of-the-sky coordinates as follows:

$$\mathbf{r}_p = R^T(\Omega, \omega, i)\mathbf{r}_o \quad (12)$$

where  $R^T(\Omega, \omega, i)$  is the transpose of the Euler rotation matrix, defined as

$$R(\Omega, \omega, i) = \begin{pmatrix} \cos \Omega \cos \omega - \sin \Omega \cos i \sin \omega & \sin \Omega \cos \omega + \cos \Omega \cos i \sin \omega & \sin i \sin \omega \\ -\cos \Omega \sin \omega - \sin \Omega \cos i \cos \omega & -\sin \Omega \sin \omega + \cos \Omega \cos i \cos \omega & \sin i \cos \omega \\ \sin \Omega \sin i & -\cos \Omega \sin i & \cos i \end{pmatrix} \quad (13)$$

and  $\Omega$  is the *longitude of ascending node* relative to the plane-of-the-sky,  $\omega$  is the *argument of periapsis* relative to the plane-of-the-sky, and  $i$  is the *inclination* relative to the plane-of-the-sky.

To have any hope of analytically solving this problem, all possible simplifications must be made without, of course, loss of generality. One such simplification will be to define the  $x$ -axis of the plane-of-the-sky system to be the line-of-nodes. This implies that

$$\Omega = 0 \quad (14)$$

Thus  $R^T(\Omega, \omega, i)$  becomes

$$R^T(\omega, i) = \begin{pmatrix} \cos \omega & -\sin \omega & 0 \\ \sin \omega \cos i & \cos \omega \cos i & -\sin i \\ \sin \omega \sin i & \cos \omega \sin i & \cos i \end{pmatrix} \quad (15)$$

If it is now assumed that the orbital elements are constant in time,

$$\dot{\mathbf{r}}_p = R^T(\omega, i) \dot{\mathbf{r}}_o \quad (16)$$

From Fig. 1, it is seen that

$$\mathbf{r}_d = \mathbf{r}_p - \mathbf{r}_T \quad (17)$$

$$\mathbf{r}_T = \begin{pmatrix} V_N t \cos \theta \\ V_N t \sin \theta \\ V_R t + \rho_{ET} \end{pmatrix} \quad (18)$$

where:

$V_N$  is the component of the velocity of the target central body relative to the Earth that lies in the plane-of-the-sky; this shall hereinafter be referred to as the "cross-velocity"

$V_R$  is the component of the velocity of the target central body relative to the Earth that lies along the line-of-sight

$\rho_{ET}$  is the distance along the line-of-sight between the center of the Earth and the center of the target central body at the initial epoch of estimation, i.e., at  $t = 0$

$\theta$  is the angle between the cross-velocity and the line-of-nodes

One can now form the range-rate observable,  $\dot{\rho}$ , as

$$\dot{\rho} = \left( \frac{\mathbf{r}_d \cdot \dot{\mathbf{r}}_d}{\rho} \right) \quad (19)$$

where

$$\rho = |\mathbf{r}_d| \quad (20)$$

Therefore,

$$\rho \dot{\rho} = \mathbf{r}_p \cdot \dot{\mathbf{r}}_p - \mathbf{r}_p \cdot \dot{\mathbf{r}}_T - \mathbf{r}_T \cdot \dot{\mathbf{r}}_p + \mathbf{r}_T \cdot \dot{\mathbf{r}}_T \quad (21)$$

where

$$\dot{\mathbf{r}}_T = \begin{pmatrix} V_N \cos \theta \\ V_N \sin \theta \\ V_R \end{pmatrix} \quad (22)$$

It can then be shown that

$$\mathbf{r}_T \cdot \dot{\mathbf{r}}_T = V_N^2 t + V_R^2 t + V_R \rho_{ET} \quad (23)$$

$$\mathbf{r}_p \cdot \dot{\mathbf{r}}_p = a^2 n e \sin E \quad (24)$$

$$\mathbf{r}_p \cdot \dot{\mathbf{r}}_T = V_N (\xi \alpha_1 + \eta \alpha_2) + V_R (\xi \sin \omega + \eta \cos \omega) \sin i \quad (25)$$

$$\mathbf{r}_T \cdot \dot{\mathbf{r}}_p = V_N t (\dot{\xi} \alpha_1 + \dot{\eta} \alpha_2) + (V_R t + \rho_{ET}) (\dot{\xi} \sin \omega + \dot{\eta} \cos \omega) \sin i \quad (26)$$

where

$$\alpha_1 = \cos \theta \cos \omega + \sin \theta \sin \omega \cos i \quad (27)$$

$$\alpha_2 = -\cos \theta \sin \omega + \sin \theta \cos \omega \cos i \quad (28)$$

Next one has

$$\rho = \left( \mathbf{r}_d \cdot \mathbf{r}_d \right)^{\frac{1}{2}} = \left[ \mathbf{r}_p \cdot \mathbf{r}_p + \mathbf{r}_T \cdot \mathbf{r}_T - 2 \mathbf{r}_p \cdot \mathbf{r}_T \right]^{\frac{1}{2}} \quad (29)$$

This can be shown to be

$$\rho = \left[ a^2 (1 - e \cos E)^2 + \rho_{ET}^2 + 2 \rho_{ET} V_R t + (V_N^2 + V_R^2) t^2 - 2 V_N t (\xi \alpha_1 + \eta \alpha_2) - 2 (V_R t + \rho_{ET}) (\xi \sin \omega + \eta \cos \omega) \sin i \right]^{\frac{1}{2}} \quad (30)$$



For large  $\rho_{ET}$  this implies

$$\frac{1}{\rho} = \frac{1}{\rho_{ET}} \left[ 1 + \frac{2V_R t}{\rho_{ET}} - \frac{2 \sin i}{\rho_{ET}} \{ \xi \sin \omega + \eta \cos \omega \} + O\left(\frac{1}{\rho_{ET}^2}\right) \right]^{-\frac{1}{2}} \quad (31)$$

This then becomes

$$\frac{1}{\rho} \approx \frac{1}{\rho_{ET}} \left[ 1 - \frac{V_R t}{\rho_{ET}} + \frac{\sin i}{\rho_{ET}} \{ \xi \sin \omega + \eta \cos \omega \} \right] \quad (32)$$

Combining Eqs. (21), (23)–(26), and (32) yields

$$\begin{aligned} \dot{\rho} \approx & \frac{1}{\rho_{ET}} \left[ \rho_{ET} V_R - \rho_{ET} \{ \dot{\xi} \sin \omega + \dot{\eta} \cos \omega \} \sin i + V_N^2 t + a^2 n e \sin E \right. \\ & \left. - V_N \{ (\xi + \dot{\xi} t) \alpha_1 + (\eta + \dot{\eta} t) \alpha_2 \} - \{ \xi \sin \omega + \eta \cos \omega \} \{ \dot{\xi} \sin \omega + \dot{\eta} \cos \omega \} \sin^2 i \right] \end{aligned} \quad (33)$$

From the above, one can observe that

$$\{ \xi \sin \omega + \eta \cos \omega \} = a \{ (\cos E - e) \sin \omega + \beta \sin E \cos \omega \} \quad (34)$$

$$\{ \dot{\xi} \sin \omega + \dot{\eta} \cos \omega \} = \left( \frac{an}{1 - e \cos E} \right) \{ -\sin E \sin \omega + \beta \cos E \cos \omega \} \quad (35)$$

Combining Eqs. (33)–(35) produces

$$\begin{aligned} \dot{\rho} \approx & \left[ V_R - \left( \frac{an \sin i}{1 - e \cos E} \right) \{ -\sin E \sin \omega + \beta \cos E \cos \omega \} \right. \\ & + \frac{1}{\rho_{ET}} \left\{ V_N^2 t - \left( \frac{a^2 n \sin^2 i}{1 - e \cos E} \right) \{ \sin E \sin \omega - \beta \cos E \cos \omega \} \{ (\cos E - e) \sin \omega + \beta \sin E \cos \omega \} \right. \\ & \left. \left. + a^2 n e \sin E - V_N \{ (\xi + \dot{\xi} t) \alpha_1 + (\eta + \dot{\eta} t) \alpha_2 \} \right\} \right] \end{aligned} \quad (36)$$

Now one proceeds to expand the last term of Eq. (36). To do this, Kepler's Equation is needed:

$$E - e \sin E = n(t - T_P) \quad (37)$$

where  $T_P$  is the *time of periapsis passage*.

Without loss of generality, one may assume that  $T_P = 0$ . Therefore

$$nt = E - e \sin E \quad (38)$$

Combining Eqs. (5), (6), (8), (9), and (38) produces

$$\begin{aligned} & \{(\xi + \dot{\xi}t)\alpha_1 + (\eta + \dot{\eta}t)\alpha_2\} \\ &= a \left\{ \alpha_1 \left( \cos E - e - \frac{(E - e \sin E) \sin E}{1 - e \cos E} \right) + \alpha_2 \beta \left( \sin E + \frac{(E - e \sin E) \cos E}{1 - e \cos E} \right) \right\} \end{aligned} \quad (39)$$

The range-rate approximation then becomes

$$\begin{aligned} \dot{\rho} \approx & \left[ V_R - \left( \frac{an \sin i}{1 - e \cos E} \right) \left\{ -\sin E \sin \omega + \beta \cos E \cos \omega \right\} + \frac{1}{\rho_{ET}} \left\{ V_N^2 t + a^2 n e \sin E \right. \right. \\ & - \left. \left( \frac{a^2 n \sin^2 i}{1 - e \cos E} \right) \left\{ \sin E \sin \omega - \beta \cos E \cos \omega \right\} \left\{ (\cos E - e) \sin \omega + \beta \sin E \cos \omega \right\} \right. \\ & \left. \left. - V_N a \left\{ \alpha_1 \left( \cos E - e - \frac{(E - e \sin E) \sin E}{1 - e \cos E} \right) + \alpha_2 \beta \left( \sin E + \frac{(E - e \sin E) \cos E}{1 - e \cos E} \right) \right\} \right] \end{aligned} \quad (40)$$

If one wished to compute an approximation to the actual received range-rate signal, ignoring the motion of the receiving station relative to the center of the Earth, Eq. (40) would be the expression to use. In this case, however, the range-rate approximation shall be used for the development of partial derivatives only, and the only partial derivatives of interest are with respect to the spacecraft state (classical orbital elements with respect to the plane-of-the-sky). Hence, the elements of interest are  $a$ ,  $e$ ,  $T_P$ ,  $\omega$ ,  $i$ , and the final element is  $\theta$ . Note that  $\theta$  by our definition is equivalent to the *longitude of the ascending node*,  $\Omega$ .

Thus those terms in the range-rate approximation which do not contribute to the appropriate partial derivatives will hereinafter be ignored. Additionally, those terms which contribute negligible partial derivatives will also be ignored. As a result, the range-rate approximation which shall be used for the computation of the partial derivatives is

$$\begin{aligned} \dot{\rho} \approx & \left[ \left( \frac{an \sin i}{1 - e \cos E} \right) \left\{ \sin E \sin \omega - \beta \cos E \cos \omega \right\} \right. \\ & \left. - \left( \frac{V_N a}{\rho_{ET}} \right) \left\{ \alpha_1 \left( \cos E - e - \frac{(E - e \sin E) \sin E}{1 - e \cos E} \right) + \alpha_2 \beta \left( \sin E + \frac{(E - e \sin E) \cos E}{1 - e \cos E} \right) \right\} \right] \end{aligned} \quad (41)$$

The partial derivatives with respect to  $a$ ,  $e$ ,  $T_P$ ,  $\omega$ , and  $i$  shall be computed from the first term in the range-rate approximation (since these dominate partials for the same parameters taken from the second term), whereas the partial derivative with respect to  $\theta$  will be computed from the second term (this is the only term with  $\theta$  variation).

### III. Obtaining the Partial Derivatives

Rewriting Eq. (41) produces

$$\dot{\rho} \approx \left[ \left( \frac{a^{-\frac{1}{2}} \mu^{\frac{1}{2}} \sin i}{1 - e \cos E} \right) \left\{ \sin E \sin \omega - \beta \cos E \cos \omega \right\} - \left( \frac{V_{Na}}{\rho_{ET}} \right) \left\{ \alpha_1 \left( \cos E - e - \frac{(E - e \sin E) \sin E}{1 - e \cos E} \right) + \alpha_2 \beta \left( \sin E + \frac{(E - e \sin E) \cos E}{1 - e \cos E} \right) \right\} \right] \quad (42)$$

To produce the partial of the observables with respect to the classical state of the spacecraft, the partials of the eccentric anomaly,  $E$ , with respect to the classical state are needed. They are

$$\left( \frac{\partial E}{\partial a} \right) = -\frac{3(E - e \sin E)}{2a(1 - e \cos E)} \quad (43)$$

$$\left( \frac{\partial E}{\partial e} \right) = \frac{\sin E}{(1 - e \cos E)} \quad (44)$$

$$\left( \frac{\partial E}{\partial T_P} \right) = -\frac{n}{(1 - e \cos E)} \quad (45)$$

$$\left( \frac{\partial E}{\partial \omega} \right) = 0 \quad (46)$$

$$\left( \frac{\partial E}{\partial i} \right) = 0 \quad (47)$$

$$\left( \frac{\partial E}{\partial \theta} \right) = 0 \quad (48)$$

Taking partial derivatives of the observable produces

$$\left( \frac{\partial \dot{\rho}}{\partial a} \right) = \left( \frac{n \sin i}{2} \right) \left[ \frac{\{-\sin E \sin \omega + \beta \cos E \cos \omega\}}{(1 - e \cos E)} - \frac{3(E - e \sin E) \{(\cos E - e) \sin \omega + \beta \sin E \cos \omega\}}{(1 - e \cos E)^3} \right] \quad (49)$$

$$\left( \frac{\partial \dot{\rho}}{\partial e} \right) = an \sin i \left[ \frac{-\cos E \{(\cos E - e) \cos \omega - \beta \sin E \sin \omega\}}{\beta(1 - e \cos E)^2} + \frac{\sin E \{(\cos E - e) \sin \omega + \beta \sin E \cos \omega\}}{(1 - e \cos E)^3} \right] \quad (50)$$

$$\left(\frac{\partial \dot{\rho}}{\partial T_P}\right) = -an^2 \sin i \left[ \frac{(\cos E - e) \sin \omega + \beta \sin E \cos \omega}{(1 - e \cos E)^3} \right] \quad (51)$$

$$\left(\frac{\partial \dot{\rho}}{\partial \omega}\right) = an \sin i \left[ \frac{\sin E \cos \omega + \beta \cos E \sin \omega}{(1 - e \cos E)} \right] \quad (52)$$

$$\left(\frac{\partial \dot{\rho}}{\partial i}\right) = an \cos i \left[ \frac{\sin E \sin \omega - \beta \cos E \cos \omega}{(1 - e \cos E)} \right] \quad (53)$$

$$\left(\frac{\partial \dot{\rho}}{\partial \theta}\right) = -\left(\frac{V_N a}{\rho_{ET}}\right) \left[ \alpha_3 \left\{ \cos E - e - \frac{(E - e \sin E) \sin E}{1 - e \cos E} \right\} + \alpha_4 \beta \left\{ \sin E + \frac{(E - e \sin E) \cos E}{1 - e \cos E} \right\} \right] \quad (54)$$

where

$$\alpha_3 = -\sin \theta \cos \omega + \cos \theta \sin \omega \cos i \quad (55)$$

$$\alpha_4 = \sin \theta \sin \omega + \cos \theta \cos \omega \cos i \quad (56)$$

Again, if one is to get through this problem, one must simplify as much as possible or risk getting lost in the algebra. A careful examination of the partials detailed in Eqs. (49)–(54) reveals that the first term of Eq. (49), other than a constant of proportionality, is the same as Eq. (53). Therefore it may be eliminated from Eq. (49) by a change of variable. Let

$$a' = a + \left(\frac{2a}{\tan i}\right) i = a + (2a \cot i) i \quad (57)$$

which implies that

$$\left(\frac{\partial \dot{\rho}}{\partial a'}\right) = \left(\frac{\partial \dot{\rho}}{\partial a}\right) \left(\frac{\partial a}{\partial a'}\right) + \left(\frac{\partial \dot{\rho}}{\partial i}\right) \left(\frac{\partial i}{\partial a'}\right) \quad (58)$$

From Eq. (57), one has

$$\left(\frac{\partial a}{\partial a'}\right) = 1 \quad (59)$$

$$\left(\frac{\partial i}{\partial a'}\right) = \left(\frac{1}{2a \cot i}\right) = \left(\frac{\tan i}{2a}\right) \quad (60)$$

Therefore,

$$\left(\frac{\partial \dot{\rho}}{\partial a'}\right) = \left(\frac{-3n \sin i}{2}\right) \left[ \frac{(E - e \sin E) \{(\cos E - e) \sin \omega + \beta \sin E \cos \omega\}}{(1 - e \cos E)^3} \right] \quad (61)$$

To obtain the correct partials in Eq. (58), one must carefully discern which of the variables in Eq. (57) are differentiable parameters and which are not; it is then clear that Eq. (61) represents a considerable simplification of the problem.

## IV. Accumulating the Information Array

The normal, Gaussian way by which to obtain the covariance of errors in a set of estimated parameters,  $\Gamma$ , is stated as follows:

$$\Gamma = [\tilde{\Gamma}^{-1} + A^T W A]^{-1} \quad (62)$$

where  $\tilde{\Gamma}$  is the a priori error in the covariance,  $A$  is the vector partial derivative of the observable (at the time of observation) with respect to the parameters being estimated, and  $W$  is the weighting matrix. If it is assumed that the error in the measurement of the observable at time  $t_i$  is uncorrelated with the error in the measurement of the observable at time  $t_j$ , for all  $i$  and  $j$ , then  $W$  becomes a diagonal matrix of dimension equal to the number of observables. From minimum-variance concepts, the diagonal elements of the weighting matrix may be shown to be just the inverse of the variance of error in the observable (it is assumed the error process of the observable is stationary and hence that the variance of error in the observable is constant and unchanging for every data point).

If one assumes the existence of no a priori information, then the error covariance sought, assuming discrete observations, may be formed as the inverse of the information array:

$$\Gamma = \left[ \frac{1}{\sigma_\rho^2} \sum_{n=1}^N \left( \frac{\partial \rho_n}{\partial \mathbf{X}} \right)^T \left( \frac{\partial \rho_n}{\partial \mathbf{X}} \right) \right]^{-1} \quad (63)$$

where  $\mathbf{X}$  is the vector of estimated parameters (in this case this is the spacecraft state in classical elements relative to the plane-of-the-sky), which shall be ordered in the following way:

$$\mathbf{X} = (p_1, p_2, p_3, p_4, p_5, p_6)^T = (\theta, a', i, e, T_P, \omega)^T$$

and  $N$  is the number of data points (observations).

In this case, to simplify the accumulation, continuous observations are assumed, and hence the sums involved will become integrals. For further simplicity, the accumulation shall be performed over a full orbit period in the time domain  $P$  (in the manner in which "real" data is taken, in equal increments of time), beginning at  $-P/2$  and ending at  $+P/2$ . Since the observable and the partials are all formulated in terms of the eccentric anomaly,  $E$ , it is logical to perform the accumulation with this parameter as the independent variable. From Eq. (37), one has

$$dt = \frac{(1 - e \cos E)}{n} dE \quad (64)$$

Thus a general information array element,  $I_{ij}$ , may be stated as

$$I_{ij} = \frac{1}{(\alpha \sigma_\rho^2)} \int_{-\pi}^{\pi} \left( \frac{\partial \rho}{\partial p_i} \right) \left( \frac{\partial \rho}{\partial p_j} \right) \frac{(1 - e \cos E)}{n} dE \quad (65)$$

where  $\alpha$  is a scale factor chosen such that the information array,  $I_{ij}$ , obtained continuously is the same as that formed discretely.

To ascertain the value of  $\alpha$ , let  $\mathbf{X}$  consist of only one parameter, with its attendant partial derivative being a constant,  $K$ . Then the discrete information is just

$$I_{11} = \left[ \frac{K^2 N}{\sigma_\rho^2} \right] \quad (66)$$

The continuous information, however, is

$$I_{11} = \left[ \frac{K^2 P}{\alpha \sigma_\rho^2} \right] \quad (67)$$

where  $P$  is the orbit period (in seconds).

Therefore, for the two informations to be equal, one must have

$$\alpha = \left( \frac{P}{N} \right) \quad (68)$$

This, however, implies that

$$\alpha = \Delta t \quad (69)$$

where  $\Delta t$  is the discrete sample rate (in seconds).

In the standard case of two-way, coherent range rate as the orbital data type, the discrete sample rate is 60 seconds.

Thus the elements of the information array may now be stated using the continuous formulation as

$$I_{ij} = \frac{1}{(\sigma_\rho^2 \Delta t)} \int_{-\pi}^{\pi} \left( \frac{\partial \dot{\rho}}{\partial p_i} \right) \left( \frac{\partial \dot{\rho}}{\partial p_j} \right) \frac{(1 - e \cos E)}{n} dE \quad (70)$$

For simplicity, scaled information array elements,  $I'_{ij}$ , shall be stated where

$$I'_{ij} = \left( \frac{\sigma_\rho^2 n \Delta t}{\pi} \right) I_{ij} \quad (71)$$

Note that Eq. (71) implies that when the determinant of the full  $6 \times 6$  information array is computed, to obtain a correct value, the result must be multiplied by

$$\left( \frac{\sigma_\rho^2 n \Delta t}{\pi} \right)^6$$

Similarly for the cofactors of dimension  $5 \times 5$ , the result must be multiplied by

$$\left( \frac{\sigma_{\rho}^2 n \Delta t}{\pi} \right)^5$$

Since the information array is symmetric, only the upper diagonal portion of the array shall be stated.

If one examines Eq. (42), it can be seen that the range-rate value obtained for a certain value of  $\omega$  is the negative of that for  $\omega + \pi$ . The partials will also be of opposite signs. The information array elements will, however, be the same. Hence the information array must be periodic in  $\omega$ , but go through two full periods as  $\omega$  goes from zero to 360 degrees.

With the complicated nature of the partials, there is a need for a very specific table of integrals to perform the accumulation. Such a table is provided in the Appendices. With the use of the Appendices, the elements of the information array can be straightforwardly computed.

Due to the complexity of the information array, any further simplification will be most welcome. We are only interested in obtaining the variance (standard deviation) of  $\theta$ , which is the error in the line-of-nodes relative to the plane-of-the-sky. Any element in the inverted information array (the covariance),  $C_{ij}$ , can be stated as

$$C_{ij} = \left[ \frac{(-1)^{i+j} Co D_{ij}}{D} \right] \quad (72)$$

where  $D$  is the determinant of the full  $6 \times 6$  information array,  $I'$ , with the appropriate constant factors included, and  $Co D_{ij}$  is the cofactor of the  $ij$ th element of the information array, which is just the determinant of the remaining  $5 \times 5$  array obtained by removing row  $i$  and column  $j$  from the full  $6 \times 6$  information array, with also the appropriate constant factors included.

In this case, we are interested in  $C_{11}$  only. This is just

$$C_{11} = \sigma_{\theta}^2 = \frac{Co D_{11}}{D} \quad (73)$$

If it is assumed that there are removable coefficients,  $a_i$ , in each of the  $i$ th rows and  $i$ th columns of  $I'$ , and similar coefficients,  $b_i$ , coming from the  $5 \times 5$  matrix whose determinant is the  $Co D_{11}$ , then

$$C_{11} = \sigma_{\theta}^2 = \left( \frac{Co D'_{11}}{D'} \right) \frac{\prod_{j=1}^5 b_j^2}{\prod_{i=1}^6 a_i^2} = \frac{1}{a_1^2} \left( \frac{Co D'_{11}}{D'} \right) \quad (74)$$

where  $Co D'_{11}$  and  $D'$  are the corresponding determinants with common factors removed.

One can now state the information array,  $I'$ , to compute  $D'$  and  $Co D'_{11}$ . The information array, with coefficients removed, is shown below.

$$I'_{11} = \left[ \left\{ \frac{\pi^2}{3} + 1 - 2e + 2e^2 \right\} (\alpha_3^2 + \alpha_4^2) + \left( \frac{2}{e^2} \right) \left\{ \beta S(\gamma) + 2\beta^2 \ln \left\{ \frac{1+\beta}{2(1+e)} \right\} - \frac{\pi^2}{6} (2-e^2) - \beta^3 + 3e\beta^2 + e - \frac{1}{2}(3-5\beta^2) \right\} (\alpha_4^2 - \alpha_3^2) \right] \quad (75)$$

where

$$(\alpha_3^2 + \alpha_4^2) = \sin^2 \theta + \cos^2 \theta \cos^2 i \quad (76)$$

$$(\alpha_4^2 - \alpha_3^2) = 2 \sin 2\omega \sin \theta \cos \theta \cos i - \cos 2\omega (\sin^2 \theta - \cos^2 \theta \cos^2 i) \quad (77)$$

$$I'_{12} = \left[ \left\{ \frac{1}{\beta^2} S(\gamma) - \frac{2}{\beta} \ln(1+\gamma) - (1+\beta) + \frac{2\beta}{(1+e)} \right\} \sin \theta - \frac{2\beta}{e^2} \left\{ (2 \ln(1+\gamma) + 3 \ln \left\{ \frac{1+\beta}{2(1+e)} \right\}) + 2e + \frac{1}{2}(1-\beta)(1-3\beta) \right\} \times (\sin \theta \cos 2\omega - \cos \theta \cos i \sin 2\omega) \right] \quad (78)$$

$$I'_{13} = \beta \left[ \sin \theta + \left\{ \gamma^2 + \frac{2}{e} + \frac{2}{e^2} \ln \left\{ \frac{1+\beta}{2(1+e)} \right\} \right\} (\sin \theta \cos 2\omega - \cos \theta \cos i \sin 2\omega) \right] \quad (79)$$

$$I'_{14} = \left[ \left( \frac{-1}{e\beta} \right) \left\{ \ln \left\{ \frac{1+\beta}{2(1+e)} \right\} + 2 - \beta(1+\beta) \right\} \sin \theta + \left( \frac{-1}{e^3\beta} \right) \left\{ (2-e^2) \ln \left\{ \frac{1+\beta}{2(1+e)} \right\} + (2e^2-1)(1-\beta) + 2e - e^4 \right\} (\sin \theta \cos 2\omega - \cos \theta \cos i \sin 2\omega) \right] \quad (80)$$

$$I'_{15} = \left[ \left( \frac{1+2e}{1+e} \right) \cos \theta \cos i - \left( \frac{2+e-\beta(1+2e)}{(1+e)(1+\beta)} \right) (\sin \theta \sin 2\omega + \cos \theta \cos i \cos 2\omega) \right] \quad (81)$$

$$I'_{16} = \beta \left[ \cos \theta \cos i - \left( \gamma^2 + \frac{2}{e} + \frac{2}{e^2} \ln \left\{ \frac{1+\beta}{2(1+e)} \right\} \right) (\sin \theta \sin 2\omega + \cos \theta \cos i \cos 2\omega) \right] \quad (82)$$

$$I'_{22} = \left[ \left\{ \frac{(2+e^2)}{2\beta^5} S(\gamma) - \frac{6}{\beta^4} \ln(1+\gamma) + \frac{3}{\beta^4} (1-e) - \frac{1}{2\beta^3} (\beta^2 + 2\beta + 3) \right\} - \cos 2\omega \left\{ \frac{e^2}{4\beta^5} S(\gamma) + \frac{(2-5e^2)}{3e^2\beta^4} \ln(1+\gamma) + \frac{(1-e)(2+e^2)}{6e^2\beta^4} + \frac{1}{4\beta} - \frac{(1+\beta)(3-\beta)}{12e^2\beta^3} \right\} \right] \quad (83)$$

$$I'_{23} = \left[ \frac{-e}{(1+e)} + \left\{ \frac{1}{1+\beta} - \frac{e}{1+e} \right\} \cos 2\omega \right] \quad (84)$$

$$I'_{24} = \left[ \frac{1}{2e} \left\{ \frac{1}{\beta} - \frac{1}{(1+e)^2} \right\} + \frac{\cos 2\omega}{12e^3} \left\{ \frac{1}{\beta} (5 - 14\beta + 9\beta^2) + \frac{2e^2}{(1+e)^2} \right\} \right] \quad (85)$$



$$I'_{25} = \left( \frac{1}{12e^2\beta^2} \right) \left[ 1 - 3\beta^2 + \frac{2(1+3e)\beta^3}{(1+e)^3} \right] \sin 2\omega \quad (86)$$

$$I'_{26} = \left[ \frac{e}{1+e} - \frac{1}{1+\beta} \right] \sin 2\omega \quad (87)$$

$$I'_{33} = 1 - \left( \frac{1-\beta}{1+\beta} \right) \cos 2\omega \quad (88)$$

$$I'_{34} = -\frac{(1-\beta)}{e\beta(1+\beta)} \cos 2\omega \quad (89)$$

$$I'_{35} = 0 \quad (90)$$

$$I'_{36} = \left( \frac{1-\beta}{1+\beta} \right) \sin 2\omega \quad (91)$$

$$I'_{44} = \frac{1}{4\beta^3(1+\beta)} \left[ 2(3+\beta) + \frac{(1-\beta)^2}{(1+\beta)} \cos 2\omega \right] \quad (92)$$

$$I'_{45} = \left( \frac{e}{4\beta^4} \right) \sin 2\omega \quad (93)$$

$$I'_{46} = \frac{(1-\beta)}{e\beta(1+\beta)} \sin 2\omega \quad (94)$$

$$I'_{55} = \frac{1}{4\beta^5} \left[ 2(2+e^2) - e^2 \cos 2\omega \right] \quad (95)$$

$$I'_{56} = \frac{1}{\beta^2} \quad (96)$$

$$I'_{66} = 1 + \left( \frac{1-\beta}{1+\beta} \right) \cos 2\omega \quad (97)$$

where the removable coefficients, of course, turn out to be the coefficients of the individual partial derivatives, which are

$$a_1 = \left( \frac{-V_N a}{\rho_{ET}} \right) \quad (98)$$

$$a_2 = \left( \frac{-3n \sin i}{2} \right) \quad (99)$$

$$a_3 = an \cos i \quad (100)$$

$$a_4 = an \sin i \quad (101)$$

$$a_5 = -an^2 \sin i \quad (102)$$

$$a_6 = an \sin i \quad (103)$$

From these scale factors and Eq. (71), one can infer that the error in  $\theta$  may be written as

$$\sigma_\theta = \sigma_\rho \left[ \frac{\rho_{ET}(\mu)^{\frac{1}{2}}(\Delta t)^{\frac{1}{2}}}{V_N(a)^{\frac{1}{2}}(\pi)^{\frac{1}{2}}} \right] \left[ \frac{f(e, \omega)}{g(\theta, i, e, \omega)} \right]^{\frac{1}{2}} \quad (104)$$

where  $f(e, \omega)$  is the cofactor of  $D'_{11}$ , and  $g(\theta, i, e, \omega)$  is just the determinant,  $D'$ .

## V. Inverting the Information Array

Since the interest is in a single element of the inversion, only those computations required in Eq. (74) shall be performed. Before plunging into the general elliptic case, the inversion for the special case of the circular orbit (i.e.,  $e = 0$ ) shall be demonstrated.

### A. Circular Case

After much work, the elements of the information array (in the limit as  $e$  tends to zero) can be shown to be:

$$I'_{11} = \left( \frac{\pi^2}{3} + 1 \right) (\alpha_3^2 + \alpha_4^2) - \frac{1}{2} (\alpha_4^2 - \alpha_3^2) \quad (105)$$

$$I'_{12} = \frac{\pi^2}{3} \sin \theta \quad (106)$$

$$I'_{13} = \sin \theta + \frac{1}{2} (\sin \theta \cos 2\omega - \cos \theta \cos i \sin 2\omega) \quad (107)$$

$$I'_{14} = \sin \theta - \frac{1}{3} (\sin \theta \cos 2\omega - \cos \theta \cos i \sin 2\omega) \quad (108)$$

$$I'_{15} = \cos \theta \cos i - \frac{1}{2} (\sin \theta \sin 2\omega + \cos \theta \cos i \cos 2\omega) \quad (109)$$

$$I'_{16} = \cos \theta \cos i - \frac{1}{2} (\sin \theta \sin 2\omega + \cos \theta \cos i \cos 2\omega) \quad (110)$$

$$I'_{22} = \frac{\pi^2}{3} - \frac{1}{2} \cos 2\omega \quad (111)$$

$$I'_{23} = \frac{1}{2} \cos 2\omega \quad (112)$$

$$I'_{24} = 1 - \frac{1}{3} \cos 2\omega \quad (113)$$

$$I'_{25} = -\frac{1}{2} \sin 2\omega \quad (114)$$

$$I'_{26} = -\frac{1}{2} \sin 2\omega \quad (115)$$

$$I'_{33} = I'_{44} = I'_{55} = I'_{56} = I'_{66} = 1 \quad (116)$$

$$I'_{34} = I'_{35} = I'_{36} = I'_{45} = I'_{46} = 0 \quad (117)$$

To perform the inversion in this special, circular case, it is necessary to remove the parameter  $\omega$  from the estimation since the argument of periapsis is indeterminate. Additionally the parameter  $T_P$  is removed from the estimation since the partials are large and the same as those of  $\omega$ , only with the opposite sign. This rank reduction corresponds to assuming that the true anomaly of the spacecraft at epoch is perfectly known, which is realistic in comparison to the error in  $\theta$ . With a large amount of algebraic manipulation, the four-parameter inversion yields an error in  $\theta$  of

$$\sigma_\theta = \sigma_{\dot{\rho}} \left[ \frac{\rho_{ET}(\mu)^{\frac{1}{4}}(\Delta t)^{\frac{1}{2}}}{V_N(a)^{\frac{1}{2}}(\pi)^{\frac{1}{2}}} \right] \left[ \frac{f(\omega)}{g(\theta, i, \omega)} \right]^{\frac{1}{2}} \quad (118)$$

where

$$f(\omega) = \left[ \frac{\pi^2}{3} - 1 + \frac{1}{6} \cos 2\omega - \frac{13}{36} \cos^2 2\omega \right] \quad (119)$$

$$g(\theta, i, \omega) = \left( \frac{\cos^2 \theta \cos^2 i}{9} \right) \left[ \pi^4 - \pi^2 \cos 2\omega - \frac{13}{12} \pi^2 \cos^2 2\omega - \frac{47}{6} + \frac{61}{8} \cos 2\omega - \frac{31}{6} \cos^2 2\omega \right] \quad (120)$$

It is now apparent from Eqs. (118)-(120) that the use of range-rate data for state estimation of a circular orbit about a distant planet can yield singular results. Singular state errors will result if

$$\theta = 90 \text{ deg, } 270 \text{ deg, and/or} \quad (121)$$

$$i = -90 \text{ deg, } 90 \text{ deg, and/or} \quad (122)$$

$$V_N = 0 \quad (123)$$

This latter condition of singularity may occur when the spacecraft is orbiting a superior planet (at the beginning and end points of retrograde motion). These are important results, which are significantly at variance with orbit determination experience for highly eccentric trajectories, which appear well behaved and certainly not singular. Note that even with zero eccentricity, the results of Eqs. (119) and (120) imply a functional dependence upon  $\omega$ . This is an artifact of the definition of the coordinate system at time zero. If we assume  $\omega = 90$  deg, then the maximum variation in the error in  $\theta$  for any other value of  $\omega$  can be shown to be 13.6 percent. Thus the error in  $\theta$  can be said to be only "weakly" dependent upon  $\omega$ .

## B. Elliptic Case

Many thanks go to R. J. Muellerschoen of the Navigation Systems Section for his help in using *MATHEMATICA* to attempt the  $6 \times 6$  inversion. Even with a large amount of available memory, the problem quickly consumed it all and more. Thus it appears unlikely that a "reasonable" functionality for  $f$  and  $g$  can be obtained in the general elliptic case. However, in having the information array, we have performed the accumulation in a general manner which will save much time. A large number of orbit-determination studies as a function of geometry relative to the plane-of-the-sky can be made quickly and cheaply.

## VI. Results

### A. Circular Case

Venus Orbiting Imaging Radar (VOIR), the progenitor of Magellan, was extensively studied many years ago. It was in these studies that the circular orbit singularities were first observed. The nominal orbit just happened to have conditions that were near-singular. As a result, orbit determination errors were obtained that were outrageously large (larger than the nominal orbit). This led to the use of several different computer programs to examine whether there were program errors or not. The two main programs used to investigate the state estimation were the Orbit Determination Program (ODP) and the Sequential Orbit Determination (SOD) Program. Both of these pieces of software yielded the same, poor orbit determination behavior and so confirmed the belief that we had inadvertently stumbled upon singularities in the orbit determination process for circular orbiters. The important parameters defining the nominal VOIR orbit were:

Epoch	March 17, 1982
Semi-major axis, $a$	6552 km
Eccentricity, $e$	$1.52 \times 10^{-4}$
Inclination, $i$	78.17 deg (relative to the plane-of-the-sky)
Argument of periapsis, $\omega$	170.7 deg (relative to the plane-of-the-sky)
Cross-velocity, $V_N$	14.413 km/sec
Target range, $\rho_{ET}$	$8.36765 \times 10^7$ km
Angle, $\theta$	85 deg
Period, $P$	1.624 hr

The computer solutions were assumed to have as a data base one full orbit of coherent, two-way range-rate data. A slight problem complicating exact comparison with the theory is that the computer programs employed assumed that when the spacecraft was not visible to the Earth tracking station, there would be no data available. This eliminated about 36 minutes of data (depending on the nominal inclination relative to the plane-of-the-sky) out of a total of 98 minutes available. The theory, however, assumes that a full orbit of data is available no matter what. As a result the displayed errors are not in perfect agreement. The trends as functions of the angle  $\theta$  and the inclination  $i$ , however, are in excellent agreement and as seen in Figs. 2 and 3, clearly demonstrate the singular behavior of the circular orbit. The best place to compare theory and computer results is in Fig. 3 where the value of the inclination relative to the plane-of-the-sky is small. In this configuration, a full orbit of data would be available to both theory and computer results, yielding agreement to about 30 percent. As the inclination increases, the agreement between theory and computer results degrades as more data is occulted. The error parameter displayed in these figures is

$$|\sigma_x| = r\sigma_\theta = a\sigma_\theta \quad (124)$$

### B. Elliptic Case

The orbit determination studies for the general elliptic case are yet to be published.<sup>1</sup>

---

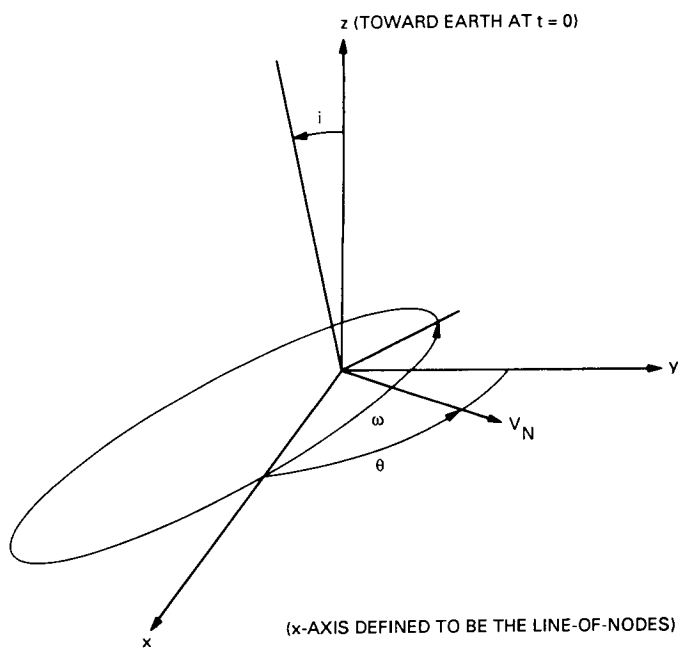
<sup>1</sup> For additional information on the initial studies which helped to delineate the circular orbit problems, see the following: J. Ellis and R. K. Russell, "Earth-Based Determination of a Near-Circular Orbit About a Distant Planet," JPL Technical Memorandum 391-406 (internal document), March 9, 1973; R. K. Russell, "Improved Orbit Determination of Low-Altitude, Nearly Circular Planetary Orbiters," Engineering Memorandum 391-561 (internal document), April 25, 1974; R. K. Russell, "Analytic Orbit Determination, A General Methodology and a Successful Example," JPL Engineering Memorandum 314-59 (internal document), August 9, 1976; and L. J. Wood "Orbit Determination Singularities in the Doppler Tracking of a Planetary Orbiter," *Journal of Guidance, Control, and Dynamics*, volume 9, number 4, p. 485, July-August 1986.

## VII. Conclusions

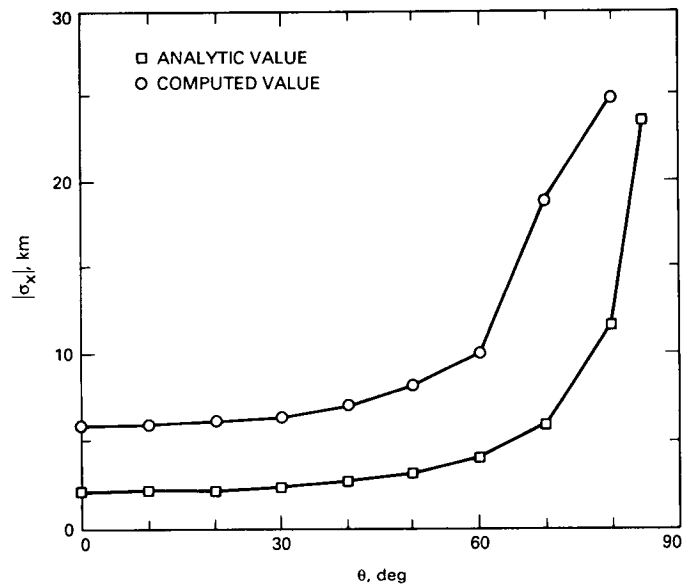
It has been seen that circular orbits about distant planets may suffer singularities in over-all position error estimation. These singularities are due to orbit inclination, placement of the line-of-nodes, and insignificant cross-velocity at the start and end of retrograde motion when orbiting a superior planet.

Even though these conclusions appear to yield poor state estimation, one should not be unduly alarmed inasmuch as the stated conditions for singularity are not maintained for extended periods during typical mission scenarios. However, mission analysts should be aware of these potential pitfalls and realize that spuriously large results for circular orbiters can be obtained and are not the result of incorrect assumptions or faulty software.

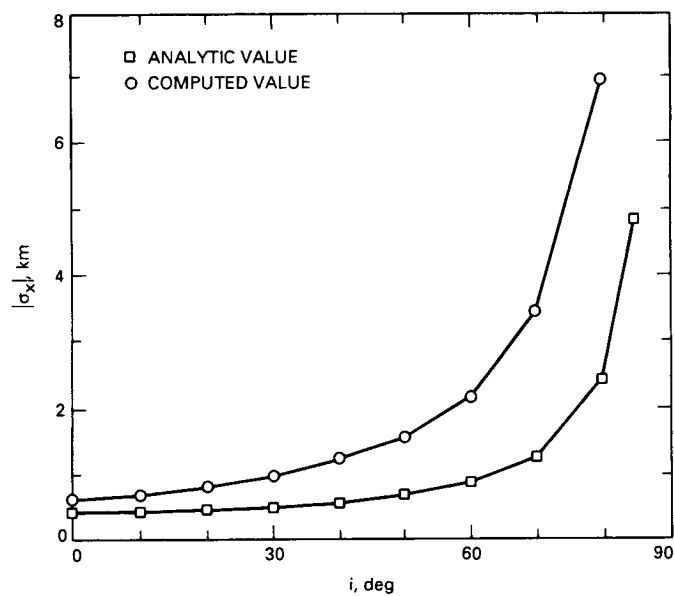
The general elliptic problem appears so involved that analytic inversion at this time is just not feasible, and in any case the resulting expression for the position error would likely be so lengthy that any understanding would be lost in the maze.



**Fig. 1. Spacecraft orbital geometry in the plane of the sky.**



**Fig. 2. Norm of position error versus  $\theta$  (with inclination to the plane of the sky = 78.17 deg).**



**Fig. 3. Norm of position error versus inclination,  $i$ , to the plane of the sky (with  $\theta = 0$  deg).**

# Appendix A

## Required Integrals I

To perform the accumulation of the information array, a large number of difficult integrals of mixed trigonometric and algebraic functions and their powers are required. Contained herein are the necessary results. Note that due to the limits of integration, all integrals of odd functions will vanish.

An important, general integral, from which some of the results contained herein are derived, is

$$I(a, b, n) = \int_{-\pi}^{\pi} \frac{dE}{(a + b \cos E)^{n+1}} = \frac{2\pi}{(a^2 - b^2)^{\frac{n+1}{2}}} P_n \left( \frac{a}{\sqrt{a^2 - b^2}} \right) \quad (\text{A-1})$$

where  $P_n$  are the Legendre polynomials.

This integral is found in [1; p. 383, 3.661, 4].

For simplicity and convenience, we shall define:

$$\beta = \sqrt{1 - e^2} \quad (\text{A-2})$$

$$\gamma = \left( \frac{1 - \beta}{e} \right) \quad (\text{A-3})$$

The integrals necessary to perform the accumulation are:

$$\int_{-\pi}^{\pi} (1 - e \cos E) dE = 2\pi \quad (\text{A-4})$$

$$\int_{-\pi}^{\pi} (1 - e \cos E) \cos E dE = -\pi e \quad (\text{A-5})$$

$$\int_{-\pi}^{\pi} (1 - e \cos E) \cos^2 E dE = \pi \quad (\text{A-6})$$

$$\int_{-\pi}^{\pi} dE = 2\pi \quad (\text{A-7})$$

$$\int_{-\pi}^{\pi} \cos E dE = 0 \quad (\text{A-8})$$

$$\int_{-\pi}^{\pi} \cos^2 E dE = \pi \quad (\text{A-9})$$

$$\int_{-\pi}^{\pi} \cos^3 E dE = 0 \quad (\text{A-10})$$

$$\int_{-\pi}^{\pi} E \sin E dE = 2\pi \quad (\text{A-11})$$

$$\int_{-\pi}^{\pi} E \sin E \cos E dE = -\left(\frac{\pi}{2}\right) \quad (\text{A-12})$$

$$\int_{-\pi}^{\pi} \frac{dE}{(1 - e \cos E)} = \left( \frac{2\pi}{\beta} \right) \quad (\text{A-13})$$

$$\int_{-\pi}^{\pi} \frac{\cos E dE}{(1 - e \cos E)} = \left( \frac{2\pi}{e\beta} \right) (1 - \beta) = \left( \frac{2\pi}{\beta} \right) \gamma \quad (\text{A-14})$$

$$\int_{-\pi}^{\pi} \frac{\cos^2 E dE}{(1 - e \cos E)} = \left( \frac{2\pi}{e^2\beta} \right) (1 - \beta) = \left( \frac{2\pi}{e\beta} \right) \gamma \quad (\text{A-15})$$

$$\int_{-\pi}^{\pi} \frac{\cos^3 E dE}{(1 - e \cos E)} = \left( \frac{\pi}{e^3\beta} \right) (1 - \beta)^2 (2 + \beta) = \left( \frac{\pi}{e\beta} \right) \gamma^2 (2 + \beta) \quad (\text{A-16})$$

$$\int_{-\pi}^{\pi} \frac{\cos^4 E dE}{(1 - e \cos E)} = \left( \frac{\pi}{e^4\beta} \right) (1 - \beta)^2 (2 + \beta) = \left( \frac{\pi}{e^2\beta} \right) \gamma^2 (2 + \beta) \quad (\text{A-17})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E dE}{(1 - e \cos E)} = - \left( \frac{2\pi}{e} \right) \ln \left\{ \frac{1 + \beta}{2(1 + e)} \right\} \quad (\text{A-18})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E \cos E dE}{(1 - e \cos E)} = - \left( \frac{2\pi}{e} \right) \left[ 1 + \frac{1}{e} \ln \left\{ \frac{1 + \beta}{2(1 + e)} \right\} \right] \quad (\text{A-19})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E \cos^2 E dE}{(1 - e \cos E)} = \left( \frac{\pi}{2e^2} \right) \left[ e - 4 \left[ 1 + \frac{1}{e} \ln \left\{ \frac{1 + \beta}{2(1 + e)} \right\} \right] \right] \quad (\text{A-20})$$

For a full development of the following integral, see Appendix B.

$$\int_{-\pi}^{\pi} \frac{E^2 dE}{(1 - e \cos E)} = \left( \frac{2\pi}{\beta} \right) S(\gamma) \quad (\text{A-21})$$

where

$$S(\gamma) = \left[ \frac{\pi^2}{3} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n \gamma^n}{n^2} \right] \quad (\text{A-22})$$

Continuing,

$$\int_{-\pi}^{\pi} \frac{E^2 \cos^2 E dE}{(1 - e \cos E)} = \left( \frac{2\pi}{e^2\beta} \right) \left[ S(\gamma) + \beta \left( 2e - \frac{\pi^2}{3} \right) \right] \quad (\text{A-23})$$

$$\int_{-\pi}^{\pi} \frac{dE}{(1 - e \cos E)^2} = \left( \frac{2\pi}{\beta^3} \right) \quad (\text{A-24})$$

$$\int_{-\pi}^{\pi} \frac{\cos E dE}{(1 - e \cos E)^2} = \left( \frac{2\pi e}{\beta^3} \right) \quad (\text{A-25})$$

$$\int_{-\pi}^{\pi} \frac{\cos^2 E dE}{(1 - e \cos E)^2} = \left( \frac{2\pi}{e^2\beta^3} \right) (1 - \beta)(1 + \beta - \beta^2) = \left( \frac{2\pi}{e\beta^3} \right) \gamma(1 + \beta - \beta^2) \quad (\text{A-26})$$



$$\int_{-\pi}^{\pi} \frac{\cos^3 E dE}{(1 - e \cos E)^2} = \left( \frac{2\pi}{e^3 \beta^3} \right) (1 - \beta)^2 (1 + 2\beta) = \left( \frac{2\pi}{e \beta^3} \right) \gamma^2 (1 + 2\beta) \quad (\text{A-27})$$

$$\int_{-\pi}^{\pi} \frac{\cos^4 E dE}{(1 - e \cos E)^2} = \left( \frac{\pi}{e^4 \beta^3} \right) (1 - \beta)^2 (2 + 4\beta - 2\beta^2 - \beta^3) = \left( \frac{\pi}{e^2 \beta^3} \right) \gamma^2 (2 + 4\beta - 2\beta^2 - \beta^3) \quad (\text{A-28})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E dE}{(1 - e \cos E)^2} = \left( \frac{2\pi}{e} \right) \left[ \frac{1}{\beta} - \frac{1}{1+e} \right] = \left( \frac{2\pi}{\beta^2} \right) (1 - \gamma) \quad (\text{A-29})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E \cos E dE}{(1 - e \cos E)^2} = \left( \frac{2\pi}{e^2} \right) \left[ \frac{1}{\beta} - \frac{1}{1+e} + \ln \left\{ \frac{1+\beta}{2(1+e)} \right\} \right] \quad (\text{A-30})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E \cos^2 E dE}{(1 - e \cos E)^2} = \left( \frac{2\pi}{e^3} \right) \left[ \frac{1}{\beta} - \frac{1}{1+e} + e + 2 \ln \left\{ \frac{1+\beta}{2(1+e)} \right\} \right] \quad (\text{A-31})$$

$$\int_{-\pi}^{\pi} \frac{E^2 dE}{(1 - e \cos E)^2} = \left( \frac{2\pi}{\beta^3} \right) \left[ S(\gamma) - 4\beta \ln(1 + \gamma) \right] \quad (\text{A-32})$$

$$\int_{-\pi}^{\pi} \frac{dE}{(1 - e \cos E)^3} = \left( \frac{\pi}{\beta^5} \right) (3 - \beta^2) \quad (\text{A-33})$$

$$\int_{-\pi}^{\pi} \frac{\cos E dE}{(1 - e \cos E)^3} = \left( \frac{3\pi e}{\beta^5} \right) \quad (\text{A-34})$$

$$\int_{-\pi}^{\pi} \frac{\cos^2 E dE}{(1 - e \cos E)^3} = \left( \frac{\pi}{\beta^5} \right) (3 - 2\beta^2) \quad (\text{A-35})$$

$$\int_{-\pi}^{\pi} \frac{\cos^3 E dE}{(1 - e \cos E)^3} = \left( \frac{\pi}{e^3 \beta^5} \right) (1 - \beta)^2 (3 + 6\beta + 2\beta^2 - 2\beta^3) = \left( \frac{\pi}{e \beta^5} \right) \gamma^2 (3 + 6\beta + 2\beta^2 - 2\beta^3) \quad (\text{A-36})$$

$$\int_{-\pi}^{\pi} \frac{\cos^4 E dE}{(1 - e \cos E)^3} = \left( \frac{3\pi}{e^4 \beta^5} \right) (1 - \beta)^2 (1 + 2\beta - 2\beta^3) = \left( \frac{3\pi}{e^2 \beta^5} \right) \gamma^2 (1 + 2\beta - 2\beta^3) \quad (\text{A-37})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E dE}{(1 - e \cos E)^3} = \left( \frac{\pi}{e} \right) \left[ \frac{1}{\beta^3} - \frac{1}{(1+e)^2} \right] \quad (\text{A-38})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E \cos E dE}{(1 - e \cos E)^3} = \left( \frac{\pi}{e^2} \right) \left[ \frac{1}{\beta^3} (1 - 2\beta^2) + \frac{(1+2e)}{(1+e)^2} \right] \quad (\text{A-39})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E \cos^2 E dE}{(1 - e \cos E)^3} = \left( \frac{\pi}{e^3} \right) \left[ \frac{1}{\beta^3} (1 - 4\beta^2) + \frac{(3+4e)}{(1+e)^2} - 2 \ln \left\{ \frac{1+\beta}{2(1+e)} \right\} \right] \quad (\text{A-40})$$

$$\int_{-\pi}^{\pi} \frac{E^2 dE}{(1 - e \cos E)^3} = \left( \frac{\pi}{\beta^5} \right) \left[ (3 - \beta^2) S(\gamma) - 12\beta \ln(1 + \gamma) - 2e\beta(1 - \gamma) \right] \quad (\text{A-41})$$

$$\int_{-\pi}^{\pi} \frac{E^2 \cos E dE}{(1 - e \cos E)^3} = \left( \frac{\pi}{e \beta^5} \right) \left[ 3(1 - \beta^2) S(\gamma) - 4\beta(3 - 2\beta^2) \ln(1 + \gamma) - 2e\beta(1 - \gamma) \right] \quad (\text{A-42})$$

$$\int_{-\pi}^{\pi} \frac{E^2 \cos^2 E dE}{(1 - e \cos E)^3} = \left( \frac{\pi}{e^2 \beta^5} \right) \left[ (3 - 5\beta^2 + 2\beta^4) S(\gamma) - 4\beta(3 - 4\beta^2) \ln(1 + \gamma) - 2e\beta(1 - \gamma) \right] \quad (\text{A-43})$$

$$\int_{-\pi}^{\pi} \frac{dE}{(1 - e \cos E)^4} = \left( \frac{\pi}{\beta^7} \right) (5 - 3\beta^2) \quad (\text{A-44})$$

$$\int_{-\pi}^{\pi} \frac{\cos E dE}{(1 - e \cos E)^4} = \left( \frac{\pi e}{\beta^7} \right) (5 - \beta^2) \quad (\text{A-45})$$

$$\int_{-\pi}^{\pi} \frac{\cos^2 E dE}{(1 - e \cos E)^4} = \left( \frac{\pi}{\beta^7} \right) (5 - 4\beta^2) \quad (\text{A-46})$$

$$\int_{-\pi}^{\pi} \frac{\cos^3 E dE}{(1 - e \cos E)^4} = \left( \frac{\pi e}{\beta^7} \right) (5 - 2\beta^2) \quad (\text{A-47})$$

$$\begin{aligned} \int_{-\pi}^{\pi} \frac{\cos^4 E dE}{(1 - e \cos E)^4} &= \left( \frac{\pi}{e^4 \beta^7} \right) (1 - \beta)^2 (5 + 10\beta - 10\beta^3 - 4\beta^4 + 2\beta^5) \\ &= \left( \frac{\pi}{e^2 \beta^7} \right) \gamma^2 (5 + 10\beta - 10\beta^3 - 4\beta^4 + 2\beta^5) \end{aligned} \quad (\text{A-48})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E dE}{(1 - e \cos E)^4} = \left( \frac{\pi}{3e} \right) \left[ \frac{(3 - \beta^2)}{\beta^5} - \frac{2}{(1 + e)^3} \right] \quad (\text{A-49})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E \cos E dE}{(1 - e \cos E)^4} = \left( \frac{\pi}{3e^2} \right) \left[ \frac{1}{\beta^5} (3 - 4\beta^2) + \frac{(1 + 3e)}{(1 + e)^3} \right] \quad (\text{A-50})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E \cos^2 E dE}{(1 - e \cos E)^4} = \left( \frac{\pi}{3e^3} \right) \left[ \frac{1}{\beta^5} (3 - 7\beta^2 + 6\beta^4) - \frac{2}{(1 + e)^3} (1 + 3e + 3e^2) \right] \quad (\text{A-51})$$

$$\int_{-\pi}^{\pi} \frac{dE}{(1 - e \cos E)^5} = \left( \frac{\pi}{4\beta^9} \right) (35 - 30\beta^2 + 3\beta^4) \quad (\text{A-52})$$

$$\int_{-\pi}^{\pi} \frac{\cos E dE}{(1 - e \cos E)^5} = \left( \frac{5\pi e}{4\beta^9} \right) (7 - 3\beta^2) \quad (\text{A-53})$$

$$\int_{-\pi}^{\pi} \frac{\cos^2 E dE}{(1 - e \cos E)^5} = \left( \frac{\pi}{4\beta^9} \right) (35 - 35\beta^2 + 4\beta^4) \quad (\text{A-54})$$

$$\int_{-\pi}^{\pi} \frac{\cos^3 E dE}{(1 - e \cos E)^5} = \left( \frac{5\pi e}{4\beta^9} \right) (7 - 4\beta^2) \quad (\text{A-55})$$

$$\int_{-\pi}^{\pi} \frac{\cos^4 E dE}{(1 - e \cos E)^5} = \left( \frac{\pi}{4\beta^9} \right) (35 - 40\beta^2 + 8\beta^4) \quad (\text{A-56})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E dE}{(1 - e \cos E)^5} = \left( \frac{\pi}{4e} \right) \left[ \frac{(5 - 3\beta^2)}{\beta^7} - \frac{2}{(1 + e)^4} \right] \quad (\text{A-57})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E \cos E dE}{(1 - e \cos E)^5} = \left( \frac{\pi}{12e^2} \right) \left[ \frac{1}{\beta^7} (15 - 21\beta^2 + 4\beta^4) + \frac{2}{(1 + e)^4} (1 + 4e) \right] \quad (\text{A-58})$$

$$\int_{-\pi}^{\pi} \frac{E \sin E \cos^2 E dE}{(1 - e \cos E)^5} = \left( \frac{\pi}{12e^3} \right) \left[ \frac{1}{\beta^7} (15 - 33\beta^2 + 20\beta^4) - \frac{2}{(1+e)^4} (1 + 4e + 6e^2) \right] \quad (\text{A-59})$$

$$\int_{-\pi}^{\pi} \frac{E^2 dE}{(1 - e \cos E)^5} = \left( \frac{\pi}{12\beta^9} \right) \left[ 3(35 - 30\beta^2 + 3\beta^4)S(\gamma) - 20\beta(21 - 11\beta^2) \ln(1 + \gamma) \right. \\ \left. - 2e\beta(1 - \gamma)(57 + 12\beta - 10\beta^2) + 2e\beta^2(12 + \beta) \right] \quad (\text{A-60})$$

$$\int_{-\pi}^{\pi} \frac{E^2 \cos E dE}{(1 - e \cos E)^5} = \left( \frac{\pi}{12e\beta^9} \right) \left[ 15(1 - \beta^2)(7 - 3\beta^2)S(\gamma) - 4\beta(105 - 115\beta^2 + 16\beta^4) \ln(1 + \gamma) \right. \\ \left. - 2e\beta(1 - \gamma)(57 + 12\beta - 38\beta^2 - 4\beta^3) + 2e\beta^2(12 + \beta - 4\beta^2) \right] \quad (\text{A-61})$$

$$\int_{-\pi}^{\pi} \frac{E^2 \cos^2 E dE}{(1 - e \cos E)^5} = \left( \frac{\pi}{12e^2\beta^9} \right) \left[ 3(1 - \beta^2)(35 - 35\beta^2 + 4\beta^4)S(\gamma) - 4\beta(105 - 175\beta^2 + 68\beta^4) \ln(1 + \gamma) \right. \\ \left. - 2e\beta(1 - \gamma)(57 + 12\beta - 66\beta^2 - 8\beta^3 + 12\beta^4) + 2e\beta^2(12 + \beta - 8\beta^2) \right] \quad (\text{A-62})$$

## Appendix B

### Required Integrals II

Out of all these difficult integrals, only one class is not expressible in closed form. These are integrals of the form

$$I_{n,m} = \int_{-\pi}^{\pi} \frac{E^2 \cos^m E dE}{(1 - e \cos E)^n} \quad (\text{B-1})$$

We shall confine our interest to the lowest order of the above, namely where  $n = 1$  and  $m = 0$ . The other integrals can be simply obtained from this one, which is

$$I = \int_{-\pi}^{\pi} \frac{E^2 dE}{1 - e \cos E} \quad (\text{B-2})$$

If the square function is expanded in a Fourier series over the interval  $-\pi$  to  $\pi$ , one obtains

$$E^2 = \frac{\pi^2}{3} - 4 \left[ \cos E - \frac{1}{2^2} \cos 2E + \frac{1}{3^2} \cos 3E - \dots \right], \quad -\pi \leq E \leq \pi \quad (\text{B-3})$$

This now can be made to yield the integral we are seeking:

$$\int_{-\pi}^{\pi} \frac{E^2 dE}{1 - e \cos E} = \frac{\pi^2}{3} \int_{-\pi}^{\pi} \frac{dE}{1 - e \cos E} + 4 \int_{-\pi}^{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \left( \frac{\cos nE}{1 - e \cos E} \right) dE \quad (\text{B-4})$$

$$= \frac{2\pi^3}{3\beta} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \int_{-\pi}^{\pi} \frac{\cos nE dE}{1 - e \cos E} \quad (\text{B-5})$$

Thus,

$$\int_{-\pi}^{\pi} \frac{\cos nE dE}{1 - e \cos E} = 2 \int_0^{\pi} \frac{\cos nE dE}{1 - e \cos E} \quad (\text{B-6})$$

This latter integral can be found in [1; p. 366, 3.613] as

$$\int_0^{\pi} \frac{\cos nE dE}{1 - e \cos E} = \frac{\pi}{\beta} \left( \frac{1 - \beta}{e} \right)^n \quad (\text{B-7})$$

52-32 ✓  
239812 ✓  
310

547636 p31

N90-12788

# A Statistical Study of Radio-Source Structure Effects on Astrometric Very Long Baseline Interferometry Observations

J. S. Ulvestad  
Tracking Systems and Applications Section

*Errors from a number of sources in astrometric Very Long Baseline Interferometry (VLBI) have been reduced in recent years through a variety of methods of calibration and modeling. Such reductions have led to a situation in which the extended structure of the natural radio sources used in VLBI is a significant error source in the effort to improve the accuracy of the radio reference frame. In the past, work has been done on individual radio sources to establish the magnitude of the errors caused by their particular structures. This article reports the results of calculations on 26 radio sources in which an effort is made to determine the typical delay and delay-rate errors for a number of sources having different types of structure.*

*It is found here that for single observations of the types of radio sources present in astrometric catalogs, group-delay and phase-delay scatter in the 50-100 psec range due to source structure can be expected at 8.4 GHz on the intercontinental baselines available in the DSN. Delay-rate scatter of  $\sim 5 \times 10^{-15}$  sec sec<sup>-1</sup> (or  $\sim 0.002$  mm sec<sup>-1</sup>) is also expected. If such errors mapped directly into source position errors, they would correspond to position uncertainties of  $\sim 2-5$  nrad (0.5-1 milliarcsec, or 0.1-0.3 microdeg), similar to the best position determinations in the current JPL VLBI catalog. With the advent of wider bandwidth VLBI systems on the large DSN antennas, the system noise will be low enough so that the structure-induced errors will be a significant part of the error budget. Several possibilities for reducing the structure errors are discussed briefly, although it is likely that considerable effort will have to be devoted to the structure problem in order to reduce the typical error by a factor of two or more.*

## I. Introduction

Very Long Baseline Interferometry (VLBI) is a radio astronomical technique which can provide exceptionally

high resolution ( $\sim 5$  nrad) for observations of compact natural radio sources. It involves simultaneous observations of a radio source by two or more observatories that may be separated by a large fraction of the Earth's diameter, or

Thus, finally

$$\int_{-\pi}^{\pi} \frac{\cos nEdE}{1 - e \cos E} = \frac{2\pi}{\beta} S(\gamma) \quad (\text{B-8})$$

where

$$S(\gamma) = \left[ \frac{\pi^2}{3} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \gamma^n \right] \quad (\text{B-9})$$

$$\beta = \sqrt{1 - e^2} \quad (\text{B-10})$$

$$\gamma = \left( \frac{1 - \beta}{e} \right) \quad (\text{B-11})$$

## Reference

- [1] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, Corrected and Enlarged Edition, New York: Academic Press, Inc., 1980.

by more than an Earth diameter if one of the observatories is in space. The uses of VLBI currently fall into two major categories. The astrophysical uses involve learning about the objects studied because of interest in their physics, the structures, changes, and motions of compact components in natural radio sources. The other major use for VLBI is astrometric and geodetic in nature; it involves the study of sources not for their own sake, but for their use as reference beacons in studies of Earth orientation and plate tectonics as well as spacecraft navigation. The two categories of VLBI investigations are complementary in the sense that improved knowledge of the properties of natural radio sources can be used to better understand and utilize their performance as celestial beacons.

The natural radio sources used as beacons are very distant, extragalactic objects. For the purposes of this article, they will often be referred to by the generic name of *quasars*, although they may also include distant radio galaxies, BL Lacertae objects, and some sources whose specific optical counterparts have not yet been identified. The quasars are so distant that they can be used to define an inertial reference frame. Spacecraft motions within the solar system and motions on the surface of the Earth can be found with respect to that reference frame.

Most astrometric VLBI work has proceeded under the assumption that the quasars are point radio sources with positions that are constant in time and do not shift as observation parameters change. Although it has long been known that this is only an approximation, it has been an appropriate assumption as long as the errors it introduces have been significantly smaller than the errors introduced by other aspects of the VLBI observations. Some other error sources in astrometric VLBI are imperfect knowledge of Earth orientation parameters, station location uncertainties, the static and fluctuating troposphere, VLBI instrumentation, and tidal effects. More detailed analyses or expected levels of these and other errors can be found in [1-3] and will not be repeated here. VLBI has now reached the stage where the above errors can be calibrated or modeled well enough so that the residual errors on interferometric delay measurements on intercontinental baselines are in the  $\leq 100$  psec range.

With the use of VLBI systems that enable high sampling rates and large spanned bandwidths (e.g., the Mark III system, [4]), expected delay errors using DSN antennas are considerably less than 100 psec. On intercontinental

baselines, errors of approximately 150 psec of delay correspond to position uncertainties of about 1 milliarcsec [1]. However, many of the quasars used for astrometric and geodetic VLBI show structure on scales of several milliarcseconds. This structure is known to change with time, and its effects on VLBI observables also depend on the baseline length and orientation in a given observation. Thus, it has been thought that delay errors caused by source structure are at roughly the 100 psec level on intercontinental baselines. This is particularly true in the case of VLBI catalogs that are used for spacecraft navigation. Because of the need for a large number of radio sources distributed around the sky with a high density of sources in the ecliptic plane, it has not been possible to be selective enough to eliminate all the radio sources in which structure may have an effect of  $\sim 100$  psec.

The purpose of the work described in this article is to use the intensity distributions of real radio sources to estimate the magnitude of the error that is made by observing these resolved radio sources. Maps of the structure of a number of compact radio sources are used to study their effects on source position estimates. Rough error estimates for several different quantities are made and their dependence on several parameters is also studied.

## II. Important Observed Quantities in Astrometric VLBI

Observed quantities in astrometric VLBI include group delay (or bandwidth synthesis delay), phase delay, delay rate, and correlated amplitude [1]. The delay is defined as the difference between the arrival times of a radio signal at two stations. In the case of quasars, the radio signal is a noise-like, broad-band signal rather than a narrow-band tone or set of tones as is the case for a spacecraft. The component of delay that is sought for understanding quasar and observing-station locations is the *geometric delay*, which is the delay caused by the combination of the geometry of the baseline and the direction to the radio source. Various effects mentioned in Section I can cause errors in the determination of the geometric delay.

Most of the important observed quantities have been defined and explained in [5], and the reader is referred to that article for more detail than given below. Traditionally, the *group delay* has been used and is defined as

$$\tau_g = \frac{\partial \phi}{\partial \omega} \quad (1)$$

$$V = \frac{S_c}{S_t} \quad (5)$$

Here,  $\phi$  is the relative phase of the radiation between the two stations and  $\omega$  is the angular frequency ( $\omega = 2\pi\nu$ , where  $\nu$  is the sky frequency in units of Hz). Although the group delay is formally defined as a derivative, it is actually determined as a differential over a small frequency range:

$$\tau_g \approx \frac{\Delta \phi}{\Delta \omega} \quad (2)$$

A second quantity, not traditionally used in astrometric VLBI, is the *phase delay*, which is defined by the formula

$$\tau_\phi = \frac{\phi}{\omega} \quad (3)$$

Phase delay is generally considered to have the potential for increased accuracy relative to group delay because of the "leverage" that is provided by having the total sky frequency in the denominator. However, phase delay traditionally has not been used because of the difficulty of making the correct a priori resolution of cycle ambiguities [5].

The *delay rate* is a third important astrometric quantity; it is currently used along with group delay in most astrometric VLBI. This quantity is defined as the time rate of change of the phase delay:

$$\dot{\tau}_\phi = \frac{\partial \tau_\phi}{\partial t} = \frac{1}{\omega} \frac{\partial \phi}{\partial t} \quad (4)$$

The delay rate will be used for VLBI tracking of the Magellan spacecraft.

A fourth quantity that is often ignored in astrometric VLBI will be introduced here, because it can provide important clues to the appearance of errors caused by radio-source structure. That quantity is the *correlated amplitude* of the radio source signal. The correlated amplitude (or visibility amplitude) is defined and discussed extensively in many works such as [1]. It may be expressed in units of flux density, as a correlation coefficient, or as an amplitude relative to the maximum possible. In this article, the visibility amplitude will be used, and is defined by

Here,  $S_c$  is the correlated amplitude (or correlated flux) and  $S_t$  is the total amplitude (or total flux).

For a point radio source, the visibility amplitude has a maximum value of one. When radio emission comes from an angular area whose dimensions are comparable to  $\lambda/D$ , where  $\lambda$  is the observing wavelength and  $D$  is the baseline length, the radio source is said to be "resolved." That is, it shows interference properties different from those exhibited by a point radio source. The emission from different directions causes a "washing out" of the interference pattern, because the cross-correlations of radiation from different parts of the source interfere destructively for the two observing stations. Then the visibility amplitude (or fringe visibility) is reduced. Since the amount of resolution changes as Earth rotation alters the baseline orientation, the visibility amplitude for a specific radio source varies with time on a given baseline. This amplitude is an important indicator of the magnitude of the source-structure effect on measured delays and delay rates, as will be described further below.

### III. Review of Source-Structure Effects on Astrometric VLBI Observations

Extensive analysis has been done to determine how the non-pointlike structure of radio sources affects the quantities observed in astrometric VLBI [6, 7], and will not be repeated here. Typically, the structure-induced group-delay error is largest for a given quasar when the visibility amplitude is lowest, or when the interferometer baseline "heavily resolves" the radio source.

In order to examine trends in the structure-induced delays, it is useful to introduce the concepts of the  $(u, v)$  plane and the interferometer hour angle (IHA). (Again, see [1].) For a single telescope, the hour angle is the east-west angle of the source relative to the local meridian. When a source transits the local meridian moving from east to west, its hour angle is said to be zero. One hour earlier, the hour angle was  $-15$  deg, or  $-1$  hour; one hour after transit, the hour angle is  $+15$  deg, or  $+1$  hour. The IHA for two separated telescopes is defined analogously, with the "local" meridian being defined as the meridian at the midpoint in longitude between the two observing stations.



The  $(u, v)$  plane is the plane perpendicular to the direction from the Earth's center to the radio source, with  $u$  increasing to the east and  $v$  increasing to the north. As the Earth rotates, the projected interferometer baseline in the  $(u, v)$  plane changes, which changes the interferometer response to any extended structure in the radio source. The evolution of the  $(u, v)$  coordinates of the baseline is calculable solely from the geometry of the locations of the observing stations and the celestial coordinates of the radio source. These coordinates are normally measured in units of wavelengths of the observing frequency, so that  $1/\sqrt{u^2 + v^2}$  gives the number of wavelengths in an interference fringe, which is the approximate spatial resolution (in radians) of the interferometer. Figures 1 and 2 show plots of the  $(u, v)$  points sampled for two different sources, one on the California–Spain baseline and one on the California–Australia baseline.

The structure phase delay tends to be constant when the structure group delay is small, in regions of the  $(u, v)$  plane where the source visibility amplitude is high. However, this phase delay is not necessarily the same as it is for other projections of the same baseline that also have high visibility. If the  $(u, v)$  track for a given source and station pair passes through a low-visibility region between the two high-visibility regions, the phase delay will change substantially in the low visibility region. This means that the structure-induced delay rate should also be large in regions of low visibility. Examples of such behavior will be discussed in Section V.

#### IV. Calculations of Source-Structure Errors for Real Radio Sources

The investigation of source structure using VLBI has made substantial progress in the last 10 years, as improved calibration, mapping, and modeling procedures have been developed. It has become possible for astrophysicists to map large numbers of compact radio sources for a variety of purposes. One such investigation that was begun more than 10 years ago is a 5-GHz mapping survey of strong extragalactic sources that has been carried out by the Caltech VLBI group [8, 9]. This survey is intended to make maps of approximately 45 sources at several epochs to study their overall properties and their structural changes with time. First-epoch models of 26 sources were kindly made available by T. Pearson to enable the author to study

the statistics of the structure effects on measured group delay and phase delay. VLBI maps of many of these objects are shown in [9].

The available models of compact VLBI sources are derived by measuring the complex source visibilities at a large number of  $(u, v)$  points, Fourier-transforming the data to make maps of total intensity versus position in the sky, then deconvolving these maps into superpositions of point sources. The fluxes and positions of a set of such point sources form a model of a single complex radio source. More detailed information on mapping and modeling procedures is available in [10].

For each of the 26 sources studied in this work, the  $(u, v)$  tracks on the Deep Space Network baselines between California (Goldstone) and Spain (Madrid) and/or California and Australia (Tidbinbilla) were computed. (Since the survey sources are fairly far north, they are all visible on the baseline to Spain, but many are not visible on the baseline to Australia.) The  $(u, v)$  tracks were sampled every 3 min, which is roughly the duration of a typical astrometric VLBI observation of an individual source. The source models were then used to compute the group delay, phase delay, and delay rate caused by structure at each sampled point, using the N-point-source formalism described in [6]. All quantities were computed relative to the flux-density centroid (“center of gravity” of the source flux) of each individual source, and computations were made only at 8.4 GHz.

Most astrometric VLBI makes use of observations at both 2.3 GHz and 8.4 GHz, but corrections for charged-particle propagation lead to the 8.4-GHz data having approximately 13 times greater weight than the 2.3-GHz data. Flux-density centroids at the two frequencies may differ by as much as 5 nrad, leading to a possible error of up to about 10 psec in the charged-particle correction to the delays. The interference phenomena will also differ significantly, with larger source sizes at 2.3 GHz being partly offset by the poorer spatial resolution at the lower frequency (see below). Again, this may be absorbed in the charged-particle correction, and will probably cause an error of no more than about 30 psec of delay in the worst cases. As seen from the results presented in Section V, this means that most of the effect of source structure can be found by doing studies at 8.4 GHz alone.

A number of additional assumptions were made in computing the effects caused by structure. First, it was

assumed that the source structures at 8.4 GHz were the same as those at 5 GHz. It is well known that this is not true, since there are substantial variations in spectral index in different parts of VLBI sources. Typically, there is one radio component that is identified as the "core" of the quasar by virtue of the fact that its spectrum is "flat" or "inverted" (i.e., the component strength at 8.4 GHz is greater than or equal to its strength at 5 GHz). The rest of the VLBI source typically shows a "steep" (flux falls off with increasing frequency) spectrum. Based on 5-GHz maps alone, it is not always possible to identify the core component. Most sources also do not have detailed spectral information available, so it is not valid to attempt to correct for the source spectrum in making calculations at 8.4 GHz.

Another important approximation arises in computing source structure effects along  $(u, v)$  tracks that were not actually sampled in the VLBI observations used in making the original source maps. Use of the models derived from baselines that are generally shorter than the DSN intercontinental baselines amounts to making an extrapolation in the  $(u, v)$  plane to determine the source visibility. Since a baseline between Owens Valley, California and Bonn, Germany was used in making most of the maps, the extrapolation is not extremely large in making calculations on the Goldstone-Madrid baseline. However, a substantial extrapolation is involved in making calculations on the Goldstone-Tidbinbilla baseline.

Finally, there is an approximation in the use of the derivative rather than the differential in computation of the group delay. For bandwidth synthesis observations using the Mark II VLBI system and the DSN, the maximum spanned bandwidth is typically 40 MHz. This is less than 0.5 percent of the 8.4-GHz observing frequency, so approximating the differential as a derivative is fairly good. However, the spanned bandwidth of  $\sim 400$  MHz typically used in the Mark III VLBI system approaches 5 percent of the observing frequency. Then, using the derivative at a single point is a less accurate approximation. Although the group delay may be very large at a single frequency, it can be significantly smaller at a frequency 300 MHz away. Then, the use of the derivative instead of the differential would tend to overestimate the effect due to structure. For phase delay, the difference between the smaller and larger spanned bandwidths is unimportant.

The data do not exist to enable us to make a quantitative study of the errors introduced by the above ap-

proximations. Qualitatively, the use of 5-GHz maps for calculations made at 8.4 GHz will probably lead to an overestimate of the group delay caused by structure. Since sources tend to be more dominated by a single component at 8.4 GHz than they are at 5 GHz, they show less of an effect due to the steep-spectrum extended emission. The neglect of the effects at 2.3 GHz would cause an underestimate of the structure effects, but the low weighting of this frequency in the charged-particle correction means that this underestimate is of relatively minor importance. The extrapolation in the  $(u, v)$  plane means that some components that might be resolved on long baselines at 8.4 GHz are not separated in the models used, leading to a possible underestimate of the effects due to structure. Finally, as mentioned already, the approximation of a differential as a derivative is likely to lead to an overestimate of the group-delay error in some cases.

It is clear that the results of the calculations presented below can be used only as statistical indicators of the magnitude of the source-structure errors and cannot be used to make detailed statements about specific cases. Overall, the magnitude of the estimated errors is probably within a factor of 1.5–2 of the "real" errors induced by structure. This error estimate is derived by procedures such as using known spectra of parts of sources that have been mapped at several frequencies to convert from their structures at 5 GHz to those estimated structures at 8.4 GHz, and then recomputing the structure-induced delay errors.

## V. Statistical Results

### A. Some General Considerations

A variety of statistical quantities can be computed from the source structure effects found for the individual sources in the sample. The relevance of each quantity depends on the specific circumstances under which a source is observed and the way the VLBI data are analyzed. For example, if a single observation is made of a source for which there is no a priori information about structure, the scatter about zero structure effect ("root-mean-square about zero") will give the typical deviation of quantities from those measured at a nominal position given by the source centroid. But if a large number of measurements of a source have been made at a variety of  $(u, v)$  points, the a priori source position may be offset from its flux centroid by an amount governed by the average structure

effects in the previous observations. Then, the traditionally computed root-mean-square would give the scatter on top of an average bias that has been built into the a priori values. Of course, since structure effects may be absorbed by other parameters in the multi-parameter fit that gives source positions, the a priori position will not be exactly that given by the centroid plus the mean structure effect. Further, the offset from the centroid would depend on which astrometric quantities were used in deriving the a priori position.

A further consideration arises from the fact that the largest group-delay and delay-rate effects occur at points with low visibility amplitudes. However, some of these points will not exist in the final data set, because the low amplitudes may mean that interference fringes are not detected, and no valid measurement is acquired. For example, a source with a total flux density of 0.8 Jy and a visibility amplitude of 0.1 is probably too weak to show fringes in a Mark II VLBI experiment using a 34-m and a 70-m antenna. If weak fringes are detected, the low-amplitude points will be lightly weighted because of their low signal-to-noise ratio. The overall reduced importance of the low-visibility points, where the largest offsets from the flux centroid would appear, makes it likely that the a priori position of a source will be very close to its flux centroid. Therefore, the scatter about zero structure effect should be the most meaningful quantity in estimating expected errors. Although scatters about both zero and about the mean have been computed for the source-structure effects, the scatter about zero will be the quantity quoted throughout the remainder of this section. The reader should keep in mind the fact that the structure errors are not Gaussian-distributed, so the scatters quoted below should not be thought of as the parameters of a standard error distribution.

## B. General Results for Deep Space Network Baselines

Table 1 lists the 26 sources that were used in making calculations of the errors caused by source structure. This table includes a classification of the type of structure each source contains, roughly corresponding to those given and described in [8, 9]. Note that these 26 sources fall into 5 different morphological classes; those divisions will be used below in determining which types of sources cause the greatest errors.

Figures 3–18 give samples of the computed structure effects along the  $(u, v)$  tracks on DSN baselines for several different sources. Each figure shows the visibility amplitude and structure-induced delays and rates for a source, as functions of IHA. Note, as expected, that the largest group-delay and delay-rate errors occur near visibility minima, while the phase delay changes rapidly at these points and thus is different on opposite sides of the visibility minima. As mentioned previously, recall that all quantities are computed relative to the flux centroid of each source.

Figures 3–6 show results for a double radio source on the baseline between California and Australia. This source is visible on the given baseline for approximately 4 hours. There are two regions where the visibility amplitude is below 0.05; the structure-induced group-delay and delay-rate errors at this point are impressively large. The phase-delay error has a large derivative at these points, but a large phase-delay error due to structure is not accumulated over the 4 hours of tracking.

Figures 7–10 show results for a circumpolar, double source on the baseline between California and Spain. The separation of the main components in this source is more than 50 nrad, which is many times larger than the resolution of the baseline. There are a large number of minima in the plot of visibility amplitude, and each corresponds to a spike in group delay and in delay rate. The phase delay changes substantially at each minimum and accumulates at the rate of approximately 100 psec per hour of observing, with a maximum value of over 700 psec caused by the source structure.

Figures 11–14 show results for a circumpolar asymmetric source that is fairly well dominated by its unresolved core. The visibility amplitude never drops below about 0.4. The corresponding group-delay and phase-delay errors are always below 100 psec, in contrast to the results shown for the relatively large double source in Figs. 7–10.

Figures 15–18 give results for another asymmetric source on the baseline between California and Australia. The only points below 0.1 visibility amplitude occur between 3.2 and 3.3 hours IHA, with other points below 0.2 occurring about an hour later. The group-delay and delay-rate errors are maximum at these points. But, it should be noted that the dependence of the group delay on the visibility amplitude is dramatic at these local visibility minima. For a 33 percent decrease in visibility amplitude,

the structure-induced group delay grows by a factor of almost 3. The delay-rate error shows a much more modest increase as the visibility amplitude goes down.

Table 2 shows the structure-induced scatter (about zero) in astrometric VLBI quantities on the California-Spain baseline for all 26 sources as a function of visibility amplitude. Table 3 shows similar results for the 12 sources visible on the California-Australia baseline. (Note that there are many fewer  $(u, v)$  points used on the latter baseline, so the conclusions are based primarily on the results for the California-Spain baseline.) As expected (see [6]), there is a strong dependence on visibility amplitude, with the lowest amplitude points giving the greatest errors in group delay and delay rate. The group-delay results are roughly comparable between the two baselines, but the phase-delay and delay-rate errors are much higher on the baseline to Spain. This difference can be explained by the long  $(u, v)$  tracks sampled on this baseline for circumpolar sources. Returning to Figs. 1 and 2, the plots show that a much wider variety of  $(u, v)$  points was sampled for the circumpolar source on the Spain baseline than for the other source on the Australia baseline. Since a larger variety of baseline projections was sampled, the phase contributed by source structure can wind through a larger number of cycles, giving a greater deviation from zero phase. Such an effect can be seen by examination of Figs. 7-10.

As a check on the effect caused by circumpolar sources, one can divide the Spain data according to the source declinations. There are 9 sources that are circumpolar at both antennas, having declinations above 60 deg, and 17 sources at lower declinations. Approximately 4000  $(u, v)$  points are sampled for each set of sources. Table 4 gives the scatters about zero for the circumpolar and "other" sources, and shows a striking result. Although the typical group-delay errors are actually somewhat larger for the more southerly sources, the phase-delay scatter is a factor of about 4 higher for the circumpolar objects, and the delay-rate scatter is a factor of 2 higher. This implies that a catalog that is constructed using phase-delay measurements between northern hemisphere antennas may have much more structure-induced scatter in the phase delay if a significant number of circumpolar sources are observed at a wide range of hour angles. This scatter can be reduced if such sources are only observed over a narrow range of hour angles, so that the phase does not wind through a large number of cycles. However, since the observations at a wide range of hour angles can be useful in isolating

the effects of parameters other than source structure, the costs and benefits of such a limitation are unclear. Another possibility would be to derive different source positions for different ranges of hour angle, although the individual positions would be of lower accuracy because of the reduced number of data points included for each.

Tables 5 and 6 show the effect of arbitrary data cut-offs based on source visibility amplitudes. Again, both California-Spain (Table 5) and California-Australia (Table 6) data are shown. These tables give the scatter for all those points above a specific value of the visibility amplitude. It is seen that cutting out the lower-visibility points dramatically reduces the group-delay error caused by structure, significantly reduces the delay-rate error, but has much less effect on the phase delay. This confirms the expectations that group-delay and delay-rate errors are highest at the low-visibility points, whereas phase-delay errors are less correlated with visibility amplitude.

Table 7 shows the scatters for the total set of 26 maps and for each class of sources. ("Good" and "bad" sources are defined below.) Only the California-Spain baseline is shown, as there are not enough objects visible on the California-Australia baseline to have a sufficient number of sources in each class. Here, it is important to note that there are certain classes of radio-source structures for which the errors caused by structure are substantially larger than for other classes of sources. In particular, the "compact-double" (here, "compact" typically means 20-100 nrad) and "steep-spectrum-core" sources show much larger structure errors than the other types of sources. Group-delay errors for the "unclassified" sources are large solely because of the inclusion of 3C 84, which is well known to have significant structure on a scale of about 100 nrad.

Most sources whose structure would cause delay errors of hundreds of psec are not included in VLBI catalogs [11, 12] that are used for astrometry and geodesy. Those for which observations have been attempted either have not shown strong (or any) interference fringes on the intercontinental baselines because of insufficient compactness, or have given large residuals in analyses of VLBI data. In either case, such sources would not remain in catalogs. For example, the compact double source 2021 + 614 (OW 637) was observed in early JPL VLBI experiments; because of its structure, however, it is no longer a prime observing candidate.

The compact-double sources, steep-spectrum-core sources, and 3C 84 can be defined as "bad" sources from an astrometric standpoint. The basic definition of "bad" can be made solely from their radio morphologies, without any actual calculations of the structure effects. Eliminating them from those listed in Table 1, we are left with a set of 16 sources that are probably more representative of those that have survived a sieving process for inclusion in astrometric catalogs. These are the sources defined to be "good" astrometric sources. Indeed, of these 16 sources, 8 of them are in the latest published JPL VLBI catalog [12]. Analysis of the data for the 16 "good" sources alone gives scatter in the delays and rates that is a factor of 3-7 less than for the entire set of 26 sources, as shown at the bottom of Table 7. These numbers are probably much more representative of the structure errors that may be present in the JPL VLBI catalog.

Of the 10 "bad" sources eliminated, only 3C 84 is in the latest JPL VLBI catalog. It is only present because, despite its very low visibility amplitude, the source flux is so high that the correlated flux is still detectable even at points of very low visibility. For observers on Earth, there are no other sources comparable to 3C 84 in strength and structure, so its properties cannot be used to generalize to other sources that might be in VLBI catalogs.

Tables 8 through 12 show the same results as in Tables 2 through 6, but with only the "good" sources included. The same trends with visibility and source declination are indicated, although they are somewhat less marked than for all 26 objects because of the overall reduction in the scatter for these sources. A comparison of Tables 2-3 and 8-9 shows that the good sources have many fewer low-visibility points, as would be expected. But even at the low-visibility points, the structure effects are much less severe than for the "bad" sources at points with similar visibility amplitudes. This is because the "bad" sources are generally many resolution elements in size; as pointed out in [6], for points with the same visibility amplitude, larger sources give more pronounced structure effects than do smaller sources.

Comparison of Tables 4 and 10 shows that the differences in phase delay and delay rate between circumpolar and more southerly sources is less pronounced for the "good" sources than for all sources. This is probably because the "good" sources are not as many resolution elements in size. Therefore, as the baseline rotates under

the radio sources, the phase does not wind nearly as much for the smaller sources. Figures 11-14 show results for a "good" circumpolar source and illustrate the relatively small accumulation of phase-delay error in such a case.

The results given in Table 5 for all sources and in Table 11 for "good" sources are quite different for the low visibility cutoffs, but converge to similar values in group delay and delay rate at the higher visibility cutoffs. This convergence comes about because there are few points of high visibility in the "bad" sources, so the data largely overlap. Despite the overlap, there is still a large difference in phase-delay results, even at visibilities of 0.6 and higher. This is due to the fact that the phase delay at a given  $(u, v)$  point depends on the phase-winding at all other points for the same source. Even for a high-visibility point in a "bad" source, the connected phase may wind through several cycles at visibility minima, so the residual phase delay at the high-visibility point can be several times a single X-band cycle ambiguity of 120 psec of delay. In Figs. 7-10, note that there are a number of  $(u, v)$  points with visibility amplitudes near 0.6 where large phase-delay errors have accumulated. It does not take many such points to raise the phase-delay scatter at moderate visibility amplitudes from the approximately 30-60 psec for "good" sources to the values of more than 100 psec found for the "bad" sources.

### C. Dependence of Structure Errors on Baseline Length

Since astrometry and spacecraft navigation can be done on baselines shorter than the DSN intercontinental baselines, it is of interest to estimate the delay errors caused by structure as a function of baseline length. For the 16 "good" sources, calculations have been made on the baseline between Haystack Observatory in Massachusetts and Fort Davis, Texas, as well as on the baseline between Haystack and Green Bank, West Virginia. The lengths of these baselines are 3140 km and 840 km, compared to the 8390 km of the Goldstone-Madrid baseline. These shorter baselines were actually used in making almost all the maps in the sample. Of course, the observing frequency was 5 GHz rather than 8.4 GHz, so the actual  $(u, v)$  points used in the computations were not sampled.

Figures 19-22 show the results for a single source on the baseline between Haystack and Fort Davis. Note that the visibility amplitude is considerably higher than for the

same source on the California–Australia baseline (Figs. 15–18). Figures 23–26 show a comparison for the two baselines with the same vertical scales for the errors. As expected, the typical group-delay, phase-delay, and delay-rate contributions due to the structure are considerably lower on the shorter baseline, as are the maximum excursions from zero.

Table 13 shows a comparison among the three different baselines. (Preliminary results were presented previously in [13].) The group-delay scatter caused by structure ranges from 5 psec on the shortest baseline to 83 psec on the longest baseline, with the corresponding phase-delay scatter being roughly comparable in size. Power-law fits of scatter versus baseline length give exponents of approximately 1.2 for the delays and rates over a factor of 10 in baseline length. However, the slope of the relation appears to steepen at the shorter baseline lengths; this is probably because the “good” sources studied here are all poorly resolved by baselines with lengths of about 1000 km. Table 14 shows that the slope change for such baselines is much less pronounced when all 26 sources are considered, because some of these sources are large enough to remain well resolved even on the baseline between Haystack and Green Bank. It is not appropriate to use the available source maps to investigate effects at even shorter baselines. Since shorter baselines were not used in acquiring the VLBI data that went into the models studied in this article, the observations were insensitive to the larger scale structure that is known to be present in many sources.

The estimated scaling with baseline length is not always appropriate for individual sources, but has a complicated dependence on the source structure and the interferometer baselines. Scaling to shorter baselines certainly would not work for all objects. For instance, the compact-double sources are small enough that they would be mostly unresolved by baselines of on the order of 100 km, so the structure errors in group delay would probably be comparable to those of the sources included in Table 13. Since the errors are much larger on the intercontinental baselines in the double sources, the scaling with baseline length would be different. This points out the fact that any time baseline lengths are changed by a large amount (say, a factor of about 2), different sources and classes of sources may be appropriate astrometric targets. More sources can be used on shorter baselines than on the intercontinental baselines of the DSN. Consideration of the properties of these additional sources would be necessary to determine

the magnitude of structure errors that might be expected, for instance, on the 253-km baseline between Goldstone and Owens Valley, California.

## VI. Discussion and Conclusions

A statistical investigation of the effects of compact-radio-source structure on astrometric measurements has been carried out using models of 26 different radio sources. The calculations show that for the entire sample of sources, with observations made on intercontinental baselines, the structure-induced scatters in group delay and in phase delay amount to hundreds of psec. These variations with interferometer hour angle correspond to several milliarcsec of position error. However, when sources that are unlikely to appear in astrometric catalogs are eliminated, the typical delay errors caused by structure for individual observations are reduced to between 50 and 100 psec, while delay-rate scatter is approximately  $5 \times 10^{-15}$  sec sec<sup>-1</sup>. This is somewhat less than the residuals in typical intercontinental Mark II VLBI experiments performed with DSN antennas, but is larger than the residuals for Mark III experiments. Such errors correspond to position errors of about 0.5–1 milliarcsec, comparable to the current accuracy of the best positions in the JPL VLBI catalog. In our current VLBI data, source structure is sometimes obvious in the correlated amplitudes, but is not readily apparent in delay measurements. Signs of structure-induced delay residuals have been seen in astrometric Mark III VLBI data taken by others [14, 15]<sup>1</sup> and efforts are underway to identify such residuals in DSN data.

It is interesting to note that quasar structure is unimportant for the Magellan mission to Venus. The delay-rate scatter of  $6.4 \times 10^{-15}$  sec sec<sup>-1</sup> (or 0.002 mm sec<sup>-1</sup>) shown for the California–Spain baseline in Table 13 is 1–2 orders of magnitude lower than the expected error in Magellan  $\Delta$ VLBI measurements.<sup>2</sup> The error for Magellan is dominated by the fluctuating troposphere, but there are other sources of error that are well above the expectations for source structure.

<sup>1</sup> Also see J. S. Ulvestad, “Possible source structure effects in IRIS data,” JPL Interoffice Memorandum 335.3-88-15 (internal document), Jet Propulsion Laboratory, Pasadena, California, February 3, 1988.

<sup>2</sup> J. S. Border, “An Error Analysis for Magellan Differential Delay Rate Measurements,” JPL Engineering Memorandum 335-96 (internal document), Jet Propulsion Laboratory, Pasadena, California, February 23, 1987.

The magnitude of structure-induced delay errors as a function of baseline length has also been investigated. The errors appear to increase slightly more than linearly with baseline length for the compact sources studied here. However, it must be recognized that different sources have structure on a variety of scales. Therefore, the radio sources selected to have sufficient correlated amplitude to be seen on intercontinental baselines will tend to be poorly resolved on shorter baselines, leading to the reduced structure effects. The dependence on baseline length may be different for sources that are not preselected to be highly compact.

Several possible methods could be used to minimize source-structure errors in VLBI astrometry. The ideal approach for construction of a catalog or for individual observations of a spacecraft relative to a reference quasar would be to select only sources that show evidence of having compact structures that give very little error. As mentioned previously, the JPL VLBI catalog already consists almost entirely of sources that would be in the "good" category, having typical structure errors below 100 psec of delay. A further reduction in error could be achieved by constructing a VLBI catalog using the sensitive Mark III VLBI system, because there are many more detectable sources from which to choose those with minimal structure. However, such a catalog might not be of much use in spacecraft navigation using a narrow-bandwidth VLBI system such as the current 250-kHz-bandwidth Block I system.<sup>3</sup> This is because there are only a limited number of natural radio sources that can be detected with the narrow-bandwidth system. In any case, source structure is not much of an issue with a relatively insensitive system, both because the system-noise errors are higher than the source-structure errors and because the lowest amplitude points will be weeded out due to the lack of source detection in the cross-correlation.

The newest JPL VLBI catalog contains only about 20 sources with correlated fluxes on DSN intercontinental baselines that are above 1 Jy at both 2.3 GHz and 8.4 GHz.<sup>4</sup> For spacecraft navigation with the Block I sys-

<sup>3</sup> J. B. Thomas, "An Error Analysis for Galileo Angular Position Measurements with the Block I  $\Delta$ DOR System," JPL Engineering Memorandum 335-26 (internal document), Jet Propulsion Laboratory, Pasadena, California, November 1981.

<sup>4</sup> J. S. Ulvestad and O. J. Sovers, "Preliminary VLBI Catalog for Magellan," JPL Interoffice Memorandum 335.3-89-14 (internal document), Jet Propulsion Laboratory, Pasadena, California, January 30, 1989.

tem, these sources are the only ones that can be detected on DSN baselines with two 34-m antennas. This is already too small a number of sources. Reduction of this sample size in order to minimize errors due to source structure would make the source density unacceptably small. Then, large errors could be caused by the reference quasars being angularly far from the spacecraft being navigated. Of course, the problem of a low density of quasars could be reduced by using an increased bandwidth in the VLBI system used for navigation.

Another method of reducing structure effects in building a VLBI catalog is a selection based on visibility amplitude. As expected, the group-delay and delay-rate errors are well correlated with visibility amplitude, while the phase-delay errors are less strongly related to visibility. This confirms the supposition that the removal of low-visibility data can reduce source-structure effects in current astrometric catalogs, where group delay and delay rate are the observed quantities that are used. However, in experiments where phase connection is achieved and the potentially more accurate phase delay is used, such a simple procedure will not reduce the structure effects greatly. Table 11 shows that a cutoff at a visibility amplitude of 0.3 would reduce the group-delay and delay-rate scatters by about 30 percent on the Goldstone-Madrid baseline at the expense of about 14 percent of the data points going into a catalog. However, such a cutoff would have no effect on the phase-delay scatter. It should be noted that any such selection of individual data points, or a selection of sources as described previously, would require careful amplitude calibration of the VLBI data. Such calibration generally is not carried out as a routine part of the gathering of astrometric VLBI data. It would require careful design of a system to monitor system temperatures both on and off source as well as additional steps in the data analysis process.

Another possible method of reducing source-structure errors would be to model the sources being observed and use those models to compute corrections to the data. This is a difficult approach to implement, as it relies on having good models of the *current* structures of the radio sources at DSN observing frequencies. Such models can only be derived by making observations of each source at many different hour angles on many baselines, and then going through a time-consuming data-reduction process. The observations would need to be repeated on a regular basis at intervals of much less than a year. The antenna

time is not available for making such observations of many sources. Attempting to include simple model types in a parameter-estimation scheme may also be doomed to failure, since the source-structure errors often depend on fine details of the radio morphologies that would be difficult to include in a parameter-estimation algorithm.

Yet another method of reducing source-structure errors in individual VLBI measurements for spacecraft navigation would be to observe a number of sources in a local reference frame, as described in [16]. If all reference-source (and spacecraft) observations on different dates were made at the same hour angle, source structure would cause *no* change in the derived spacecraft position between those dates, as long as the source structures did not change between observations. The assumption that structures do not change is probably good on time scales of a few months [17], but will break down on longer time scales. In some sources, the relative strengths of components will change more rapidly than their relative positions change, and this can greatly affect the magnitude of the structure effects.

If the sources in a local reference frame could be thought of as unmoving beacons, spacecraft motion could be determined relative to those beacons. Of course, for absolute spacecraft position to be determined in a single observation, the positions of the reference sources would have to be determined relative to a VLBI catalog. If the observations used in constructing that catalog were made at a variety of hour angles, the scatter caused by source structure in the catalog observations would cause an error in the determination of the absolute spacecraft position.

In summary, the source-structure errors may be an important effect in astrometric Mark III VLBI data taken with the DSN. With special effort, these errors could be reduced somewhat. However, with currently practical methods of reducing these errors, it is difficult to see how they could be made much smaller than about 30–50 psec of delay on individual observations of single sources. This would correspond to 1–2 nrad of position error (or 0.2–0.4 milliarcsec) on intercontinental baselines if the error mapped entirely into source position. Delay-rate scatter is likely to stay at least as large as  $2 \times 10^{-15}$  sec sec<sup>-1</sup>, giving a similar position error. The local-reference-frame method of determining spacecraft position has the potential for reducing the scatter significantly if observations are scheduled carefully, although this method may not reduce the error in tying such observations to a global VLBI catalog.

As mentioned previously, the errors caused by source structure may not map solely into source-position errors, but may be absorbed by other parameters in a multi-parameter fit. For example, the long periods of apparent phase-delay drift shown in Fig. 9 might map into drift of the station clocks. An important next step for any analysis would be to understand how structure effects could affect a variety of parameters in a large fit. As part of that effort, it would be important to isolate data that should be subject to large structure effects in the JPL VLBI data set; such points could then be eliminated in an effort to see how they have affected different parameters.

## Acknowledgment

I thank Brooks Thomas, Bob Treuhaft, Roger Linfield, and Chad Edwards for useful discussions and suggestions. I am also very grateful to Tim Pearson for supplying the source models used in this work.



## References

- [1] A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr., *Interferometry and Synthesis in Radio Astronomy*, New York: John Wiley and Sons, 1986.
- [2] O. J. Sovers, J. B. Thomas, J. L. Fanelow, E. J. Cohen, G. H. Purcell, Jr., D. H. Rogstad, L. J. Skjerve, and D. J. Spitzmesser, "Radio Interferometric Determination of Intercontinental Baselines and Earth Orientation Utilizing Deep Space Network Antennas: 1971 to 1980," *Journal of Geophysical Research*, vol. 89, pp. 7597-7607, 1984.
- [3] R. N. Treuhaft and G. E. Lanyi, "The Effect of the Dynamic Wet Troposphere on Radio Interferometric Measurements," *Radio Science*, vol. 22, pp. 251-265, 1987.
- [4] A. E. E. Rogers et al., "Very Long Baseline Radio Interferometry: The Mark III System for Geodesy, Astrometry, and Aperture Synthesis," *Science*, vol. 219, pp. 51-54, 1983.
- [5] C. D. Edwards, "Short Baseline Phase Delay Interferometry," *TDA Progress Report 42-91*, vol. July-September 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 46-56, November 15, 1987.
- [6] J. B. Thomas, *An Analysis of Source Structure Effects in Radio Interferometry Measurements*, JPL Publication 80-84, Jet Propulsion Laboratory, Pasadena, California, December 15, 1980.
- [7] W. D. Cotton, "Source Structure Corrections to the Geodetic Very Long Baseline Interferometry Observables," *Radio Interferometry Techniques for Geodesy*, NASA Conference Publication 2115, pp. 193-197, 1980.
- [8] T. J. Pearson and A. C. S. Readhead, "VLBI Survey of a Complete Sample of Active Nuclei and Quasars," *International Astronomical Union Symposium No. 110, VLBI and Compact Radio Sources*, edited by R. Fanti, K. Kellermann, and G. Setti, Dordrecht: D. Reidel Publishing Company, pp. 15-24, 1984.
- [9] T. J. Pearson and A. C. S. Readhead, "The Milliarcsecond Structure of a Complete Sample of Radio Sources. II. First-epoch Maps at 5 GHz," *Astrophysical Journal*, vol. 328, pp. 114-142, 1988.
- [10] T. J. Pearson and A. C. S. Readhead, "Image Formation by Self-Calibration in Radio Astronomy," *Annual Review of Astronomy and Astrophysics*, vol. 22, pp. 97-130, 1984.
- [11] C. Ma et al., "Radio Source Positions from VLBI," *Astronomical Journal*, vol. 92, pp. 1020-1029, 1986.
- [12] O. J. Sovers, C. D. Edwards, C. S. Jacobs, G. E. Lanyi, K. M. Liewer, and R. N. Treuhaft, "Astrometric Results of 1978-1985 Deep Space Network Radio Interferometry: The JPL 1987-1 Extragalactic Source Catalog," *Astronomical Journal*, vol. 95, pp. 1647-1658, 1988.
- [13] J. S. Ulvestad, "Effects of Source Structure on Astrometry and Geodesy," *International Astronomical Union Symposium No. 129, The Impact of VLBI on Astrophysics and Geophysics*, edited by M. J. Reid and J. M. Moran, Dordrecht: D. Reidel Publishing Co., pp. 429-430, 1988.

- [14] P. Charlot, J.-F. Lestrade, and C. Boucher, "3C 273 and DA 193 Mapped with Crustal Dynamics VLBI Data," *International Astronomical Union Symposium No. 129, The Impact of VLBI on Astrophysics and Geophysics*, edited by M. J. Reid and J. M. Moran, Dordrecht: D. Reidel Publishing Co., pp. 33-34, 1988.
- [15] G. Tang, B. Rönnäng, and L. Bååth, "Radio Source Structure from Geodetic VLBI Observations: 8 GHz Multi-epoch Maps of the Quasar 4C39.25," *Astronomy and Astrophysics*, vol. 185, pp. 87-93, 1987.
- [16] R. N. Treuhaft, "Deep Space Tracking in Local Reference Frames," *TDA Progress Report 42-94*, vol. April-June 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 1-15, August 15, 1988.
- [17] J. A. Zensus and T. J. Pearson (editors), *Superluminal Radio Sources*, Cambridge: Cambridge University Press, 1987.

**Table 1. Source maps used in statistical study**

IAU Name	Other Name	Structure Type [8, 9]	Data Date
0133 + 476	OC 457	Compact	Dec. 1978
0153 + 744	...	Compact double	April 1982
0212 + 735	...	Asymmetric	Sept. 1980
0316 + 413	3C 84	Unclassified	Dec. 1978
0454 + 844	...	Compact	Aug. 1981
0710 + 439	OI 417	Compact double	Dec. 1982
0711 + 356	OI 318	Compact double	Dec. 1982
0804 + 499	OJ 508	Compact	Dec. 1979
0831 + 557	4C 55.16	Steep-spectrum core	Dec. 1979
0836 + 710	4C 71.07	Asymmetric	Sept. 1980
0850 + 581	4C 58.17	Unclassified	July 1980
0859 + 470	4C 47.29	Compact	Dec. 1978
0906 + 430	3C 216	Steep-spectrum core	Dec. 1979
0923 + 392	4C 39.25	Unclassified	Dec. 1978
1458 + 718	3C 309.1	Steep-spectrum core	Dec. 1982
1624 + 416	4C 41.32	Asymmetric	July 1980
1642 + 690	4C 69.21	Asymmetric	July 1980
1652 + 398	4C 39.49	Unclassified	July 1980
1807 + 698	3C 371	Asymmetric	Dec. 1982
1823 + 568	4C 56.27	Asymmetric	Dec. 1979
1828 + 487	3C 380	Steep-spectrum core	Dec. 1978
1928 + 738	4C 73.18	Asymmetric	Sept. 1980
2021 + 614	OW 637	Compact double	Dec. 1982
2200 + 420	BL Lac	Asymmetric	Dec. 1978
2351 + 456	4C 45.51	Asymmetric	July 1980
2352 + 495	OZ 488	Compact double	Dec. 1979

**Table 2. Structure-induced scatter in VLBI observables on California-Spain baseline, as a function of visibility amplitude, for all 26 sources**

Visibility	Scatter about centroid			No. points
	Group delay, psec	Phase delay, psec	Delay rate, E-15 sec/sec	
0.0 - 0.1	2614	271	53.4	285
0.1 - 0.2	703	478	45.1	891
0.2 - 0.3	419	495	33.4	934
0.3 - 0.4	195	226	19.8	1610
0.4 - 0.5	138	216	19.3	1563
0.5 - 0.6	78	189	5.1	776
0.6 - 0.7	50	175	3.8	513
0.7 - 0.8	19	143	2.2	514
0.8 - 0.9	10	137	1.1	333
0.9 - 1.0	2	26	0.2	772

**Table 3. Structure-induced scatter in VLBI observables on California–Australia baseline, as a function of visibility amplitude, for the 12 visible sources**

Visibility	Scatter about centroid			No. points
	Group delay, psec	Phase delay, psec	Delay rate, E-15 sec/sec	
0.0 – 0.1	2273	90	68.9	39
0.1 – 0.2	600	76	24.7	65
0.2 – 0.3	258	63	18.9	80
0.3 – 0.4	217	63	11.5	121
0.4 – 0.5	101	35	4.3	154
0.5 – 0.6	45	39	2.5	66
0.6 – 0.7	23	24	1.6	85
0.7 – 0.8	24	61	1.2	13
0.8 – 0.9	9	27	0.8	56
0.9 – 1.0	1	26	0.2	134

**Table 4. Scatter of VLBI observables for circumpolar sources and other sources on California–Spain baseline, all 26 sources are included**

Sources	Scatter about centroid			No. points
	Group delay, psec	Phase delay, psec	Delay rate, E-15 sec/sec	
Circumpolar	465	385	30.3	4320
Other	665	90	15.3	3871

**Table 5. Structure-induced scatter in VLBI observables on California–Spain baseline for all 26 sources, for points above a given visibility amplitude**

Visibility cutoff	Scatter about centroid			No. points
	Group delay, psec	Phase delay, psec	Delay rate, E-15 sec/sec	
0.0	569	286	24.4	8191
0.1	298	287	22.8	7906
0.2	193	252	18.1	7015
0.3	126	189	14.3	6081
0.4	90	173	11.7	4471
0.5	46	145	3.2	2908
0.6	26	125	2.2	2132
0.7	12	103	1.4	1619
0.8	6	78	0.7	1105
0.9	2	26	0.2	772

**Table 6. Structure-induced scatter in VLBI observables on California–Australia baseline for the 12 visible sources, for points above a given visibility amplitude**

Visibility cutoff	Scatter about centroid			No. points
	Group delay, psec	Phase delay, psec	Delay rate, E-15 sec/sec	
0.0	541	50	18.2	813
0.1	216	47	10.6	774
0.2	134	43	8.1	709
0.3	109	40	5.5	629
0.4	59	32	2.7	508
0.5	23	31	1.4	354
0.6	14	29	1.0	288
0.7	8	30	0.5	203
0.8	5	27	0.5	190
0.9	1	26	0.2	134

**Table 7. Structure-induced scatter in VLBI observables for different classes of source structure, on California–Spain baseline**

Class	Scatter about centroid				No. points
	No. sources	Group delay, psec	Phase delay, psec	Delay rate, E-15 sec/sec	
All	26	569	286	24.4	8191
Steep-spectrum	4	833	612	49.3	1259
Compact double	5	469	316	30.9	1535
Asymmetric	9	101	114	7.9	3337
Unclassified	4	1244	39	12.1	907
Compact	4	15	16	1.8	1153
“Bad”	10	939	459	39.6	2964
“Good”	16	83	93	6.4	5227

**Table 8. Structure-induced scatter in VLBI observables on California–Spain baseline, as a function of visibility amplitude, for the 16 “good” sources**

Visibility	Scatter about centroid			No. points
	Group delay, psec	Phase delay, psec	Delay rate, E-15 sec/sec	
0.0 – 0.1	349	82	24.1	73
0.1 – 0.2	163	73	11.0	324
0.2 – 0.3	135	122	11.2	314
0.3 – 0.4	94	122	7.2	1051
0.4 – 0.5	48	109	4.6	1129
0.5 – 0.6	28	93	3.9	585
0.6 – 0.7	17	77	3.8	410
0.7 – 0.8	16	36	2.2	438
0.8 – 0.9	5	40	1.4	178
0.9 – 1.0	2	27	0.2	725

**Table 9. Structure-induced scatter in VLBI observables on California–Australia baseline, as a function of visibility amplitude, for the 7 visible “good” sources**

Visibility	Scatter about centroid			No. points
	Group delay, psec	Phase delay, psec	Delay rate, E-15 sec/sec	
0.0 – 0.1	876	41	55.3	7
0.1 – 0.2	187	83	24.5	17
0.2 – 0.3	78	73	10.5	29
0.3 – 0.4	154	68	9.5	52
0.4 – 0.5	100	35	3.7	129
0.5 – 0.6	47	39	2.3	47
0.6 – 0.7	23	25	1.6	76
0.7 – 0.8	24	61	1.2	13
0.8 – 0.9	8	79	0.9	6
0.9 – 1.0	1	30	0.1	107

**Table 10. Scatter of VLBI observables for circumpolar sources and other sources on California–Spain baseline, only the 16 “good” sources are included**

Sources	Scatter about centroid			No. points
	Group delay, psec	Phase delay, psec	Delay rate, E-15 sec/sec	
Circumpolar	97	115	7.0	2880
Other	61	55	5.6	2347

**Table 11. Structure-induced scatter in VLBI observables on California–Spain baseline for 16 “good” sources, for points above a given visibility amplitude**

Visibility cutoff	Scatter about centroid			No. points
	Group delay, psec	Phase delay, psec	Delay rate, E-15 sec/sec	
0.0	83	93	6.4	5227
0.1	72	93	5.8	5154
0.2	62	94	5.3	4830
0.3	53	92	4.6	4516
0.4	31	80	3.4	3465
0.5	17	62	2.7	2336
0.6	12	47	2.2	1751
0.7	10	32	1.3	1341
0.8	3	30	0.6	903
0.9	2	27	0.2	725

**Table 12. Structure-induced scatter in VLBI observables on California–Australia baseline for the 7 visible “good” sources, for points above a given visibility amplitude**

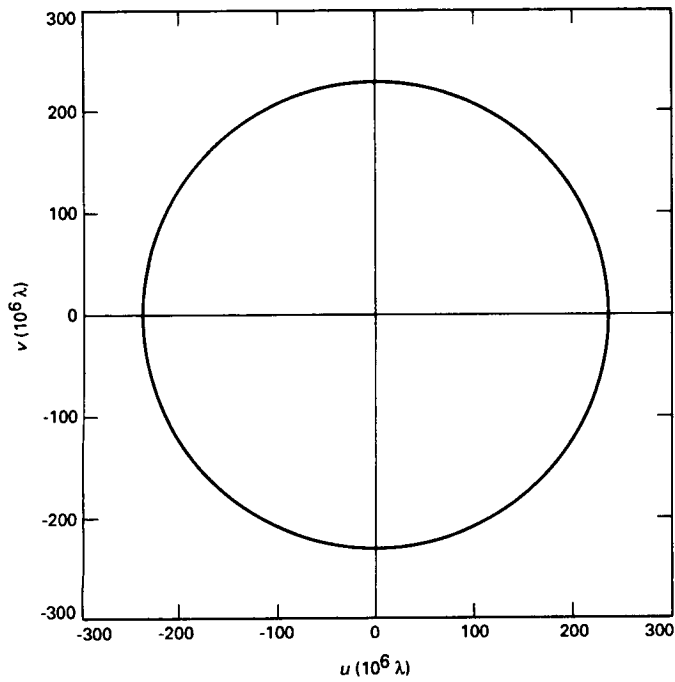
Visibility cutoff	Scatter about centroid			No. points
	Group delay, psec	Phase delay, psec	Delay rate, E-15 sec/sec	
0.0	135	45	8.8	483
0.1	85	45	6.8	476
0.2	79	43	4.6	459
0.3	79	40	4.0	430
0.4	62	35	2.4	378
0.5	25	35	1.4	249
0.6	14	33	1.1	202
0.7	8	38	0.4	126
0.8	2	34	0.2	113
0.9	1	30	0.1	107

**Table 13. Structure-induced scatter for 16 “good” sources on 3 baselines**

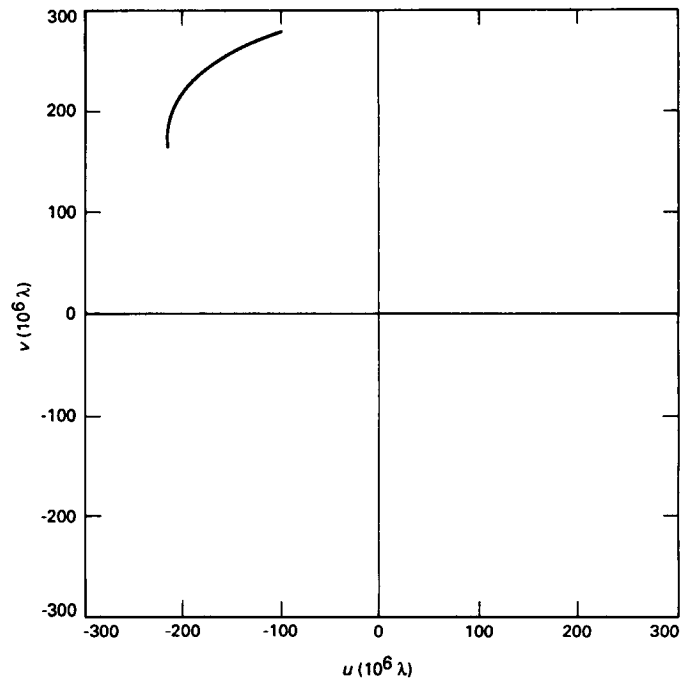
Baseline	Length, km	Scatter about centroid		
		Group delay, psec	Phase delay, psec	Delay rate, E-15 sec/sec
Goldstone–Madrid	8390	83	93	6.4
Haystack–Fort Davis	3140	28	41	2.1
Haystack–Green Bank	840	5	3	0.3

**Table 14. Structure-induced scatter for all 26 sources on 3 baselines**

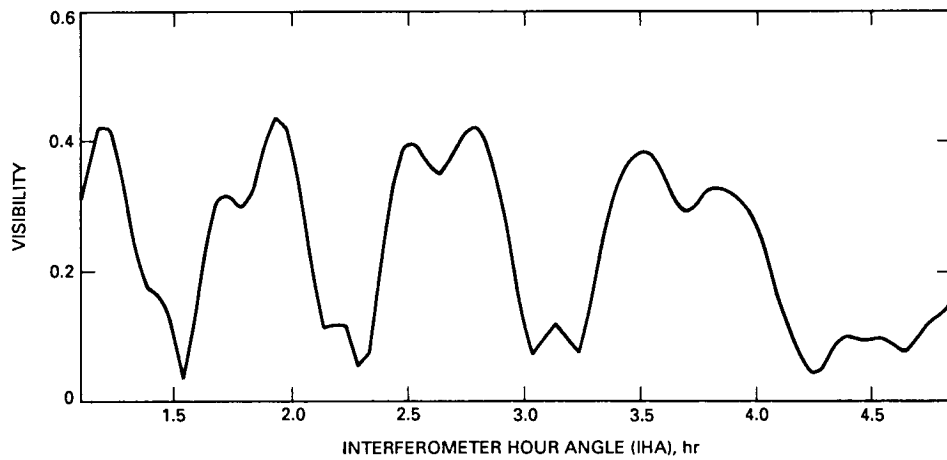
Baseline	Length, km	Scatter about centroid		
		Group delay, psec	Phase delay, psec	Delay rate, E-15 sec/sec
Goldstone–Madrid	8390	569	286	24.4
Haystack–Fort Davis	3140	104	107	8.1
Haystack–Green Bank	840	42	22	2.7



**Fig. 1. Plot of sampled  $(u,v)$  points for a circumpolar source (0153 + 744) on the California-Spain baseline.**

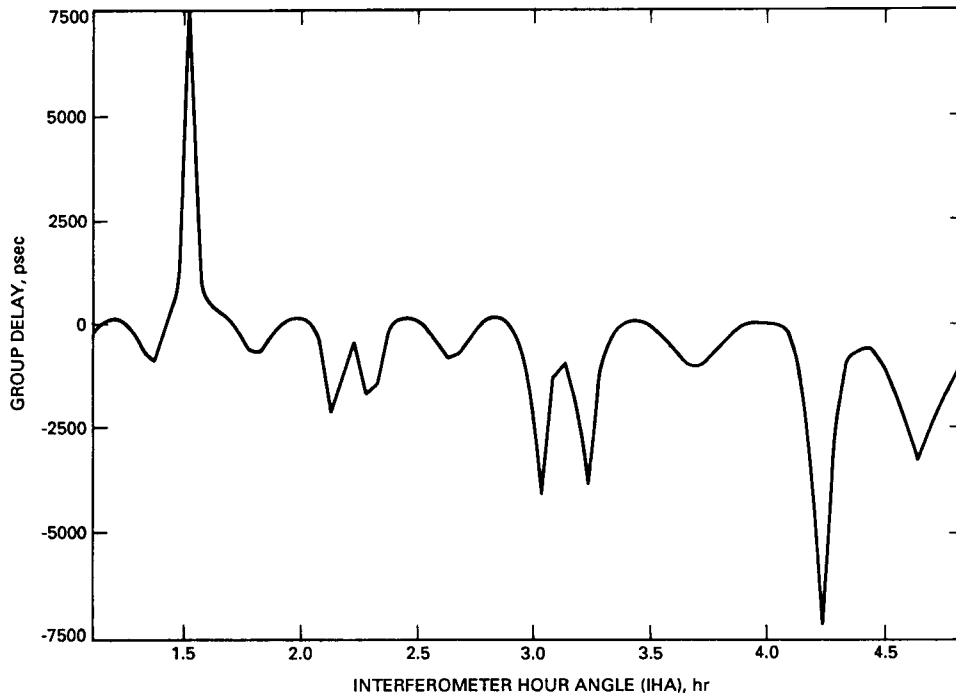


**Fig. 2. Plot of sampled  $(u,v)$  points for a lower declination source (0711 + 356) on the California-Australia baseline.**

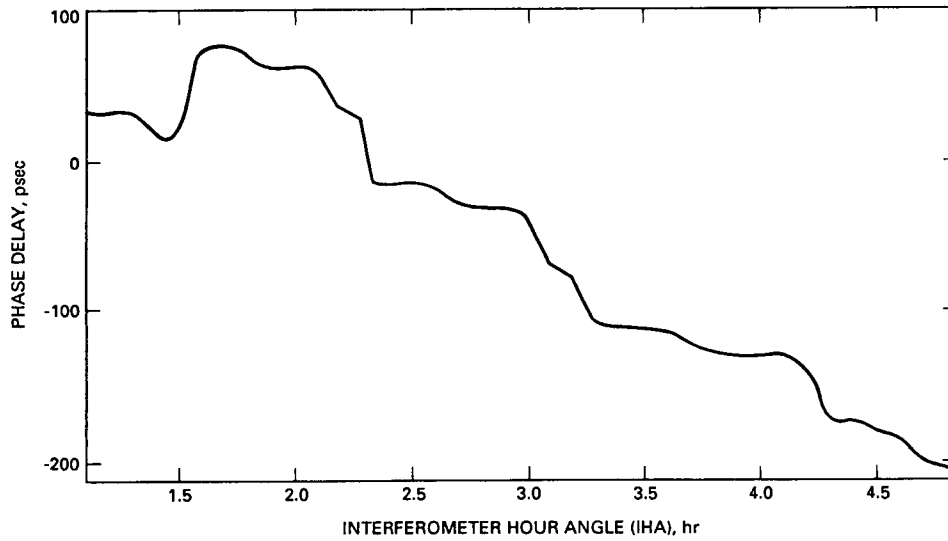


**Fig. 3. Computed visibility amplitude at 8.4 GHz for the compact double source 0710 + 439 on the California-Australia baseline.**

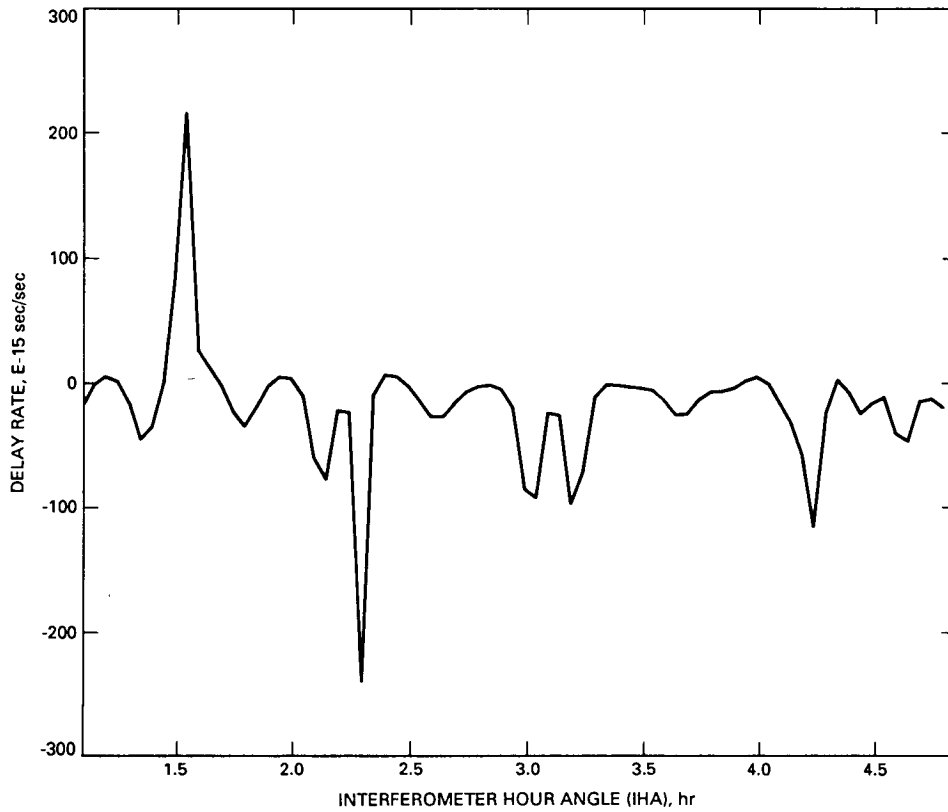




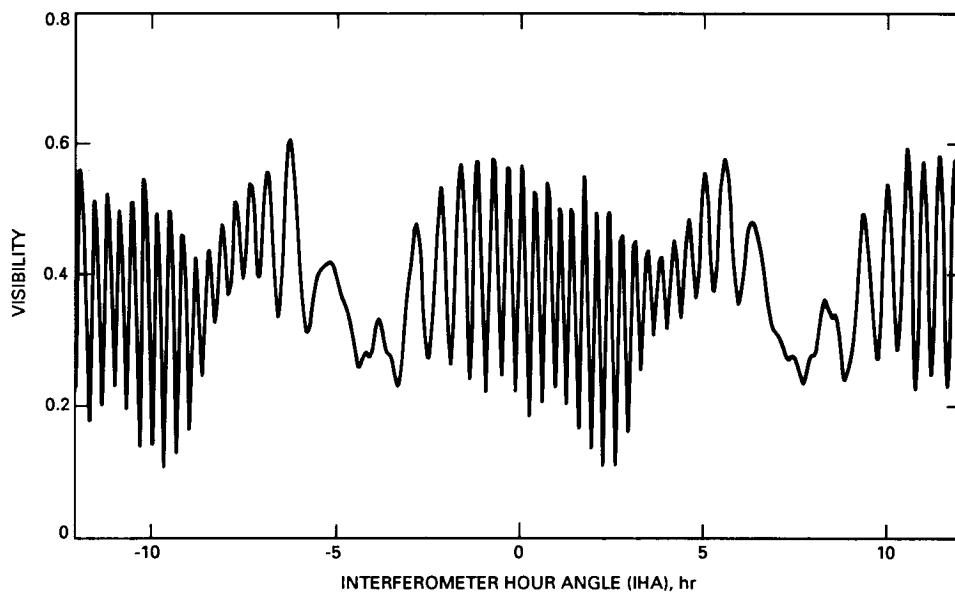
**Fig. 4. Structure-induced group delay at 8.4 GHz for the compact double source 0710 + 439 on the California-Australia baseline.**



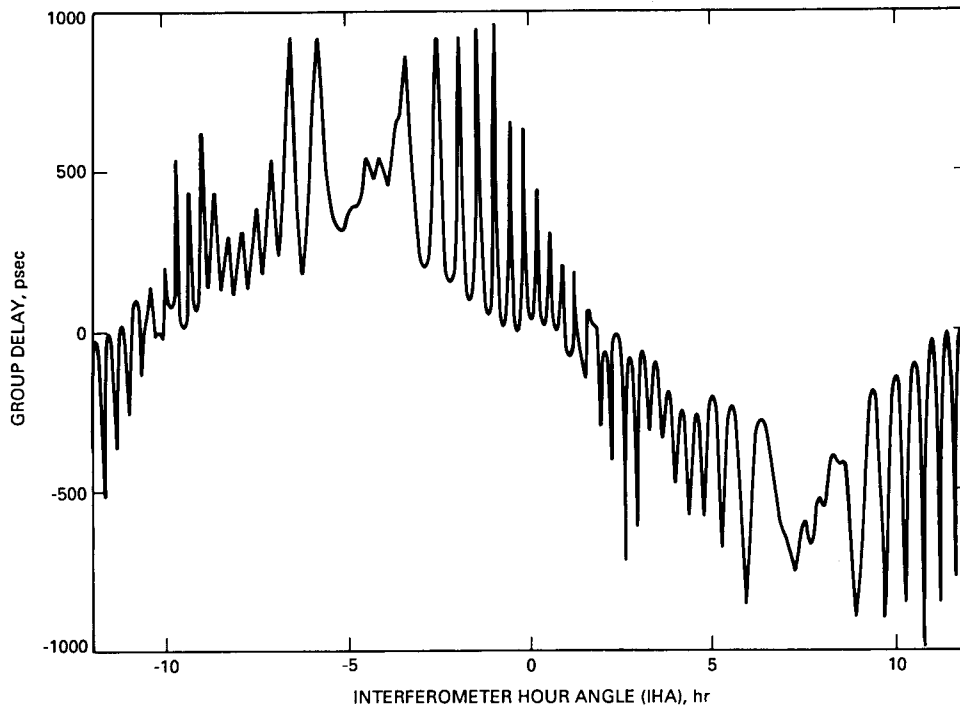
**Fig. 5. Structure-induced phase delay at 8.4 GHz for the compact double source 0710 + 439 on the California-Australia baseline.**



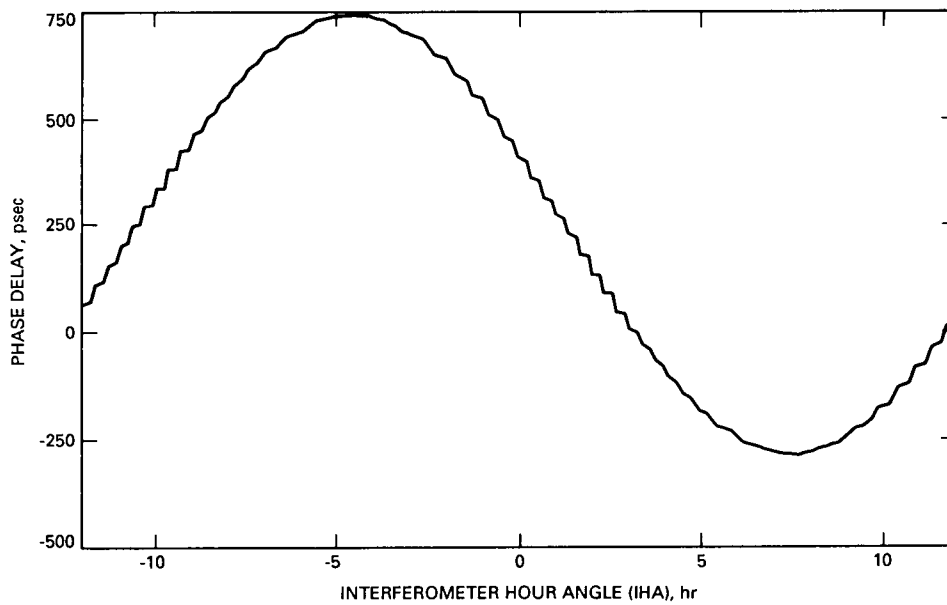
**Fig. 6. Structure-induced delay rate at 8.4 GHz for the compact double source 0710 + 439 on the California-Australia baseline.**



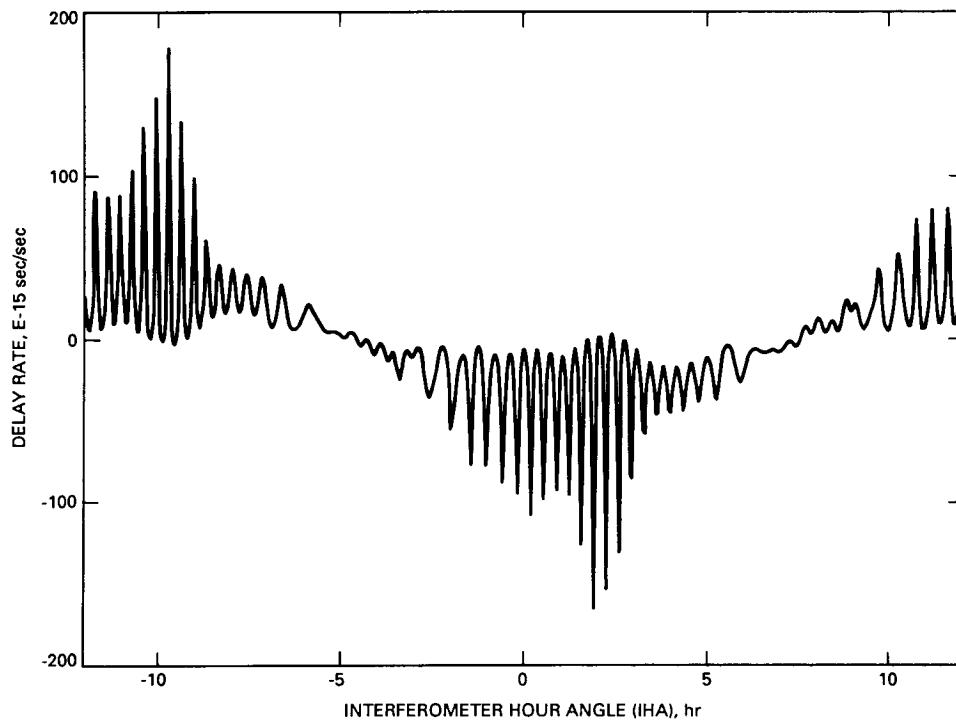
**Fig. 7. Computed visibility amplitude for the compact double source 0153 + 744 on the California-Spain baseline.**



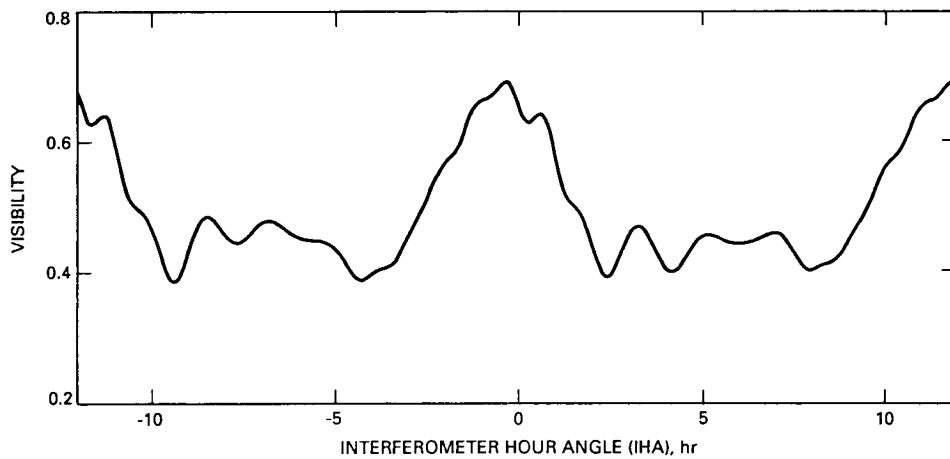
**Fig. 8. Structure-induced group delay for the compact double source 0153 + 744 on the California-Spain baseline.**



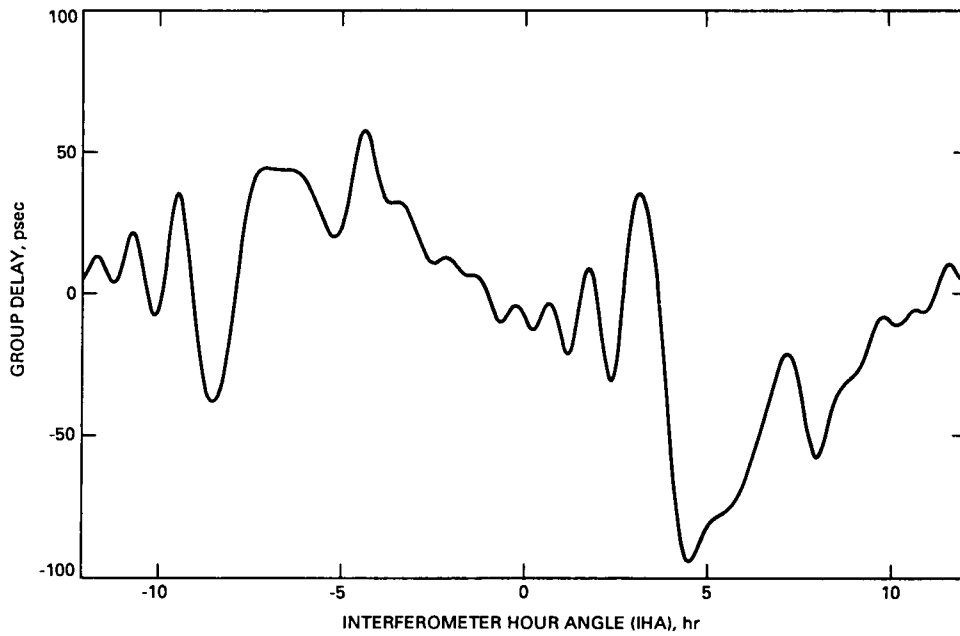
**Fig. 9. Structure-induced phase delay for the compact double source 0153 + 744 on the California-Spain baseline.**



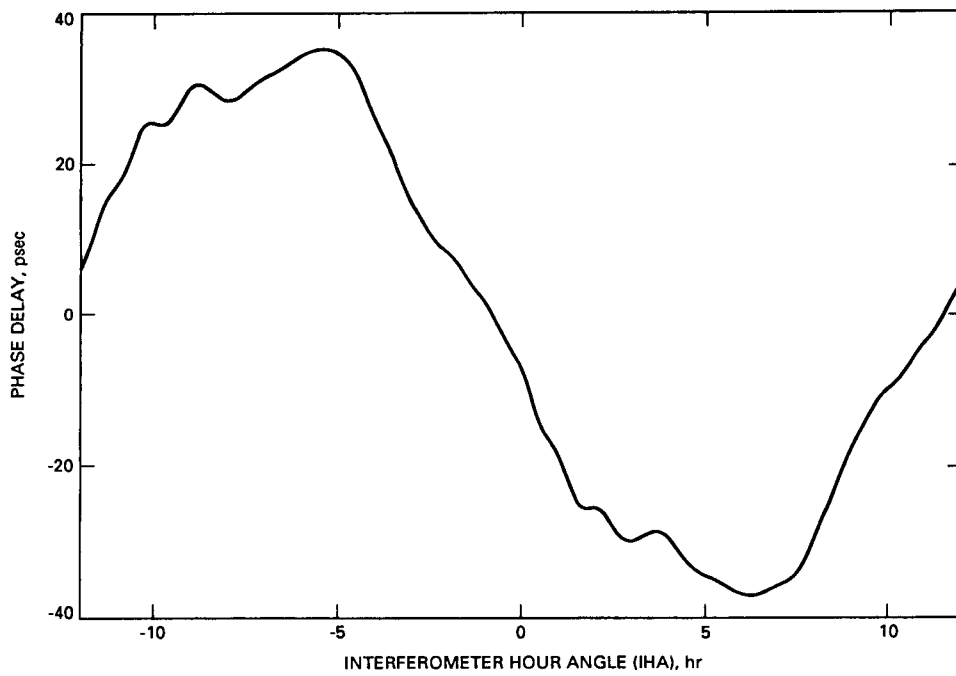
**Fig. 10. Structure-induced delay rate for the compact double source 0153 + 744 on the California-Spain baseline.**



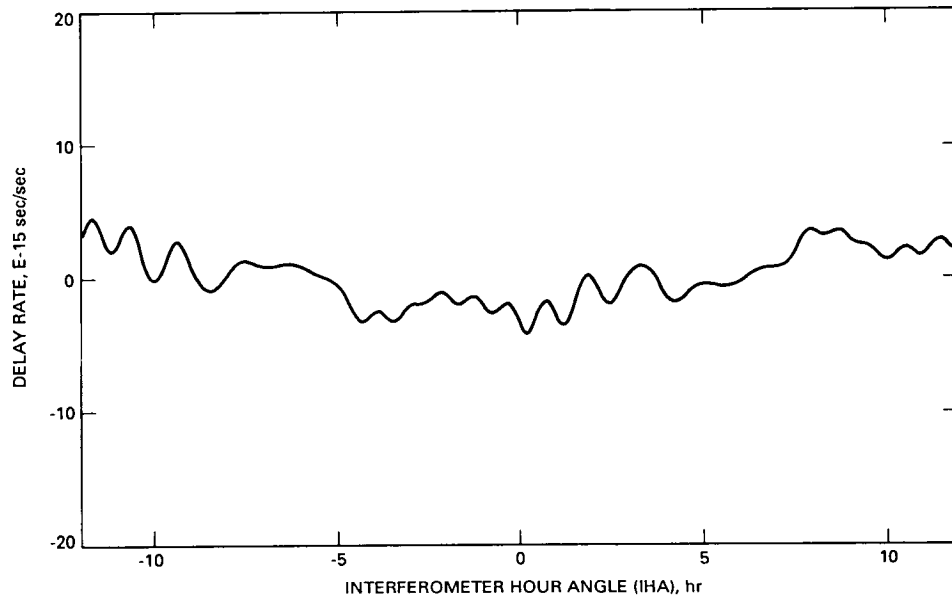
**Fig. 11. Computed visibility amplitude for the asymmetric source 1642 + 690 on the California-Spain baseline.**



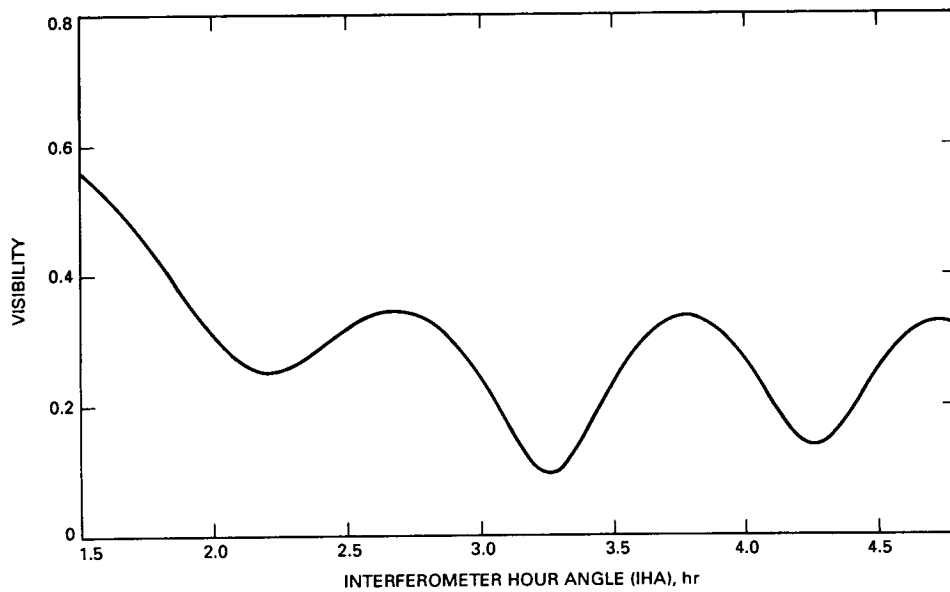
**Fig. 12. Structure-induced group delay for the asymmetric source 1642 + 690 on the California–Spain baseline.**



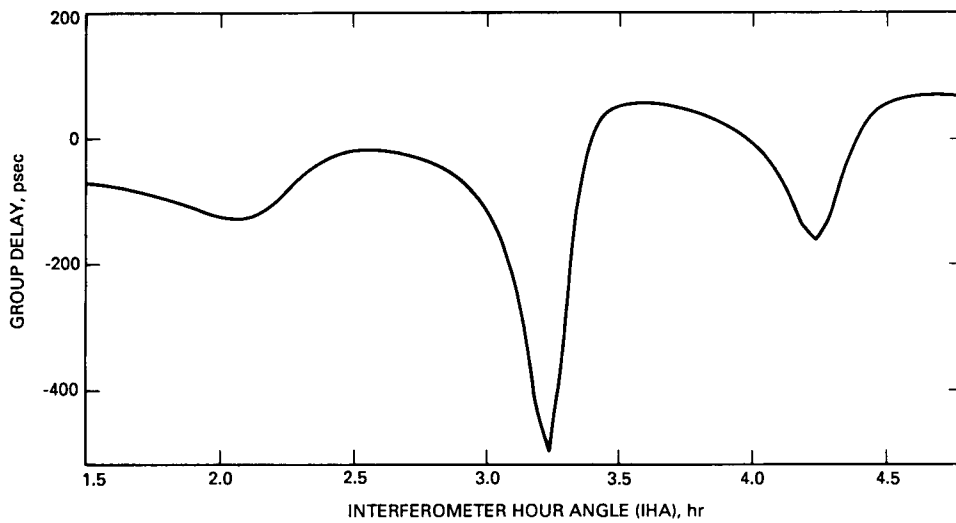
**Fig. 13. Structure-induced phase delay for the asymmetric source 1642 + 690 on the California–Spain baseline.**



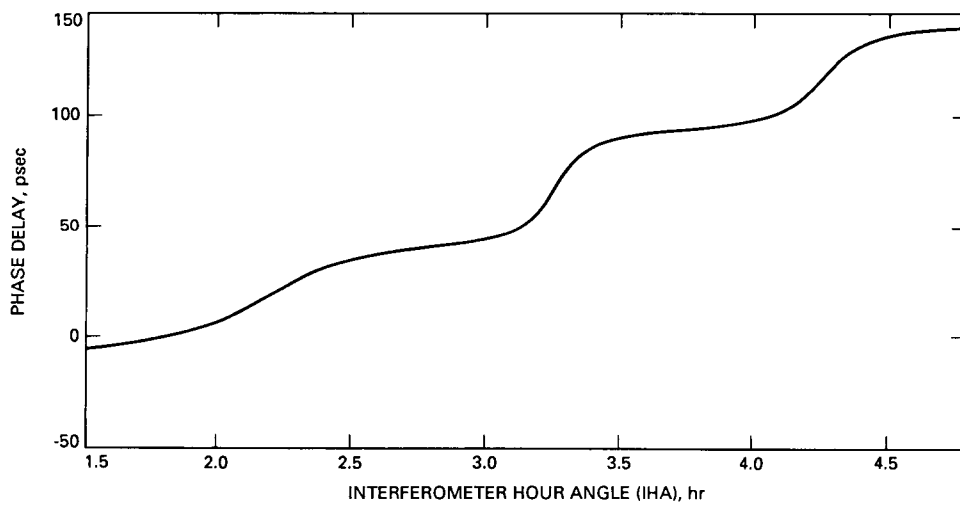
**Fig. 14. Structure-induced delay rate for the asymmetric source 1642 + 690 on the California-Spain baseline.**



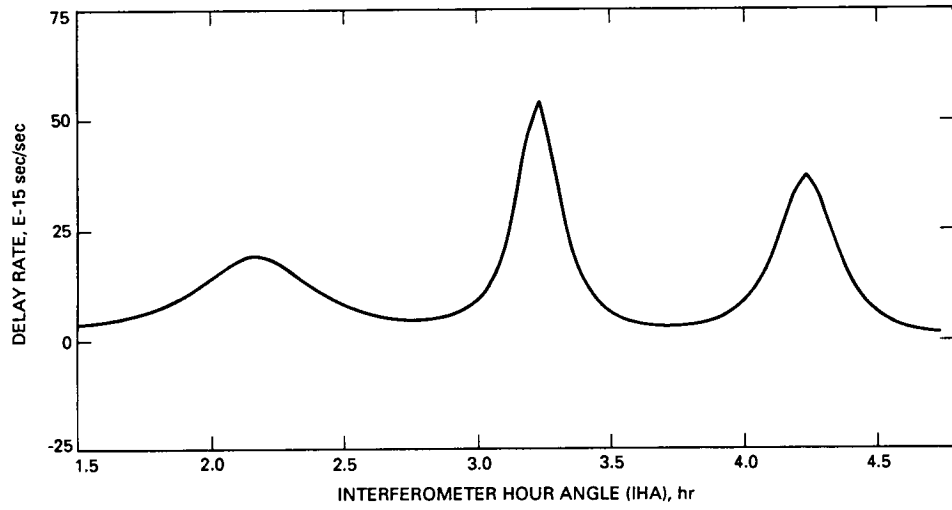
**Fig. 15. Computed visibility amplitude for the asymmetric source 2351 + 456 on the California-Australia baseline.**



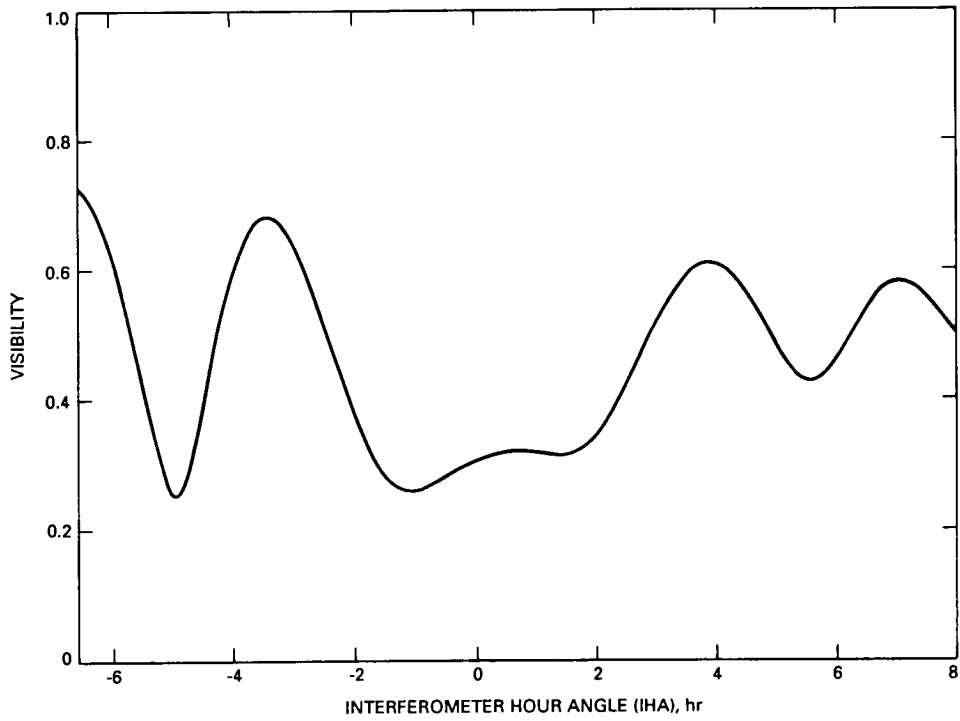
**Fig. 16. Structure-induced group delay for the asymmetric source 2351 + 456 on the California–Australia baseline.**



**Fig. 17. Structure-induced phase delay for the asymmetric source 2351 + 456 on the California–Australia baseline.**

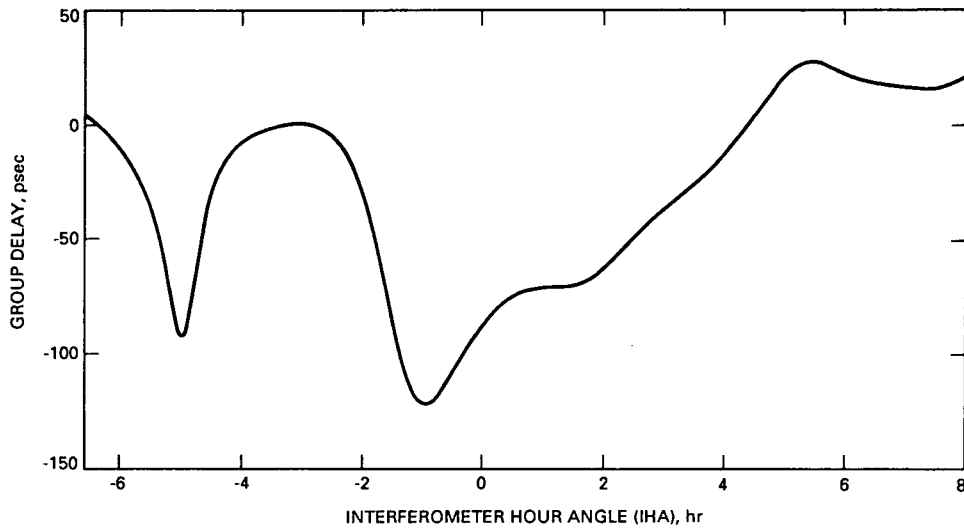


**Fig. 18. Structure-induced delay rate for the asymmetric source 2351 + 456 on the California–Australia baseline.**

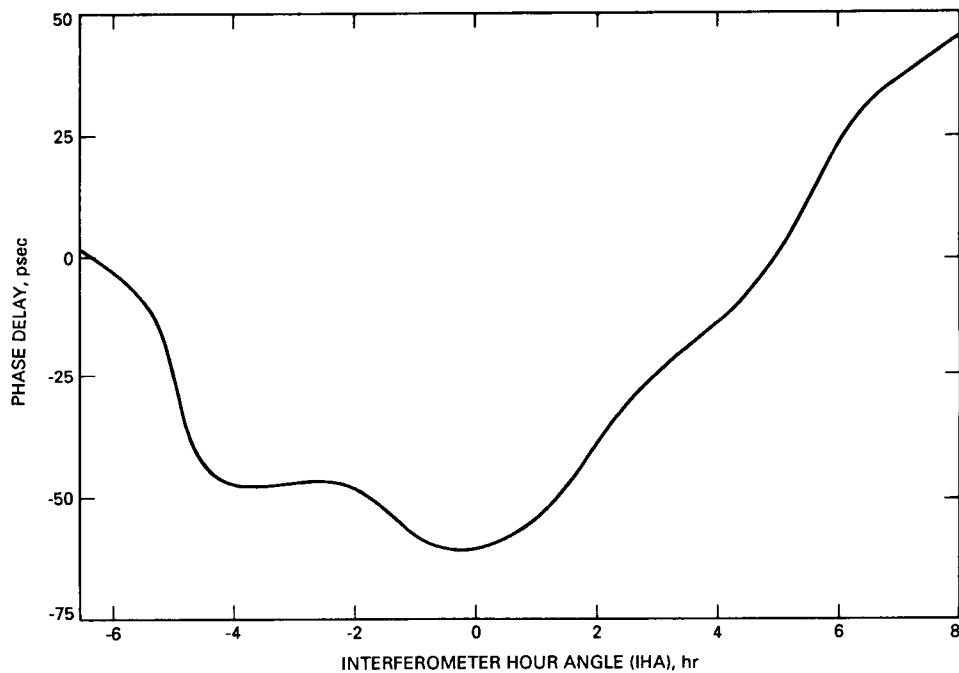


**Fig. 19. Computed visibility amplitude for the asymmetric source 2351 + 456 on the Haystack, MA–Fort Davis, TX baseline.**

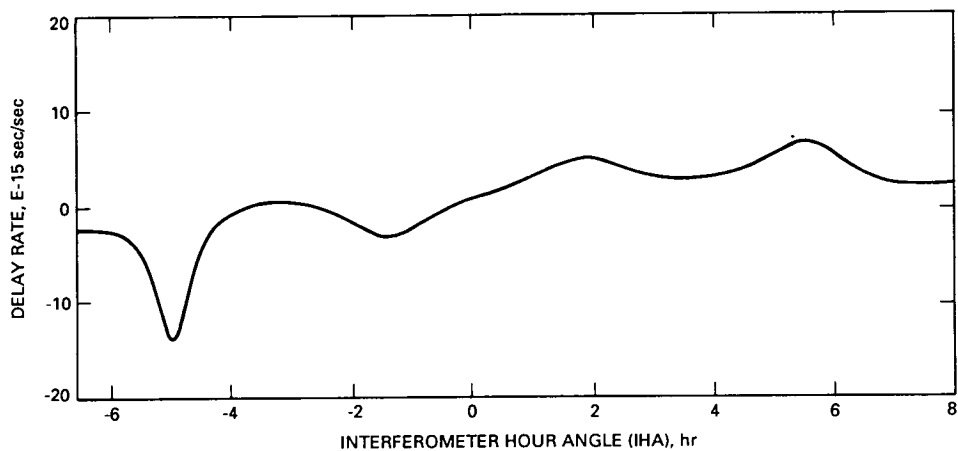




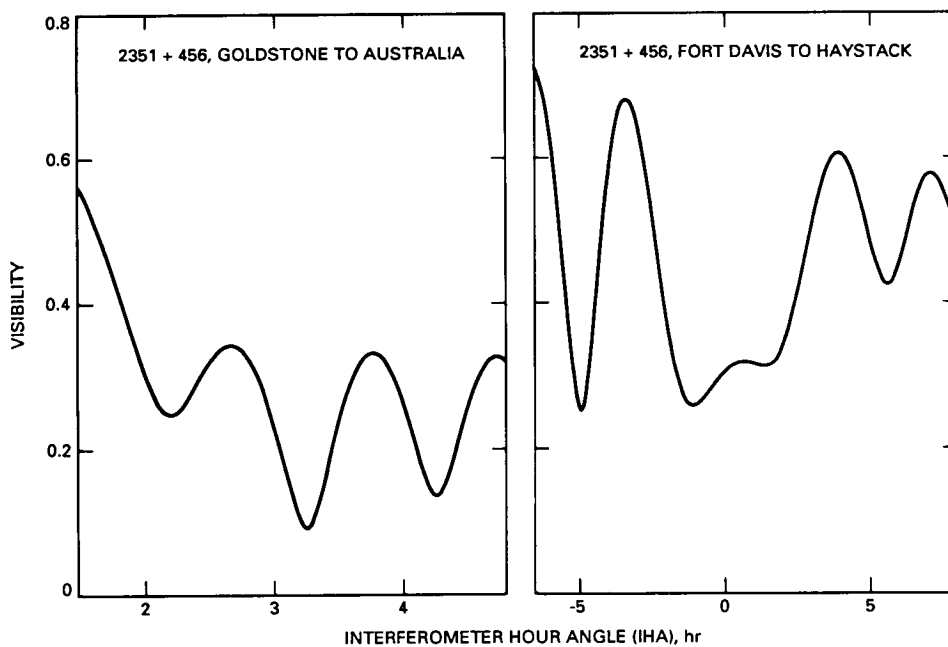
**Fig. 20. Structure-induced group delay for the asymmetric source 2351 + 456 on the Haystack, MA-Fort Davis, TX baseline.**



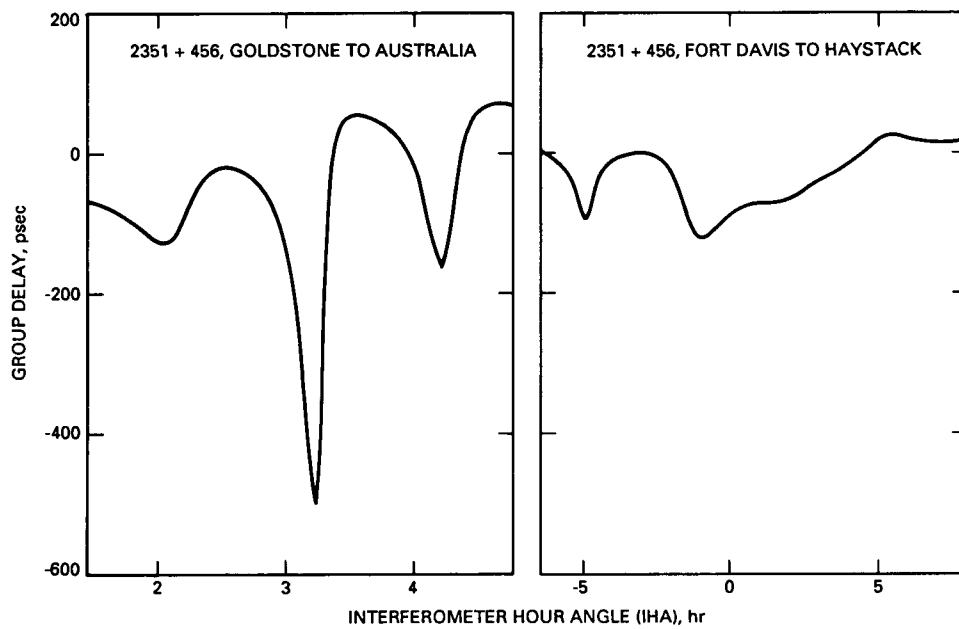
**Fig. 21. Structure-induced phase delay for the asymmetric source 2351 + 456 on the Haystack, MA-Fort Davis, TX baseline.**



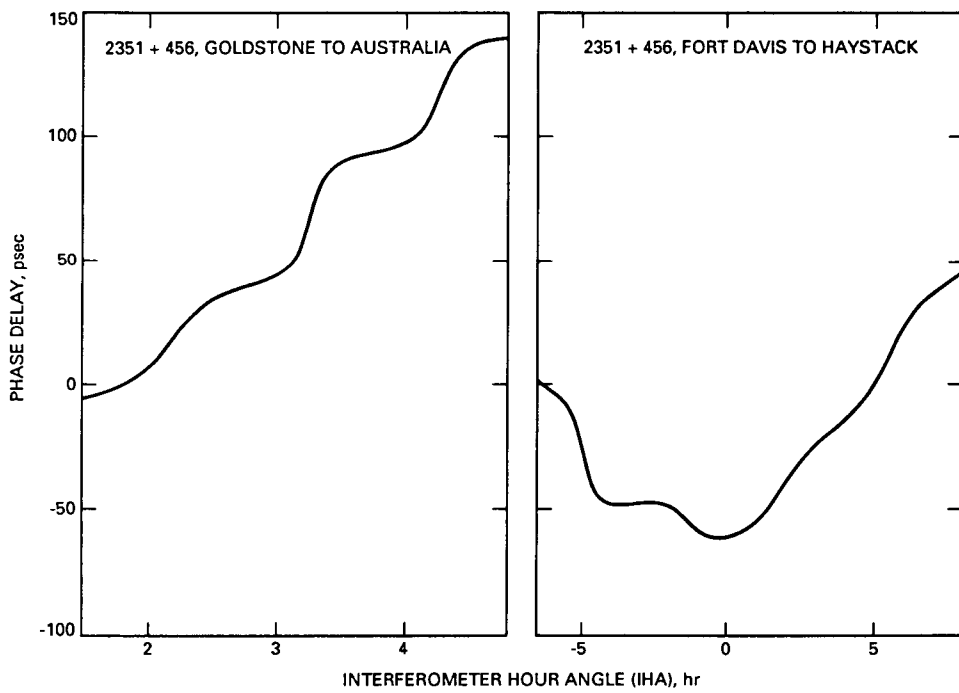
**Fig. 22. Structure-induced delay rate for the asymmetric source 2351 + 456 on the Haystack, MA-Fort Davis, TX baseline.**



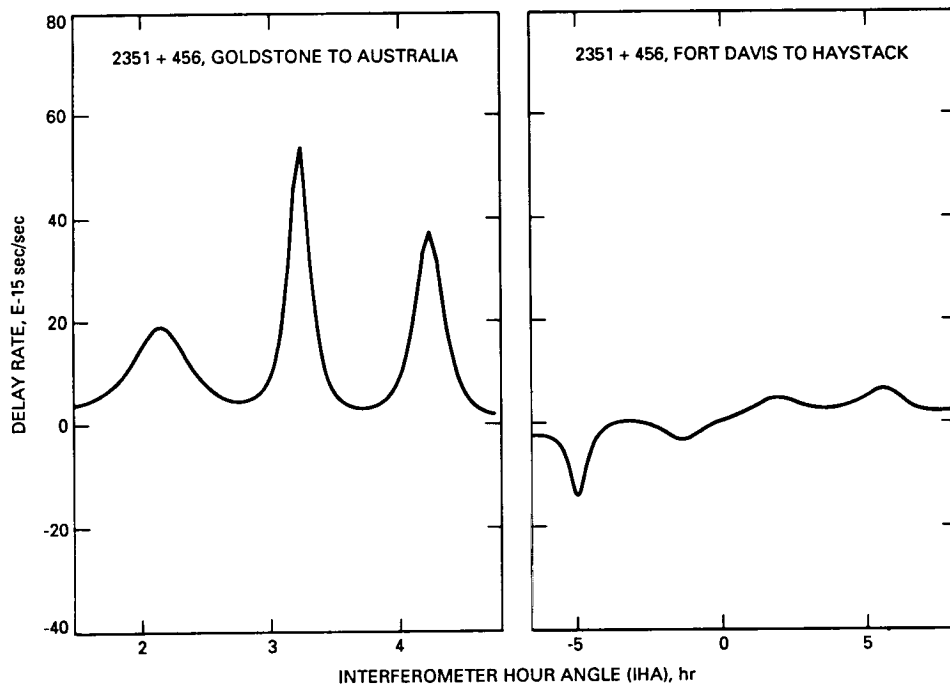
**Fig. 23. Visibility amplitudes from Figs. 15 and 19, shown on the same vertical scale.**



**Fig. 24. Structure-induced group delays from Figs. 16 and 20, shown on the same vertical scale.**



**Fig. 25. Structure-induced phase delays from Figs. 17 and 21, shown on the same vertical scale.**



**Fig. 26. Structure-induced delay rates from Figs. 18 and 22, shown on the same vertical scale.**

# An Apparatus for the Electrodynamic Containment of Charged Macroparticles

A. Williams, R. Melbourne, L. Maleki, G. Janik, and J. Prestage  
Communications Systems Research Section

*The dynamic motion of the ions contained in the trapped  $^{199}\text{Hg}^+$  frequency standard contributes to the stability of the standard. In order to study these dynamics, a macroscopic analog of the  $^{199}\text{Hg}^+$  trap is constructed. Containment of micron-sized particles in this trap allows direct visual observation of the particles' motion. Influenced by the confining fields and their own Coulomb repulsion, the particles can form stable arrays.*

## I. Introduction

JPL is involved in the research and development of new technologies which surpass the stability of standards such as the hydrogen maser. One of the technologies under development in the frequency standards laboratory (FSL) at JPL is the trapped mercury ion frequency standard.

Like the cesium beam, hydrogen maser, and other atomic frequency sources, the trapped ion standard is based on an atomic transition. When an atom makes a transition from an energy level  $E_1$  to a lower level  $E_2$ ,

a photon of frequency  $\nu = (E_1 - E_2)/h$  ( $h$  being Planck's constant) is emitted. Thus, an atomic transition creates a frequency source. However, the atomic energy levels are shifted, split, or broadened when disturbed by electric or magnetic fields, spin-exchange collisions, etc. Therefore, to obtain very stable frequency standards, it is desirable to protect the atoms from such offenders. This is the motivation for holding the atoms in an electromagnetic trap in a very high vacuum [1].

This article describes the construction and operation of a macroparticle trap which will be used to study the

dynamics of charged particles confined in a trap and the parameters that influence the details of their motion. Since the trap allows the confinement of micron-sized particles, a direct visual observation of the dynamics in the trap is possible. In particular, phenomena such as cooling, crystallization [2], and evaporation of confined particles become directly observable.

The versatility of this experimental environment permits the testing of various traps; the construction of a novel linear trap [3] for confinement of micron-sized particles will enable the first study of the dynamics of charged particles in this trap geometry. Insight into the motion of charged particles confined in such a linear trap will allow the modeling of the dynamics of ions in the linear trapped  $^{199}\text{Hg}^+$  frequency standard.

The magnitude of variation in the second-order Doppler shift comprises the largest practical limit on the frequency stability achievable with a trapped ion frequency standard [3]. Since the shift is due to the motion of ions in the trap, a model of the ion dynamics in the  $^{199}\text{Hg}^+$  trap will aid the determination of parameters which minimize variation in the second-order Doppler shift.

The organization of the article is as follows. First, in Section II, the theory of trapping charged particles in a radio frequency (RF) trap is discussed. The details and results of macroparticle trap research are described in Section III. Results obtained with a hyperbolic trap are given in Section IV, and a look ahead to other capabilities of the system is presented in Section V.

## II. Trapping Theory

The type of trap that has been often employed for frequency-standard applications is the Paul trap, one type of RF trap. It consists of three hyperbolic electrodes: two endcaps, which are a hyperboloid of two sheets, and a ring electrode, which is a hyperboloid of one sheet (Fig. 1). The trap, when assembled with the ring between the two endcaps, has characteristic dimensions in cylindrical coordinates of  $z_0$ , half the distance between the endcaps, and  $r_0$  (usually set to be equal to  $\sqrt{2} \cdot z_0$ ), half the distance across the ring. See Fig. 2 for a cross-sectional view of the assembled trap.

To operate the trap, an AC voltage in series with a DC voltage is applied between the ring and the endcaps (Fig. 2), which produces a potential

$$\phi(r, z) = \frac{V_{\text{DC}} - V_{\text{AC}} \cos \Omega t}{z_0^2} \left( z^2 - \frac{r^2}{2} \right) \quad (1)$$

inside the trap [2]. From this expression for the potential the electric field  $\mathbf{E} = -\nabla\phi$  is easily obtained:

$$\mathbf{E} = \frac{(-V_{\text{DC}} + V_{\text{AC}} \cos \Omega t)}{z_0^2} (2z\hat{z} - r\hat{r}) \quad (2)$$

It is clear that the field intensities  $E_z$  and  $E_r$  (and hence the forces  $F_z$  and  $F_r$ ) vary linearly with  $z$  and  $r$ , respectively, and that the electric field  $\mathbf{E}$  is zero at the center, or node ( $r = z = 0$ ) of the trap. It is also seen that the field intensity in the  $z$  direction is twice as strong as that in the  $r$  direction and has opposite sign. This negative sign shows that when the confining fields point toward the node in one direction, they point away in the other.

It may not be immediately obvious that an alternating potential should stably confine charged particles. The "secret" lies in the fact that the forces toward the node increase with distance away from the node. Suppose that in one half-cycle of the potential, the force on a particle is away from the node (Fig. 3). At the second half of the cycle, the particle is situated at a point of stronger field than where it started. Now the force is not only directed toward the node, it is larger in magnitude, thus leaving the particle closer to the node than it started. It should now be clear that over one period of the trapping potential, the net effect is to move the particle nearer to the node.

Stable confinement occurs in a region dependent on the charge-to-mass ratio of the particles,  $V_{\text{AC}}$ ,  $V_{\text{DC}}$ , and  $\Omega$ . When the system is operated with  $V_{\text{DC}}$  set to zero, the limit of stable containment is given by [2]

$$\left( \frac{e}{m} \right) \left( \frac{V_{\text{AC}}}{z_0^2} \right) \left( \frac{1}{\Omega^2} \right) - 0.908 < 0 \quad (3)$$

Thus, the charge-to-mass ratio can be determined by exploring the boundary of the stability region, where the particle is no longer trapped. Either  $V_{\text{AC}}$  or  $\Omega$  is held constant and the other varied in order to find the boundary of stable operation. This experimentally determined value can then be substituted into Eq. (3) in order to solve for the charge-to-mass ratio.

### III. The Macroparticle Trap

In 1959 the first visual observations of macroparticles in a Paul trap were made by Wuerker and his co-workers [2]. These investigations revealed interesting dynamic behavior when many particles were held in the trap—influenced by both the confining fields of the trap and their own Coulomb repulsion, the particles formed a stable array (Fig. 4). When the driving frequency of the trap was increased, the crystal “melted” and the particles underwent random motion in the trap. If the frequency was subsequently decreased, the particles would once again take up their stable array.

The first version of the JPL macroparticle trap consists of a hyperbolic electrode configuration with  $r_0 = 1.9$  cm. Alumina powder (particle size nominally  $5 \mu\text{m}$ ) is loaded into a small basket of stainless wire mesh. The basket is suspended from a Sonalert piezoelectric beeper which, when activated, shakes the basket and causes a fine mist of powder to fall through the mesh. This beeper-basket assembly is suspended above the trap, and the particles fall into the trap through a hole in the top endcap (Fig. 5). It is also possible to release particles from the basket without using the beeper by causing a discharge around the basket.

The entire trapping apparatus is mounted inside a custom vacuum chamber which has electrical feedthroughs on its bottom flange, a viewport for viewing in the  $r$ - $z$  plane, two illumination windows perpendicular to the viewport, and a vacuum port. A mechanical pump evacuates the chamber to a pressure of about 5 mtorr. While the trap is operated primarily in this pressure regime, operation with pressures as high as 200 mtorr has also been carried out. The particles may be illuminated by any light source with sufficient power to make them visible; this particular system has been illuminated by a 0.5-mW helium-neon laser, a 15-mW argon ion laser, and a 300-mW dye laser. A Bausch and Lomb microscope with magnification  $20\times$  completes the viewing apparatus.

The 5-micron alumina particles have enough static charge to be trapped readily by the potential generated with  $V_{AC} \approx 300$  volts at frequency  $\Omega = 2\pi 60$  Hz. The trapping potential can thus be conveniently derived from ordinary wall power (110 V, 60 Hz) passed through a variac and a step-up transformer. By adjusting the variac, the RF potential can be chosen to be any value between 0 and 440 V. The DC voltage ( $V_{DC}$ ) in series with the AC voltage is set to zero, and the bottom endcap is grounded. There

is also a DC voltage ( $V_g$ ) which can be applied to the top endcap in order to counteract the gravitational pull on the particles (Fig 2).

With this system particles are readily trapped when the pressure in the vacuum chamber is lower than about 200 mtorr. Often both positively and negatively charged particles are trapped. The sign of the DC potential  $V_g$  applied to the top electrode of the trap allows the determination of the sign of the charge of the trapped particles.

### IV. Results

Operating the trap in this way reveals behavior similar to that seen by earlier experimenters. Up to about 50 particles have been trapped by this system. Often both positively and negatively charged particles are contained. In the many-particle case, a crystal formation centered on the node and outlining the trapping fields is seen when  $V_g$  cancelled the weight of the particles (Fig. 6). The 60-Hz micromotion of the particles is seen to decrease in amplitude, as it should, when the particles are nearer to the node. With  $V_g = 0$ , the particles are pulled by gravity into the bottom of the trapping potential still in their crystalline formation. By adjusting the trapping potentials, the crystal can be broken and reformed. The particles have a wide range of charge-to-mass ratios, which can be approximately determined from the potential  $V_g$  necessary to counteract the gravitational force, on the order of  $10^{-3}$  °C/kg. Those particles whose low charge-to-mass ratios cause them to locate themselves on the lower region of the particle cloud and can be ejected from the trap by varying  $V_g$ .

### V. Other Capabilities of the System

Since  $\Omega/2\pi$  is fixed at 60 Hz in this system, the charge-to-mass ratio of the particles can also be determined by measuring the maximum  $V_{AC}$  for stable confinement and using this measured value in Eq. (3):

$$\frac{e}{m} = 0.908 \left( \frac{(1.9 \text{ cm})^2 (2\pi \cdot 60 \text{ Hz})^2}{4V_{AC}^{\max}} \right) \quad (4)$$

By replacing the hyperbolic electrodes with any other suitably sized electrode configuration, the motion of charged particles in novel trap geometries can be studied with this system. The linear trap previously mentioned

has been implemented in the macroparticle system in this way, allowing visual study of the particle motion associated with the linear trapped  $^{199}\text{Hg}^+$  frequency standard. A thorough study of the dynamics of particles in the linear trap is currently under way.

Since any dustlike or aerosol material with appropriate charge-to-mass ratio can be trapped in this system, a multitude of experiments on various substances are possible. For example, there are plans to operate the system at  $0^\circ\text{C}$  so that ice particles can be studied; such an investigation has significant application to atmospheric physics. A version of the system suitable for installation on a spacecraft could trap cosmic dust, which could subsequently be

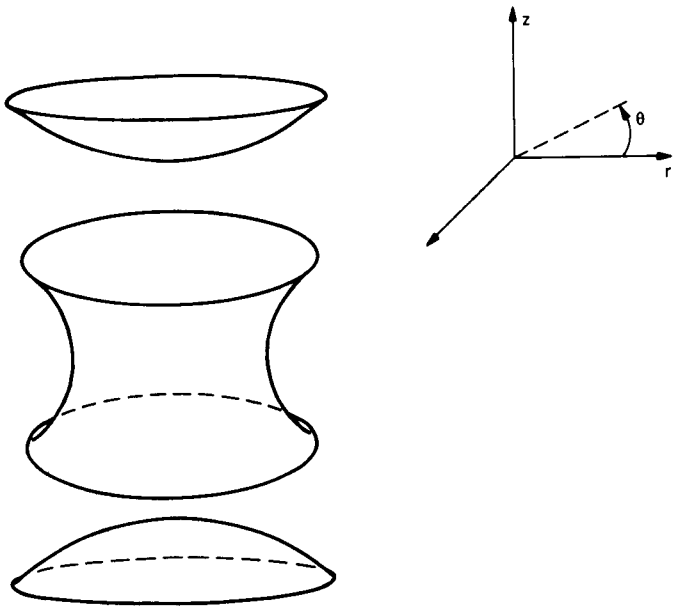
irradiated with laser light for spectroscopic study of this material.

The physics of many-particle systems can also be readily studied with use of a macroparticle trap because the number of particles in the trap can be increased (by activating the beeper) or decreased (by ejecting those particles with extreme charge-to-mass ratios). The micro-motion of the particles and the form of the crystal may be observed as the crystal is cooled by viscous drag or heated by increasing  $V_{AC}$ . Quantitative measurements on the phenomena of crystallization and melting as functions of number of particles and particle energy help to describe the nature of several-particle systems.

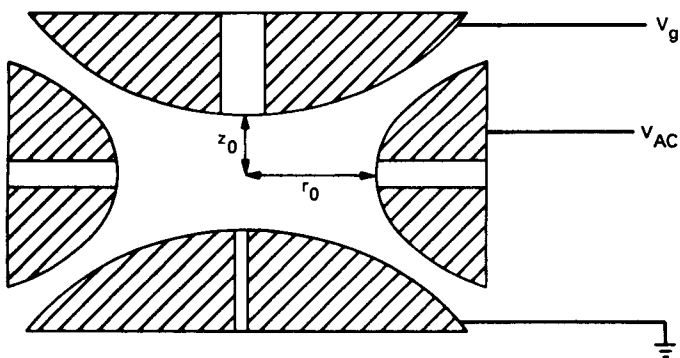
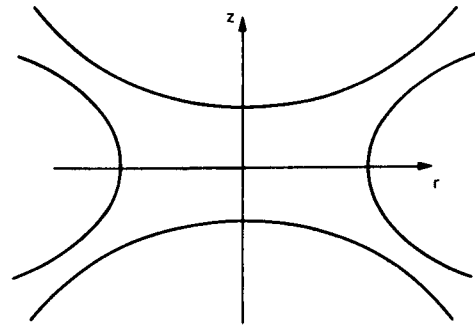
## References

- [1] L. Maleki, J. D. Prestage, and G. J. Dick, "Atomic Frequency Standards for Ultra-High-Frequency Stability," *TDA Progress Report 42-91*, vol. July-September 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 67-72, November 15, 1987.
- [2] R. F. Wuerker, H. Shelton, and R. V. Langmuir, "Electrodynamic Containment of Charged Particles," *J. App. Phys.*, vol. 30, pp. 342-349, March 1959.
- [3] J. D. Prestage, G. J. Dick, and L. Maleki, "New Ion Trap for Frequency Standard Applications," *Proc. 20th Ann. PTTI Meeting*, pp. 305-311, 1988.

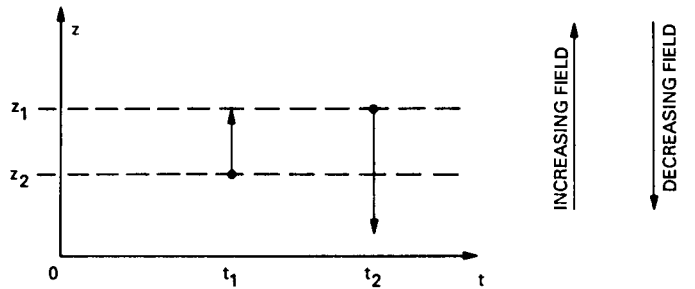




**Fig. 1. The hyperbolic electrodes of a Paul trap.**



**Fig. 2. Cross-sectional view of a Paul trap.**



**Fig. 3. One period of the AC trapping potential forces a particle toward the node ( $r = z = 0$ ).**

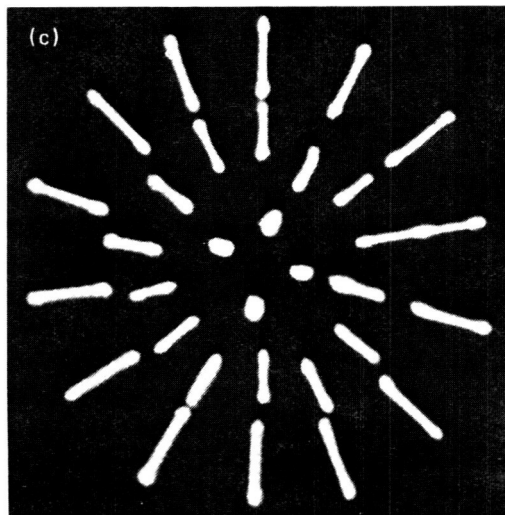
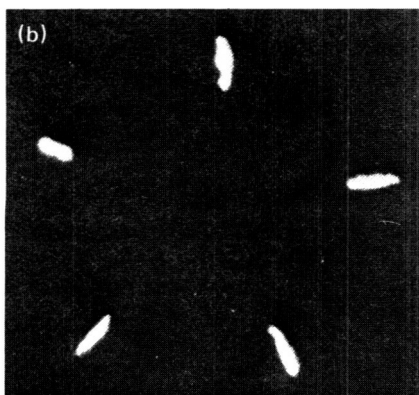
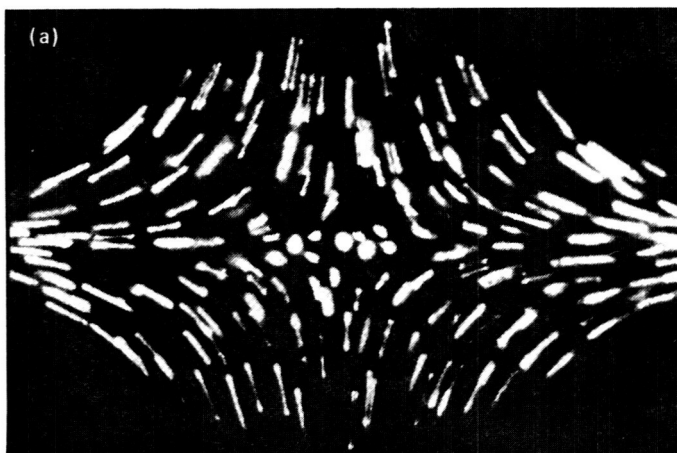


Fig. 4. Results of early work in macroparticle traps from [2], showing crystalline formations: (a) about 100 particles viewed in the  $r$ - $z$  plane, (b) 5 particles viewed in the  $r$ - $\theta$  plane, and (c) 32 particles viewed in the  $r$ - $\theta$  plane.

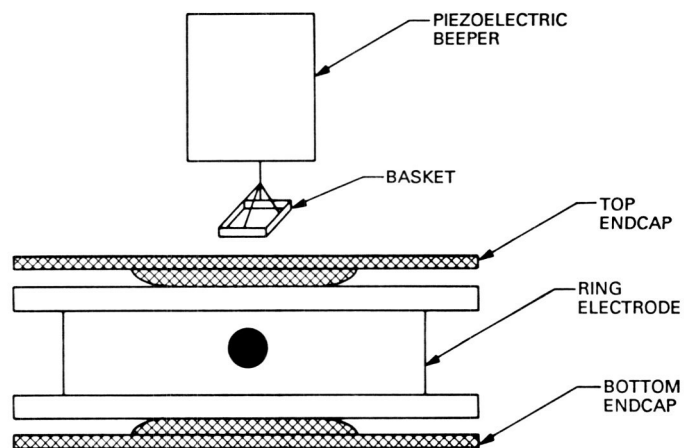


Fig. 5. Diagram of the present macroparticle introduction system, showing the beeper-basket assembly used for introduction through the top endcap.

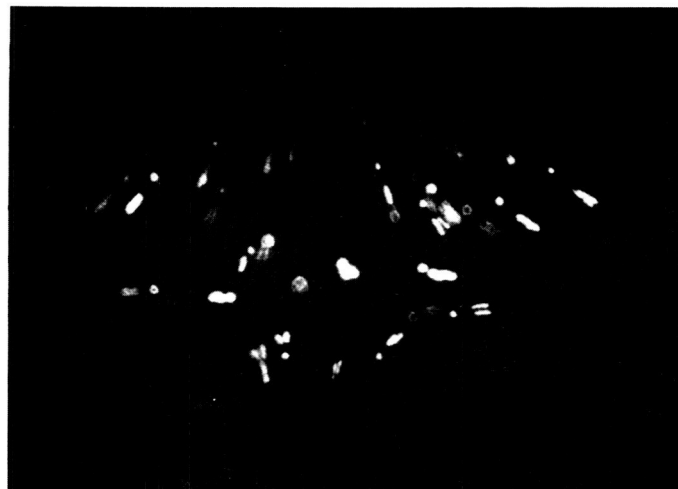


Fig. 6. An  $r$ - $z$  photograph of about 45 alumina particles in the trap described in this article.

# Spin-Lattice Relaxation and the Calculation of Gain, Pump Power, and Noise Temperature in Ruby

J. R. Lyons

Telecommunications Systems Section

*This article describes how a quantitative analysis of the dominant source of relaxation in ruby spin systems is used to make predictions of key maser amplifier parameters. The spin-lattice Hamiltonian which describes the interaction of the electron spins with the thermal vibrations of the surrounding lattice is obtained from the literature. Taking into account the vibrational anisotropy of ruby, Fermi's rule is used to calculate the spin transition rates between the maser energy levels. The spin population rate equations are solved for the spin transition relaxation times, and a comparison with previous calculations is made. Predictions of ruby gain, inversion ratio, and noise temperature as a function of physical temperature are made for 8.4-GHz and 32-GHz maser pumping schemes. The theory predicts that ruby oriented at 90 deg will have approximately 50 percent higher gain in dB and slightly lower noise temperature than a 54.7-deg ruby at 32 GHz (assuming pump saturation). A specific calculation relating pump power to inversion ratio is given for a single channel of the 32-GHz reflected wave maser.*

## I. Introduction

Ruby has been the active material employed in all DSN masers [1], and is again being considered for use at 32 GHz. In order to be able to predict the best operating point (i.e., crystal orientation, magnetic field intensity, and pumping scheme) available at a particular frequency and temperature, it is necessary to understand the relaxation behavior of the spins.

The spins in ruby are  $\text{Cr}^{3+}$  ions introduced as an impurity in the sapphire lattice. Because the spins are localized in the lattice, they obey classical particle statistics

and have a Boltzmann distribution of energies at thermal equilibrium. To achieve maser operation, a spin population inversion is created by pumping the appropriate transitions at their resonance frequencies [2]. The incoming signal is then amplified by inducing stimulated emission between the signal transition levels. The amount of pump power required to maintain the population inversion depends on the rate at which the spins relax back to lower levels. With a knowledge of the relaxation rates, one may compute the corresponding inversion ratio, gain, and noise temperature of the ruby. Conduction losses would, of course, be accounted for separately.

The dominant spin relaxation mechanism in ruby at microwave frequencies and liquid He temperatures is due to interaction of the spins with the lattice, the so-called Kronig-Van Vleck mechanism [13]. Thermal vibrations of the  $\text{Al}_2\text{O}_3$  lattice induce spin transitions via spin-orbit coupling. A so-called magneto-elastic tensor which characterizes the coupling of the  $\text{Cr}^{3+}$  ions to the lattice has been measured at 10 GHz [3]. Assuming the "direct" process of relaxation in which a single quantum of energy is exchanged between the spins and the lattice, Donoho [4] calculated the spin relaxation rates for an isolated spin based on the measured value of the magneto-elastic tensor. Based largely on Donoho's work, the calculation of relaxation rates was performed at 32 GHz for maser pumping schemes of interest.

As a means of comparing theory with experiment, Donoho also calculated the relaxation times associated with the saturation and recovery of a pair of spin levels. The calculation is performed by solving a linear set of first-order differential equations involving the level populations to determine the characteristic times associated with the return to equilibrium. The relaxation times depend on field strength, crystal orientation, and temperature, and are readily compared with measured values. Varying degrees of agreement between theory and experiment have been reported in the literature [5]. Because of the dependence of the measured relaxation times on crystal growth procedure, we are having measurements performed at the University of California, San Diego on Union Carbide ruby at frequencies of 9 GHz and 35 GHz. Note that the measurement of relaxation times is important not only in verifying the theory, but also because shorter relaxation times generally imply greater pump power requirements.

In addition to computing relaxation times, the present work computes the resulting spin populations for various amounts of pump power. Knowing the spin populations, all parameters of interest, such as ruby gain and noise temperature, may be determined. Hence, this theory provides a means of comparing in detail the different operating schemes for a 32-GHz maser at various temperatures. Experiments are being planned to measure inversion ratio, gain, and noise temperature or ruby at the 32-GHz operating points of interest.

It should be pointed out that a significant approximation has been made in assuming that the lattice interacts with an isolated spin. The Cr concentration is actually dilute ( $\sim 0.05$  percent), but if the ions cluster or if significant amounts of magnetic impurities or crystal imperfections are present, the relaxation behavior of the spins can be greatly modified [5]. In this work it is assumed that

spin-spin interactions do not significantly affect the spin relaxation rates, but experiment may prove this to be a poor assumption.

This article proceeds as follows. Section II discusses the spin and spin-lattice Hamiltonians for ruby. Section III outlines the calculation of the spin-lattice transition rate. Section IV describes the solution of the time-dependent rate equations for the relaxation times and compares them with Donoho's [4] results. Section V discusses the calculation of inversion ratio, gain, pump power, and noise temperature. Conclusions are presented in Section VI.

## II. Spin and Spin-Lattice Hamiltonians

### A. Spin Hamiltonian

The spin Hamiltonian formalism [6, 13] is a semiempirical technique for describing the interaction of the electron spin of a paramagnetic ion with the static crystal electric field and applied magnetic field. Evaluation of the spin Hamiltonian yields values for the spin energy levels and rates of spin transition between levels due to the application of an RF magnetic field. The spin-lattice Hamiltonian is also a semiempirical formalism, but it describes the interaction of the paramagnetic ion spin with the time-varying crystal electric field associated with thermal lattice vibrations.

The spin Hamiltonian for ruby is discussed in [8], but a brief review is given below for completeness. The spin Hamiltonians for several other potential maser materials have been evaluated recently [9].

The form of the spin Hamiltonian often reflects the symmetry of the surrounding crystal field. In ruby,  $\text{Cr}^{3+}$  ions enter the  $\text{Al}_2\text{O}_3$  lattice substitutionally for  $\text{Al}^{3+}$ . Each Al/Cr ion is surrounded by a distorted octahedron of  $\text{O}^{2-}$  ions such that the local site symmetry is  $C_3$ . The trigonal axis of the site, which coincides with the  $c$  axis of the ruby, is labeled the  $z$  axis, as shown in Fig. 1. The spin Hamiltonian for ruby is [10]

$$H_s = g\beta\bar{B} \cdot \bar{S} + D \left( S_z^2 - \frac{5}{4} \right) \quad (1)$$

where the spectroscopic splitting factor is approximately isotropic with the value  $g = 1.98$ ,  $\beta$  is the Bohr magneton,  $\bar{B}$  is the applied dc magnetic field, and the crystal field splitting factor is  $D = -5.73$  GHz.  $S_z$  is the  $z$  component of the vector or spin operators,

$$\bar{S} = S_x \hat{x} + S_y \hat{y} + S_z \hat{z} \quad (2)$$

where the  $S_i$  are given in standard texts on quantum mechanics [11].

The first term in Eq. (1) is the Zeeman interaction and the second term represents the interaction of the spin with the static crystal electric field via spin-orbit coupling. Because the ground-state orbital is a singlet, the orbital angular momentum is quenched [13]. Hence, the crystal field splitting may be said to arise from admixture of the spin with the orbital angular momentum of higher-lying orbitals. Applying second-order perturbation theory, one obtains a splitting term that is second order in spin, as shown by Eq. (1). Detailed accounts of the derivation of the form of Eq. (1) based on the equivalent operator method are given elsewhere [8]. The parameter  $D$  is determined empirically, as a first-principles calculation is very difficult, perhaps because the bonding is partly ionic and partly covalent.

Cr is a transition metal element with the electronic configuration  $3d^5 4s^1$ . The  $\text{Cr}^{3+}$  ion has the configuration  $3d^3$ . For the free ion,  $L = 3$  and  $S = 3/2$ , so the term is  ${}^4F_{3/2}$ . A qualitative and simplified energy splitting diagram for Cr in sapphire showing the sequence of degeneracy removing perturbations is given in Fig. 2 [13].

To determine the spin energy levels, each spin state is expressed as a sum of pure spin states, that is, quantum states in which spin-orbit coupling is neglected, as follows:

$$|\Psi_i\rangle = a_i \left| +\frac{3}{2} \right\rangle + b_i \left| +\frac{1}{2} \right\rangle + c_i \left| -\frac{1}{2} \right\rangle + d_i \left| -\frac{3}{2} \right\rangle \quad (3)$$

The eigenvalue equation

$$H_s |\Psi_i\rangle = E_i |\Psi_i\rangle \quad (4)$$

is solved for the energies  $E_i$  and eigenvectors,  $|\Psi_i\rangle$ . Knowing the eigenstates, one may calculate  $W_{ij}$ , the rate of spin transitions stimulated by the time-varying magnetic field  $\bar{H}_1$  between levels  $i$  and  $j$ , from Fermi's rule [2]:

$$W_{ij} = \frac{1}{4} \gamma^2 g_{ij}(f) |\langle \Psi_i | \bar{H}_1^* \cdot \bar{S} | \Psi_j \rangle|^2 \quad (5)$$

where  $\gamma = g\beta\mu_0/\hbar$  and  $g_{ij}(f)$  is the line shape function as a function of frequency. Plots of energy levels versus  $B$

and isofrequency plots of  $B$  versus  $\theta$  for ruby are given by Schultz-duBois [10].

An important point concerning the mixing of pure spin states should be made. For pure spin states, the selection rule  $\Delta M = \pm 1$  applies, where  $M$  is the spin quantum number equal to  $3/2$ ,  $1/2$ ,  $-1/2$ , or  $-3/2$  for  $S = 3/2$ . When spin-orbit interactions are included for a  $\text{Cr}^{3+}$  ion in a crystal, the resulting spin quantum states are a mixture of pure spin states, and as such cannot be labeled by a single value of  $M$ . Therefore, the  $\Delta M = \pm 1$  rule cannot be strictly applied. This mixing of pure spin states is essential to continuous wave (CW) maser operation.

## B. Spin-Lattice Hamiltonian

So far, the time-varying crystal field arising from lattice vibrations has been neglected. To describe the effect lattice vibrations have on the spin states, it is convenient to define a spin-lattice Hamiltonian,  $H_{SL}$  (derivations of the spin-lattice Hamiltonian are given in [14, 15]. Since the lattice vibrations may cause a slightly anisotropic perturbation of the crystal field, a reasonable form for  $H_{SL}$  is [4]

$$H_{SL} = \sum_{ij} d_{ij} S_i S_j, \quad i, j = x, y, z \quad (6)$$

where  $|d_{ij}| \ll |D|$ . For the one-phonon (direct) process of energy exchange, Van Vleck [16] assumed that  $H_{SL}$  was linear in strain. The  $d$  tensor may then be expressed as

$$d_{ij} = \sum_{kl} G_{ijkl} e_{kl}, \quad k, l = x, y, z \quad (7)$$

where  $G$  is a fourth-rank tensor having many of the symmetry properties of the classical elastic stiffness tensor relating stress and strain in a material. The tensor  $e$  is the classical strain given by

$$e_{kl} = \frac{1}{2} \left( \frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k} \right) \quad (8)$$

where  $u_k$  is the displacement of atoms in the direction  $x_k$ . The above form of  $H_{SL}$  neglects spin-spin interactions.

Calculation of the magneto-elastic-tensor  $G$  is of the same order of difficulty as calculation of the spin Hamiltonian parameter  $D$ . Therefore,  $G$  is determined experimentally. In [3]  $G$  was measured by applying a known

uniaxial stress to a ruby crystal and observing the change in the EPR spectrum. The spin-lattice Hamiltonian given by Eqs. (6) and (7) is treated as a static perturbation to the energy levels computed from the spin Hamiltonian. The measurements were performed at 10.1 GHz at room temperature. Line shifts were in the range of 10 to 50 Gauss when the crystals were strained. Five crystal samples were employed, each having a different stress axis. The components of  $G$  were obtained by least-squares analysis of the data. The Cr concentration of the samples was thought to be  $\approx 0.05$  percent. At this concentration, spin-spin interactions may have a non-negligible effect on the spectra, so

the  $G$  tensor values could be somewhat in error; however, this possibility has not been investigated.

The number of independent components in  $G$  can be reduced by symmetry arguments. The form of a fourth-rank tensor with  $C_3$  symmetry is given in [17], and is shown to have twelve independent components. Choosing trace ( $H_{SL}$ ) = 0, which shifts the corresponding energy levels by an insignificant constant, reduces the number of independent components to ten. The  $G$  tensor satisfies  $G_{ijkl} = G_{jikl} = G_{ijlk}$ , and so may be written in Voigt notation by defining the indices  $1 = xx$ ,  $2 = yy$ ,  $3 = zz$ ,  $4 = yz$ ,  $5 = xz$ ,  $6 = xy$ . In Voigt notation  $G$  has the form

$$G = \begin{bmatrix} G_{11} & G_{12} & -G_{33}/2 & G_{14} & -G_{25} & G_{16} \\ G_{12} & G_{11} & -G_{33}/2 & -G_{14} & G_{25} & -G_{16} \\ -(G_{11} + G_{12}) & -(G_{11} + G_{12}) & G_{33} & 0 & 0 & 0 \\ G_{41} & -G_{41} & 0 & G_{44} & G_{45} & G_{52} \\ -G_{52} & G_{52} & 0 & -G_{45} & G_{44} & G_{41} \\ -G_{16} & G_{16} & 0 & G_{25} & G_{14} & \frac{1}{2}(G_{11} - G_{12}) \end{bmatrix} \quad (9)$$

The  $G$  values (in GHz) used in [4] and in this work are the following:

$$\begin{aligned} G_{11} &= 124.6 & G_{41} &= -15.0 \\ G_{12} &= -35.8 & G_{25} &= 45.0 \\ G_{33} &= 181.2 & G_{52} &= 45.0 \\ G_{44} &= 54.0 & G_{16} &= 0 \\ G_{14} &= -15.0 & G_{45} &= 0 \end{aligned}$$

The factor of  $\hbar^2$  from the  $S_i S_j$  term is absorbed into the  $G$  values.

For interactions with the  $\text{Al}_2\text{O}_3$  lattice, the  $\text{Cr}^{3+}$  ions can be considered to occupy two inequivalent sites, as discussed in [3]. These sites are related by a 2-fold rotation about the  $x$  axis. For the two sites all the elements of  $G$  have the same magnitude, but  $G_{25}$ ,  $G_{52}$ ,  $G_{16}$ , and  $G_{45}$  have opposite signs. In [4] and in this work only one Cr site, specified by the elements of  $G$  given above, is ac-

counted for. Computer experiments have shown that the second site does not lead to significantly different relaxation times or inversion ratios.

The tensor  $G$  characterizes the effect a given strain has on the EPR spectrum. To relate  $G$  to the thermal lattice vibrations one must know the strain associated with a given lattice mode and also the number of phonons present (i.e., the phonon occupation number). Since the phonon wavelengths of interest are  $\sim 10^{-4}$  cm (for a sound velocity  $\sim 10^6$  cm/sec), the atoms may be assumed to undergo displacements of equal amplitude in any unit cell. (This is an approximation since the different mass ions would undergo slightly different displacements, thus modifying the local crystal field.) Then, for phonons of wavevector  $\vec{k}$  and polarization vector  $\hat{\epsilon}_p$ , the displacement may be written as

$$\vec{u}(\vec{r}) = \left( \frac{\hbar}{2M\omega} \right)^{1/2} \left( a_{\vec{k},p}^+ - a_{-\vec{k},p} \right) \hat{\epsilon}_p e^{i\vec{k}\cdot\vec{r}} \quad (10)$$

where  $M$  is the crystal mass,  $\omega$  is the angular frequency, and  $a$  and  $a^+$  are the phonon annihilation and creation operators. The phonon operators arise from an analogy with

the quantum harmonic oscillator, as discussed in any solid state physics text [18]. They have the following properties:

$$a_{k,p}^+ |n_{k,p}\rangle = \sqrt{n_{k,p} + 1} |n_{k,p} + 1\rangle \quad (11)$$

$$a_{k,p} |n_{k,p}\rangle = \sqrt{n_{k,p}} |n_{k,p} - 1\rangle \quad (12)$$

where  $n_{k,p}$  is the phonon occupation number for the state  $k, p$ , and where subscript  $k$  is the mode wavevector and  $p$  is the mode polarization (two transverse and one longitudinal). From Eqs. (8) and (10), the strain becomes

$$e_{ij} = \left(\frac{\hbar}{8M\omega}\right)^{1/2} (a_{k,p}^+ - a_{-k,p}) (\epsilon_{p,i} k_j + \epsilon_{p,j} k_i) e^{i\vec{k}\cdot\vec{r}} \quad (13)$$

where  $i, j = x, y, z$ .

### III. The Spin-Lattice Transition Rate

In determining the  $G$  tensor, the spin-lattice Hamiltonian is static since the applied strain is static. However, in modeling the interaction of the Cr spin with the thermal lattice vibrations, the strain is time periodic. The resulting  $H_{SL}$  is therefore a dynamical Hamiltonian requiring the application of time-dependent perturbation theory, specifically Fermi's rule.

We are interested in lattice modes that are on "speaking terms" with the spins, that is, vibrating at frequencies that correspond to spin transitions. Fermi's rule is used to calculate the transition probability rate,  $\omega_{ij}$ , for spin transitions from level  $i$  to level  $j$  due to spin-lattice interaction. Since ruby is vibrationally anisotropic, an integration over all possible phonon directions is performed. The three phonon polarizations (one longitudinal and two transverse) and the phonon density of states are also accounted for. Applying Fermi's rule,

$$\omega_{ij} = \frac{2\pi}{\hbar^2} \int_{4\pi} \sum_{p=1}^3 \left| \langle \Psi_i, n_{k,p} | H_{SL} | \Psi_j, n_{k,p} + 1 \rangle \right|^2 \times \rho(\omega) \frac{d\Omega}{4\pi} \quad (14)$$

where  $|\Psi_i\rangle$  is the state vector for level  $i$  as determined from Eq. (4),  $|n_{k,p}\rangle$  is the  $n$ th phonon occupation state for a phonon with wavevector  $k$  and polarization  $p$ ,  $\rho(\omega) = \partial N / \partial \omega$  is the density of phonon states, as a function of frequency, and  $d\Omega$  is an infinitesimal solid angle. In Eq. (14), the perturbing Hamiltonian  $H_{SL}$  changes the quantum state from that of a spin in state  $i$  and  $n$  phonons in state  $k, p$  to that of a spin in state  $j$  and  $n + 1$  phonons in state  $k, p$ . This corresponds to the creation of a phonon when a spin drops from a higher energy state to a lower energy state.

From [18], the density of states for each polarization in a periodic solid is

$$\rho(\omega) = V \frac{k^2}{2\pi^2} \frac{dk}{d\omega} \quad (15)$$

where  $V$  is the crystal volume. Making the Debye approximation in which the velocity of sound  $v$  is a constant for each polarization type, the dispersion relation is simply  $\omega = vk$ . Equation (15) then becomes

$$\rho(\omega) = \frac{V\omega^2}{2\pi^2 v^3} \quad (16)$$

The Debye temperature of sapphire is 980 K [19], which corresponds to a frequency of  $\approx 2 \times 10^{13}$  Hz, so the Debye approximation is assumed to be valid for the one-phonon relaxation process. Anharmonic crystal interactions and multiple phonon processes are neglected.

Consider now the matrix element in Eq. (14). Substituting Eqs. (6) and (7) for  $H_{SL}$  in Eq. (14), and separating spin and phonon operators, one obtains

$$\begin{aligned} \langle \Psi_i, n_{k,p} | H_{SL} | \Psi_j, n_{k,p} + 1 \rangle &= \left(\frac{\hbar}{8M\omega}\right)^{1/2} \\ &\times \langle n_{k,p} | a_{k,p}^+ - a_{-k,p} | n_{k,p} + 1 \rangle e^{i\vec{k}\cdot\vec{r}} \\ &\times \sum_{mnst} G_{mnst} \langle \Psi_i | S_m S_n | \Psi_j \rangle (\epsilon_{p,s} k_t + \epsilon_{p,t} k_s) \end{aligned} \quad (17)$$

where  $m, n, s, t = x, y, z$ . Applying Eqs. (11) and (12), the phonon matrix element becomes

$$\langle n_{k,p} | a_{k,p}^+ - a_{-k,p} | n_{k,p} + 1 \rangle = -\sqrt{n_{k,p} + 1} \quad (18)$$

For a large number of spins,  $n_{k,p}$  may be replaced by the average phonon occupation number as given by a Bose-Einstein distribution of phonons (spin-one quasi-particles) at the lattice temperature  $T$ ,

$$(n_{k,p})_{avg} = \frac{1}{e^{\frac{\hbar\omega}{kT}} - 1} \quad (19)$$

Substituting Eqs. (16) through (19) into Eq. (14), one obtains

$$\begin{aligned} \omega_{ij} = & \frac{\omega}{32\pi^2\rho_c\hbar} \frac{1}{1 - e^{-\frac{\hbar\omega}{kT}}} \int_{4\pi} \sum_{p=1}^3 \\ & \times \left\{ \left| \sum_{mnst} G_{mnst} \langle i | S_m S_n | j \rangle \right. \right. \\ & \left. \left. \times (\epsilon_{p,s} k_t + \epsilon_{p,t} k_s) \right|^2 \frac{1}{v_{k,p}^3} \right\} d\Omega \quad (20) \end{aligned}$$

where  $\Psi_i$  and  $\Psi_j$  have been denoted by  $i$  and  $j$ ,  $\rho_c$  is the crystal density, and  $v_{k,p}$  is the phonon velocity for mode  $k, p$ . Equation (20) is Eq. (11) in [4].

In order to simplify Eq. (20), it is convenient to use Voigt notation and to make some new variable definitions. Define the direction cosines of the wavevector  $k$  as follows:

$$l_i = \frac{k_i}{k}, \quad i = x, y, z \quad (21)$$

Writing  $G$  in Voigt notation, the summation over  $m, n, s, t$  in Eq. (20) may be written as

$$\begin{aligned} \sum_{mnst} G_{mnst} \langle i | S_m S_n | j \rangle (\epsilon_{p,s} k_t + \epsilon_{p,t} k_s) = \\ k \sum_{u,v=1}^6 G_{uv} \langle i | S_u^2 | j \rangle L_{p,v} \quad (22) \end{aligned}$$

where  $S_u^2$  and  $L_{p,v}$  are defined as follows:

$$\begin{aligned} S_1^2 = S_x S_x & & S_2^2 = S_y S_y \\ S_3^2 = S_z S_z & & S_4^2 = S_y S_z + S_z S_y \\ S_5^2 = S_z S_x + S_x S_z & & S_6^2 = S_x S_y + S_y S_x \end{aligned} \quad (23)$$

and

$$\begin{aligned} L_{p,1} = 2\epsilon_{p,x} l_x & & L_{p,2} = 2\epsilon_{p,y} l_y \\ L_{p,3} = 2\epsilon_{p,z} l_z & & L_{p,4} = 2(\epsilon_{p,y} l_z + \epsilon_{p,z} l_y) \\ L_{p,5} = 2(\epsilon_{p,z} l_x + \epsilon_{p,x} l_z) & & L_{p,6} = 2(\epsilon_{p,x} l_y + \epsilon_{p,y} l_x) \end{aligned} \quad (24)$$

Using the dispersion relation  $\omega = 2\pi f = kv_{k,p}$ , where  $f$  is the frequency of the spin transition between levels  $i$  and  $j$ , and substituting Eqs. (21) through (24) into Eq. (20) yields

$$\begin{aligned} \omega_{ij} = & \frac{\pi^2}{2\rho_c\hbar} \frac{f^3}{1 - e^{-\frac{\hbar f}{kT}}} \int_{4\pi} \sum_{p=1}^3 \\ & \times \left| \sum_{u,v=1}^6 G_{uv} \langle i | S_u^2 | j \rangle L_{p,v} \right|^2 \frac{1}{v_{k,p}^5} d\Omega \quad (25) \end{aligned}$$

Defining

$$A_{p,u} = \sum_{v=1}^6 G_{uv} L_{p,v} \quad (26)$$

the modulus-squared quantity in Eq. (25) can be written as

$$\begin{aligned} \left| \sum_{u,v=1}^6 G_{uv} \langle i | S_u^2 | j \rangle L_{p,v} \right|^2 = \\ \sum_{q,r=1}^6 A_{p,q} A_{p,r} \langle i | S_q^2 | j \rangle \langle i | S_r^2 | j \rangle^* \quad (27) \end{aligned}$$

Substituting Eq. (27) into Eq. (25) and interchanging the integral and the sum over polarization with the sum over  $q$  and  $r$ , Eq. (25) becomes



$$\omega_{ij} = \frac{\pi^2}{2\rho_c h} \frac{f^3}{1 - e^{-\frac{hf}{kT}}} \sum_{q,r=1}^6 B_{qr} \langle i | S_q^2 | j \rangle \langle i | S_r^2 | j \rangle^* \quad (28)$$

where

$$B_{qr} = \int_{4\pi} \sum_{p=1}^3 A_{p,q} A_{p,r} \frac{1}{v_{k,p}^5} d\Omega \quad (29)$$

Equation (29) is evaluated by a numerical integration over  $4\pi$  steradians. Specifically,

$$\int_{4\pi} \{ \} d\Omega = \int_0^{2\pi} \int_0^\pi \{ \} \sin\theta d\theta d\phi \quad (30)$$

where the double integral is evaluated as a double sum. Numerical experimentation shows that the double sum converges for a summation step of  $\Delta\theta = \Delta\phi = 5$  deg.

The density of sapphire is  $4.0 \text{ gm/cm}^3$  [20]. For  $G$  and  $f$  specified in units of GHz,  $v_{k,p}$  in units of  $10^5 \text{ cm/sec}$ , and  $S_q^2$  dimensionless, the constant in front of Eq. (28) becomes  $8.2 \times 10^{-7}$  and  $\omega_{ij}$  has the units of  $\text{sec}^{-1}$ .

In order to evaluate Eqs. (24) and (29), one needs the phonon velocity and polarization vector for the three phonon polarizations for each wavevector direction in the  $4\pi$  steradians. These are the quantities of interest in the so-called Christoffel equations [21], which describe elastic wave propagation in an anisotropic elastic medium. The equations are a coupled set of three equations of motion for volume elements of the crystal; Kittel [18] discusses the case of cubic symmetry.

Wachtman et al. [22] gives the appropriate equations for single-crystal  $\text{Al}_2\text{O}_3$ , which has trigonal symmetry. The elastic constants,  $c_{ij}$ , (Voigt notation) are defined by

$$\sigma_i = \sum_{j=1}^6 c_{ij} \epsilon_j \quad (31)$$

where  $\sigma$  is the stress tensor and  $\epsilon$  is the strain tensor. Values for the six independent elements of  $c_{ij}$  are determined at room temperature by a resonance method discussed in [22].

For the coordinate system of Fig. 1, the equations of motion for an elastic wave with frequency  $\omega = k \cdot v$  may be written as

$$\begin{bmatrix} A & H & G \\ H & B & F \\ G & F & C \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \rho_c v^2 \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \quad (32)$$

where  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$  are the material displacements which taken together specify the polarization vector. The matrix elements  $A, B, C$  and  $F, G, H$  are functions of the  $c_{ij}$  and the wavevector direction cosines:

$$\begin{aligned} A &= c_{11}l^2 + \frac{1}{2}(c_{11} - c_{12})m^2 + c_{44}n^2 + 2c_{14}mn \\ B &= \frac{1}{2}(c_{11} - c_{12})l^2 + c_{11}m^2 + c_{44}n^2 - 2c_{14}mn \\ C &= c_{44}(l^2 + m^2) + c_{33}n^2 \\ F &= c_{14}(l^2 - m^2) + (c_{13} + c_{44})mn \\ G &= 2c_{14}lm + (c_{13} + c_{44})ln \\ H &= \frac{1}{2}(c_{11} + c_{12})lm + 2c_{14}ln \end{aligned} \quad (33)$$

where  $l = \sin\theta \cos\phi$ ,  $m = \sin\theta \sin\phi$ , and  $n = \cos\theta$ . For each wavevector direction in Eq. (30), Eq. (32) is solved for the phonon velocity and polarization, as given by the eigenvalues and eigenvectors, respectively, of the matrix in Eq. (32).

The  $B_{qr}$  quantities defined in Eq. (29) need only be calculated once for each Cr site; the resulting values are stored in computer memory. For Cr site number 1 (i.e., for  $G$  given by Eq. 9) the matrix  $B$  has the form

$$B = \begin{bmatrix} B_{11} & B_{12} & B_{13} & B_{14} & B_{15} & B_{16} \\ B_{12} & B_{11} & B_{13} & -B_{14} & -B_{15} & -B_{16} \\ B_{13} & B_{13} & B_{33} & 0 & 0 & 0 \\ B_{14} & -B_{14} & 0 & B_{44} & 0 & -B_{15} \\ B_{15} & -B_{15} & 0 & 0 & B_{44} & B_{14} \\ B_{16} & -B_{16} & 0 & -B_{15} & B_{14} & B_{66} \end{bmatrix} \quad (34)$$

where the independent components have the following values:

$$\begin{aligned}
B_{11} &= 32.2 & B_{16} &= -0.0030 \\
B_{12} &= -13.9 & B_{33} &= 36.6 \\
B_{13} &= -18.3 & B_{44} &= 11.7 \\
B_{14} &= -3.67 & B_{66} &= 23.1 \\
B_{15} &= -14.04 & &
\end{aligned} \tag{35}$$

For Cr site number 2 the  $B_{qr}$  have the same magnitude, but  $B_{15}$  and  $B_{16}$  have opposite signs. For the  $B_{qr}$  given above,  $\omega_{ij}$  is calculated from Eq. (28) for each field intensity, angle, transition, and temperature of interest.

Finally, consider the frequency and temperature dependence of  $\omega_{ij}$ . From Eq. (28), the temperature dependence is determined by the factor  $(1 - e^{-hf/kT})^{-1}$ , for which

$$\left(1 - e^{-\frac{hf}{kT}}\right)^{-1} \approx \begin{cases} \frac{kT}{hf}, & \frac{hf}{kT} \ll 1 \\ 1, & \frac{hf}{kT} \gg 1 \end{cases} \tag{36}$$

Thus, for  $hf/kT \ll 1$ ,  $\omega_{ij}$  is proportional to temperature, or, since the relaxation time  $T_1$  is  $\sim \omega_{ij}^{-1}$ ,  $T_1 \sim T^{-1}$ . A relaxation time inversely proportional to physical temperature is a characteristic of the one-phonon relaxation process, and has been observed for many paramagnetic materials (including ruby) at liquid He temperatures [13]. The multiple-phonon relaxation processes exhibit relaxation times with significantly different dependence on temperature, e.g.,  $T_1 \propto T^{-7}$  (Raman process).

The frequency dependence of  $\omega_{ij}$  is contained in the factor  $f^3 \cdot (1 - e^{-hf/kT})^{-1}$  and also in the mixing of spin states.

Also note the strong dependence of  $\omega_{ij}$  on the velocity of sound. From Eqs. (28) and (29),  $\omega_{ij} \sim v^{-5}$ , so if long relaxation times are desired, a hard material such as sapphire is beneficial.

## IV. Spin-Lattice Relaxation Times

The spin-lattice relaxation times,  $T_1$ , are of interest primarily because they are more accessible to measurement than the transition probability rates. The relaxation times are calculated by solving a set of time-dependent rate equations for the spin populations. The rate equations, not including pump-induced transition rates, for the  $i$ th level are [2, 4]

$$\frac{dn_i(t)}{dt} = \sum_{j=1}^4 (\omega_{ji}n_j(t) - \omega_{ij}n_i(t)), \quad i = 1, \dots, 4 \tag{37}$$

where  $n_i(t)$  is the instantaneous spin population of level  $i$ . Conservation of the number of spins requires that

$$n_1 + n_2 + n_3 + n_4 = N \tag{38}$$

where  $N$ , the total number of spins, is constant. Equations (37) and (38) reduce to a system of three equations,

$$\dot{\bar{n}} = A\bar{n} + \bar{b} \tag{39}$$

where

$$\bar{n} = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} \tag{40}$$

$$\bar{b} = \begin{bmatrix} \omega_{41} \cdot N \\ \omega_{42} \cdot N \\ \omega_{43} \cdot N \end{bmatrix} \tag{41}$$

and

$$A = \begin{bmatrix} A_{11} & \omega_{21} - \omega_{41} & \omega_{31} - \omega_{41} \\ \omega_{12} - \omega_{42} & A_{22} & \omega_{32} - \omega_{42} \\ \omega_{13} - \omega_{43} & \omega_{23} - \omega_{43} & A_{33} \end{bmatrix} \tag{42}$$

where

$$\begin{aligned}
A_{11} &= -(\omega_{12} + \omega_{13} + \omega_{14} + \omega_{41}) \\
A_{22} &= -(\omega_{21} + \omega_{23} + \omega_{24} + \omega_{42}) \\
A_{33} &= -(\omega_{31} + \omega_{32} + \omega_{34} + \omega_{43})
\end{aligned} \tag{43}$$

Note that

$$\omega_{ji} = e^{\frac{-hf}{kT}} \omega_{ij} \quad (44)$$

where  $f = f_i - f_j$ , as can be shown by manipulation of Eq. (28).

Experimentally, the usual procedure for measuring  $T_1$  is to saturate a pair of levels with a strong microwave pulse and observe the recovery of the same or another pair of levels as the spins return to thermal equilibrium. For non-interacting spins (i.e., no cross-relaxation) the recovery is exponential, with  $T_1$  being identified as the time constant of the decay. A theoretical  $T_1$  is determined by solving Eq. (39) for an appropriate set of initial conditions. For example, if levels one and two are saturated, then the initial conditions would be

$$\bar{n}(t=0) = \begin{bmatrix} \frac{1}{2}(N_1 + N_2) \\ \frac{1}{2}(N_1 + N_2) \\ N_3 \end{bmatrix} \quad (45)$$

where  $N_i$  is the thermal equilibrium spin population of level  $i$ . The  $N_i$  satisfy a Boltzmann distribution,

$$\frac{N_i}{N_j} = e^{\frac{-hf}{kT}} \quad (46)$$

and also satisfy the conservation relation,

$$N_1 + N_2 + N_3 + N_4 = N \quad (47)$$

The solution to Eq. (39) is straightforward and may be found in [23]:

$$\bar{n}(t) = \alpha_1 \bar{c}_1 e^{\lambda_1 t} + \alpha_2 \bar{c}_2 e^{\lambda_2 t} + \alpha_3 \bar{c}_3 e^{\lambda_3 t} - A^{-1} \bar{b} \quad (48)$$

where

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = (\bar{c}_1 \ \bar{c}_2 \ \bar{c}_3)^{-1} \cdot (\bar{n}(0) + A^{-1} \bar{b}) \quad (49)$$

and the  $\bar{c}_i$  are the eigenvectors of  $A$  (written as column vectors) with the corresponding eigenvalues  $\lambda_i$ . The normalized population difference between levels  $i$  and  $j$  is

$$\frac{n_i - n_j}{N_i - N_j} = 1 + A_1 e^{-t/T_1} + A_2 e^{-t/T_2} + A_3 e^{-t/T_3} \quad (50)$$

where the  $T_i = -1/\lambda_i$  are the relaxation times, and the  $A_i$  are determined from the  $\alpha_i$  and  $\bar{c}_i$ . Thus, the recovery of each pair of levels is characterized by three relaxation times. In fact, the same three relaxation times characterize the recovery between any pair of levels for a given initial condition; however, the amplitudes,  $A_i$ , will vary for the different transitions.

Donoho [4] plots  $T_i$  and  $A_i$  as a function of  $\theta$  and transition for a frequency  $f = 9.3$  GHz and a physical temperature  $T = 4.2$  K. In [4] there are also several plots of  $T_1$  versus frequency (1–10 GHz) for particular angles and transitions. Generally it was found that only one or two relaxation times were important, usually the longest two. These calculations were repeated as a check on the computer code. Table 1 compares two sets of  $T_i$  and  $A_i$  values from [4] with the author's calculations for an angle  $\theta = 54.7$  deg. The  $T_i$  from the author's calculations are approximately a factor of one-half less than Donoho's results, which leads to better agreement between theory and experiment at 10 GHz [5, 24]. More troublesome, however, is the lack of consistent agreement among the  $A_i$  values. Although the cause of the disagreement is unknown, the disagreement is not considered significant in light of the consistent relaxation times computed.

Previous comparisons of experimental and theoretical values of  $T_1$  have generally not shown good agreement even for very dilute ruby ( $\sim 0.01$  percent) [5]. For example, the relaxation times may be several times shorter than the theory predicts and may not have the angular dependence displayed by Donoho's results. However, Standley and Vaughan [24] made measurements on ruby grown by the so-called vapor phase modification to the flame-fusion technique, and found considerably better agreement with Donoho's theory for concentrations up to 0.2 percent. It is thus likely that the source of the disagreement for the rubies not grown by the vapor phase method is dependent on the crystal growth process. If the Cr ions are not uniformly distributed, then cross-relaxation between exchange-coupled pairs or clusters is a possibility. Crystal imperfections and other magnetic impurities (e.g.,  $\text{Fe}^{3+}$ ) could similarly affect the relaxation behavior of the Cr ions. Standley and Vaughan [5] discuss these possibilities further. To account for these effects, Donoho's theory

would most likely have to be modified to include spin-spin interactions. Previous attempts to incorporate spin-spin effects have not been very successful [30].

Professor Sheldon Shultz at the University of California, San Diego is presently under contract with JPL to perform  $T_1$  measurements at 9 GHz and 35 GHz on our Czochralski-grown ruby. These measurements are being made to determine the optimum Cr concentration in a 32-GHz traveling wave maser (TWM) as a function of temperature, and also to better understand the relaxation mechanism in ruby. Previous measurements [2] have indicated a strong dependence of inversion ratio and gain on Cr concentration, particularly at very low temperatures (e.g.,  $\sim 2$  K). This effect is a result of spin-spin interactions, and cannot be explained by Donoho's theory.

## V. Calculation of Gain, Pump Power, and Noise Temperature for Ruby

Gain, pump power requirements, and noise temperature are important considerations in maser design and operation. In principle, each of these quantities can be predicted given an accurate knowledge of the spin-lattice transition probability rates,  $\omega_{ij}$ .

### A. Inversion Ratio

The gain in the ruby is directly proportional to the population difference of the inverted levels. The steady-state populations of the levels can be determined as a function of pump power from the steady-state rate equations:

$$\dot{n}_i = 0 = \sum_{j=1}^4 [\omega_{ji}n_j - \omega_{ij}n_i + W_{ij}(n_j - n_i)]$$

$$i = 1, 2, 3, 4 \quad (51)$$

where  $W_{ij}$  is the stimulated transition rate due to RF pumping given by Eq. (5). Applying conservation of spins, Eq. (51) may be reduced to a set of three equations,

$$A'\bar{n} = \bar{b} \quad (52)$$

where  $\bar{n}$  is given by Eq. (40),

$$\bar{b} = -N \begin{bmatrix} \omega_{41} \\ \omega_{42} + W_{24} \\ \omega_{43} + W_{34} \end{bmatrix} \quad (53)$$

and

$$A' = A + \begin{bmatrix} -(W_{12} + W_{13}) & W_{12} & W_{13} \\ (W_{12} - W_{24}) & -(W_{12} + 2W_{24}) & -W_{24} \\ (W_{13} - W_{34}) & -W_{34} & -(W_{13} + 2W_{34}) \end{bmatrix} \quad (54)$$

where  $A$  is given by Eq. (42). The population vector  $\bar{n}$  is determined by matrix inversion. Note that the  $W_{ij}$  terms present in  $\bar{b}$  and  $A'$  are restricted to  $W_{13}$ ,  $W_{24}$ ,  $W_{12}$ , and  $W_{34}$ , corresponding to the transitions that would be pumped in the double pumping schemes shown in Fig. 3. (The choice of these pumping schemes is discussed in [9].)

It is assumed that the signal transition is far from saturation, so  $W_{\text{signal}}$  can be neglected compared to  $W_{\text{pump}}$  and  $\omega_{ij}$ . (Actually,  $W_{\text{signal}}$  could be  $> \omega_{ij}$  for some transitions, but it is assumed this has negligible effect on the populations.)

A convenient measure of the inverted population difference is the inversion ratio, defined for the signal transition occurring between levels  $i$  and  $j$  ( $i > j$ ) as

$$I = \frac{n_i - n_j}{N_j - N_i} \quad (55)$$

where  $n_i$  and  $n_j$  are determined from Eq. (52). Plots of  $I$  versus  $W_{\text{pump}}$  may be made with temperature, frequency, angle, etc. as parameters.

Figure 4 shows  $I$  versus  $W_{\text{pump}}$  for  $f = 32$  GHz,  $\theta = 54.7$  deg, and for a range of temperatures. Figure 5 is a similar family of curves for  $\theta = 90$  deg. Note the opposite dependence of  $I$  on temperature for the two pumping schemes when the pumping rates are near saturation. This is a result of the settling of spins into the lowest energy state as  $hf/kT$  increases.

The inversion ratio corresponding to saturation of the pumped transitions can easily be read off of plots of  $I$  versus  $W_{\text{pump}}$ . However, if pump saturation is expected, then the saturation inversion ratio can be determined directly from a single steady-state rate equation. A simple algebraic equation then results for  $I$ .

Figure 6 shows  $I$  versus temperature assuming pump saturation for the 8.5-GHz maser pumping scheme [1]. Siegman [2] defines the ‘‘optimum’’ and ‘‘equal’’ inversion ratios by rewriting the expression for  $I$  in terms of a pseudo-relaxation time,

$$T_{ij} = (\omega_{ij} + \omega_{ji})^{-1} \quad (56)$$

and then either optimizing  $I$  with respect to the  $T_{ij}$ , or simply setting all the  $T_{ij}$  equal and calculating the resulting  $I$ . (Siegman does not discuss the theory of spin-lattice relaxation quantitatively, and therefore has to make assumptions about the  $\omega_{ij}$ .) Note that in Fig. 6 the inversion ratio for saturation is close to the optimum value possible for this pumping scheme. Figure 6 will be compared with measurements being performed at JPL.

Figures 7 and 8 show predicted values  $I$  versus temperature for two pumping schemes for a 32-GHz maser. Note that the assumption of pump saturation may not be correct for the higher pump frequencies.

Many measurements of inversion ratio have been performed by R. Clauss et al. [25, 26] at JPL. Table 2 compares some of these measured values with predictions from the spin-lattice relaxation theory. In most cases, it can be assumed that the pump transitions were saturated. The agreement between measurement and theory is moderate, with predicted values often being 30 percent or more different from the measured values.

Inversion ratio measurements have also been performed by Moore and Neff [27] at  $\theta = 54.7$  deg. Their results are shown in Fig. 9 with the theoretical value overlaid. The agreement between specific values varies, but the trend of decreasing inversion ratio with increasing signal frequency is predicted by the theory. Note the large scatter in Moore and Neff’s measured values of  $I$  at frequencies below 20 GHz.

The source of disagreement between the measured and predicted values of  $I$  is most likely due to spin-spin interactions, as discussed in Section IV.

## B. Gain of Ruby

Expressions for maser gain are given in [2]. Considering only first-order temperature-dependent terms, the gain in dB is

$$G_{\text{dB}} \propto \Delta n_{\text{signal}} \cdot \sigma_{\text{signal}}^2 \quad (57)$$

where  $\Delta n_{\text{signal}} = I \Delta N_{\text{signal}}$  and where  $\Delta N_{\text{signal}}$  is determined from the Boltzmann relations. Figure 10 shows  $G_{\text{dB}}$  (normalized to  $G_{\text{dB}}$  at 4.2 K) versus temperature for the JPL X-band maser pumping scheme. Since  $hf/kT \ll 1$  (approximately) even at  $T = 1.5$  K, little departure from the expected  $T^{-1}$  behavior (dashed line) is seen [2].

For  $f = 32$  GHz,  $hf/kT \approx 1$  for  $T = 1.5$  K, so some deviation of  $G_{\text{dB}}$  from a  $T^{-1}$  dependence is expected. Figure 11 shows the thermal equilibrium population difference (normalized to the population difference at 4.2 K) for the maser configurations shown in Figs. 3(a) and 3(b). Note the strong departure from the  $T^{-1}$  curve for the pumping schemes employing the 2-3 signal transition. However, the temperature dependence of  $G_{\text{dB}}$  is proportional to  $I \cdot \Delta N$ . Figure 7 displays an inversion ratio that increases as  $T$  decreases, so the product  $I \cdot \Delta N$  will not deviate greatly from a  $T^{-1}$  dependence.

Figure 12 shows normalized gain as a function of temperature for the two pumping scenarios corresponding to Figs. 7 and 8. Both curves are normalized to the gain at 4.2 K for the scenario employing the 2-3 signal transition, and both assume pump saturation. The plots indicate that the 90-deg scheme (1-2 signal transition) will yield greater ruby gain for all temperatures in the range of 1.5 K to 4.2 K.

## C. Pump Power

After the pump transition rate  $W_{ij}$  is determined for a given inversion ratio, the corresponding pump power can be calculated from  $W_{ij}$ . To simplify the expression for  $W_{ij}$  as given by Eq. (5), Siegman [2] defines the vector

$$\bar{\sigma}_{ij} = \langle \Psi_i | \bar{S} | \Psi_j \rangle \quad (58)$$

where  $\bar{\sigma}_{ij}$  is a measure of the strength of the  $i, j$  transition. Equation (5) then becomes

$$W_{ij} = \frac{1}{4} \gamma^2 g(f) \left| \bar{H}_1^* \cdot \bar{\sigma}_{ij} \right|^2 \quad (59)$$

The maximum value of  $\sigma_{ij}^2$  occurs when  $\bar{H}_1$  and  $\bar{S}$  are parallel. Siegman further defines

$$\bar{\sigma}_{ij} = \frac{1}{2} (\alpha_{ij} \hat{x} + \sqrt{-1} \beta_{ij} \hat{y} + \gamma_{ij} \hat{z}) \quad (60)$$

The vector components  $\alpha_{ij}$ ,  $\beta_{ij}$ ,  $\gamma_{ij}$  are determined from the spin Hamiltonian. The optimum polarization of  $\bar{H}_1$  is then easily found by maximizing  $\bar{H}_1^* \cdot \bar{\sigma}_{ij}$ .

Rather than outlining a general procedure for calculating pump power, a specific calculation for a single channel of the 32-GHz reflected wave maser (RWM) [28] is given. The RWM consists of 8 channels of ruby-filled waveguide in which both the signal and pump power propagate in the TE<sub>10</sub> mode. (Note that at the pump frequency the waveguide can support higher-order modes; here it is assumed that none of these is excited.)

The geometry of a single waveguide channel is shown in Fig. 13. Note that the  $z$  axis is defined to be parallel to the dc field  $\bar{B}$ , unlike Fig. 1 in which the  $z$  axis is parallel to the crystal  $c$  axis. Thus, the  $\bar{\sigma}_{ij}$  vectors computed from the spin Hamiltonian given by Eq. (1) must be rotated through an angle  $\theta = 54.7$  deg to apply to Fig. 13.

The RWM employs the push-pull pumping scheme (Fig. 3b) at the double-pump angle  $\theta = 54.7$  deg. The pump transitions are the 1-3 and 2-4 at 66 GHz, and the signal transition is the 2-3 at 32 GHz. The  $\bar{\sigma}_{ij}$  components are the following:

$$\begin{aligned} \alpha_{13} &= -0.291 & \beta_{13} &= 0.303 & \gamma_{13} &= -0.198 \\ \alpha_{24} &= -0.257 & \beta_{24} &= 0.236 & \gamma_{24} &= -0.198 \\ \alpha_{23} &= -1.96 & \beta_{23} &= 1.96 & \gamma_{23} &= 0.085 \end{aligned} \quad (61)$$

Concentrating on the 1-3 pump transition, note that

$$\bar{\sigma}_{13} \approx \frac{\alpha_{13}}{2} \left( -\hat{x} + j\hat{y} - \frac{2}{3}\hat{z} \right) \quad (62)$$

where  $\alpha_{13} \approx 0.30$ .

The magnetic field of the TE<sub>10</sub> rectangular waveguide mode is [29]

$$\begin{aligned} \bar{H}_1 &= \frac{E_0}{Z_d} \left( j \frac{\lambda_d}{2a} \cos \frac{\pi y}{a} \hat{x} + \sqrt{1 - \left( \frac{\lambda_d}{2a} \right)^2} \sin \frac{\pi y}{a} \hat{y} \right) \\ &\times e^{j(\beta x + \omega t)} \end{aligned} \quad (63)$$

where  $E_0$  is the electric-field amplitude,  $Z_d = \sqrt{\mu_0/\epsilon}$  is the dielectric impedance,  $\lambda_d = \lambda_{\text{freespace}}/\sqrt{\epsilon_r}$  is the dielectric wavelength, and  $\beta$  is the mode propagation constant. For ruby  $\epsilon_r \approx 10$ , so  $Z_d \approx 120\Omega$  and  $\lambda_d \approx 0.0566$  in. The RWM guide dimensions are  $a = 0.10$  in. and  $b = 0.05$  in. [28]. From Eq. (63),  $\bar{H}_1$  is circularly polarized for

$$\tan \frac{\pi y}{a} = \frac{\frac{\lambda_d}{2a}}{\sqrt{1 - \left( \frac{\lambda_d}{2a} \right)^2}} \quad (64)$$

which has the solutions  $y \approx 0.1a$  and  $y \approx 0.9a$ . Forming the product  $\bar{H}_1^* \cdot \bar{\sigma}_{13}$ , it is found that the pump RF field is most strongly coupled to the spins at  $y \approx 0.1a$ .

Neglecting losses and ruby absorption, the time-average power flow down the waveguide is [29]

$$P_{\text{avg}} = \frac{Z_{\text{TE}}}{2} \int_{\text{cross section}} |H_t|^2 dA \quad (65)$$

where  $H_t$  is the transverse component of  $H_1$  and

$$Z_{\text{TE}} = \frac{Z_d}{\sqrt{1 - \left( \frac{\lambda_d}{2a} \right)^2}} \quad (66)$$

Evaluating Eq. (65) and solving for  $E_0$ , one obtains

$$E_0 = \left( \frac{4 \cdot P_{\text{avg}} \cdot Z_{\text{TE}}}{a \cdot b} \right)^{1/2} \quad (67)$$

Assuming  $P_{\text{avg}} = 50$  mW per channel [28], an electric-field amplitude  $E_0 = 2.78 \times 10^3$  V/m is obtained in the guide.

Substituting Eq. (62) into Eq. (59) and noting that  $H_1$  has no  $z$  component,

$$W_{13} = \frac{1}{16} \alpha_{13}^2 \gamma^2 g(f) \left| \overline{H}_1^* \cdot (\hat{x} - j\hat{y}) \right|^2 \quad (68)$$

Since  $\overline{H}_1$  varies as a function of  $y$ , an average value of the modulus-squared quantity in Eq. (68) is required. Evaluating the average, one obtains

$$\frac{1}{a} \int_0^a \left| \overline{H}_1^* \cdot (\hat{x} - j\hat{y}) \right|^2 dy = \frac{1}{2} \left( \frac{E_0}{Z_d} \right)^2 \quad (69)$$

Equation (68) averaged over the waveguide cross section becomes

$$\left\langle W_{13} \right\rangle_{\text{c.s.}} = \frac{1}{32} \alpha_{13}^2 \gamma^2 g(f) \left( \frac{E_0}{Z_d} \right)^2 \quad (70)$$

To evaluate Eq. (70), one needs to know the line shape function  $g(f)$ . For an unbroadened Lorentzian line of width  $\Delta f_L$ , the line shape at the pump frequency is [2]

$$g(f_p) \approx \frac{2}{\pi \Delta f_L} \quad (71)$$

where  $\Delta f_L \approx 60$  MHz for ruby.

For a line broadened by a stagger-tuned dc field and assuming that the pump sweep rate  $\gg$  the pump level relaxation time, the line shape is approximately [2]

$$g(f) = \begin{cases} \frac{1}{\frac{\partial f_p}{\partial f_s} \Delta f_s} & f_p - \frac{\Delta f_s}{2} < f < f_p + \frac{\Delta f_s}{2} \\ 0 & \text{elsewhere} \end{cases} \quad (72)$$

where  $\Delta f_p$  and  $\Delta f_s$  are the pump and signal linewidths, and  $\partial f_p / \partial f_s$  is evaluated by differencing or by application of the Hellman-Feynman theorem [7]. For the 32-GHz

pumping scheme used in the RWM,  $\partial f_p / \partial f_s \approx 2$ . A typical broadened linewidth for the RWM is  $\Delta f_s = 400$  MHz.

Evaluating Eq. (70) for the above values, and noting that  $\gamma = 2.2 \times 10^5$  m/Coulomb,

$$W_{13} \approx \begin{cases} 780 \text{ sec}^{-1} & \text{unbroadened} \\ 90 \text{ sec}^{-1} & \text{broadened} \end{cases} \quad (73)$$

From Fig. 4, the corresponding inversion ratios at  $T = 4.5$  K are

$$I \approx \begin{cases} 1.3 & \text{unbroadened} \\ 0.8 & \text{broadened} \end{cases} \quad (74)$$

These values compare favorably with the corresponding experimental values. Moore and Neff [27] found the maximum inversion ratio to be  $I = 1.1$  to  $1.2$ . (The saturation value of  $I$  in Fig. 4 is  $I \approx 1.5$ ). Shell and Neff [28] determined the inversion ratio of the RWM broadened to 400 MHz to be  $I \approx 0.7$  to  $0.8$ .

## D. Noise Temperature

Amplifier noise temperature  $T'_a$  is plotted versus physical temperature  $T$  with the resistive loss quantity  $\beta$  as a parameter for the two 32-GHz pumping schemes shown in Figs. 7 and 8. Following Stelzreid [12], the amplifier noise power is

$$P_{n,\text{amp}} = G k T'_a B \quad (75)$$

where  $G$  is the amplifier gain and  $B$  is the bandwidth. To make this definition valid for  $hf/kT'_a \gtrsim 1$ , define

$$T'_a = \frac{\frac{hf}{k}}{e^{\frac{hf}{kT'_a}} - 1} \quad (76)$$

where  $T_a$  is defined by the black-body noise power

$$P_{n,\text{amp}} = G \frac{hfB}{e^{\frac{hf}{kT_a}} - 1} \quad (77)$$

Noise in masers is discussed thoroughly in [2]. The two primary noise mechanisms are spontaneous emission of photons by the ruby spins and conductor losses in the microwave circuitry. The noise power of a TWM is [2]

$$P_{n,amp} = (G - 1) \left( \frac{\alpha_m}{\alpha_0 - \alpha_m} P_n(-T_m) + \frac{\alpha_0}{\alpha_m - \alpha_0} P_n(T) \right) \quad (78)$$

where

$$P_n(T_0) = \frac{hfB}{e^{\frac{hf}{kT_0}} - 1} \quad (79)$$

and

$$G = e^{2(\alpha_m - \alpha_0)L} \quad (80)$$

where  $\alpha_m$  and  $\alpha_0$  are the ruby gain and forward loss coefficients, respectively, of the TWM, and  $L$  is the structure length.  $T_m = |T_s|$  is the magnitude of the spin temperature defined by the ratio of the inverted spin populations:

$$\frac{n_i}{n_j} = e^{-\frac{hf_{ij}}{kT_s}} \quad i > j \quad (81)$$

For  $G \gg 1$ , Eqs. (75), (78), and (79) yield a noise temperature

$$T'_a = \frac{hf}{k(1 - \beta)} \left[ \frac{1}{1 - e^{-\frac{hf}{kT_m}}} + \frac{\beta}{e^{\frac{hf}{kT}} - 1} \right] \quad (82)$$

where  $\beta = \alpha_0/\alpha_m$ . Figures 14 and 15 show  $T'_a$  versus  $T$  with  $\beta$  as a parameter and assuming pump level saturation. The figures display a slightly lower  $T'_a$  at a given physical temperature and  $\beta$  for the 90-deg orientation. The difference arises from the lower  $T_m$  for the 90-deg pumping scheme. Note that at  $T = 1.5$  K the effect of  $\beta$  is diminished. Estimates of  $\beta$  for a 32-GHz TWM are not

available at this time. For the RWM, Shell and Neff [28] estimate  $\beta \sim 0.1$ .

## VI. Conclusions

This article shows that by accounting for the rate of spin relaxation from higher energy levels, predictions of many basic maser parameters such as gain, pump power, and noise temperature can be made. For ruby, the dominant source of spin relaxation is interaction of the  $\text{Cr}^{3+}$  ions with thermal vibrations of the  $\text{Al}_2\text{O}_3$  lattice. Based on measurements found in the literature which characterize the strength of the coupling between the spins and the lattice, quantitative estimates of the spin relaxation rates have been made.

Relaxation time measurements are presently underway at 9 GHz and 35 GHz as a check of the theory and to determine the best orientations and Cr concentration of the ruby crystal for a 32-GHz maser. Previous relaxation time measurements have shown varying degrees of agreement with the theory, perhaps due to the presence of impurities or due to clustering of Cr ions. To account for impurities or clustering, the theory would have to be modified to account for spin-spin interactions.

Comparison of predicted inversion ratios with measured values also shows varying degrees of agreement, but the agreement is generally good. More detailed experimental studies of inversion ratio and gain for various levels of pump power and as a function of physical temperature for the 54.7-deg and 90-deg ruby orientations are necessary at 32 GHz.

The theory predicts that at 32 GHz and assuming pump saturation, the gain for the 90-deg orientation (1-2 signal transition) should be nearly 50 percent higher in dB than for the 54.7-deg orientation (2-3 signal transition). The theory also predicts a slightly lower noise temperature for the 90-deg orientation. Although specific calculations have not been performed, it is expected that the 90-deg orientation will require  $\approx 50$  percent greater pump power than the 54.7-deg orientation.

The usefulness of the present theory remains to be seen. It may be found that spin-spin interactions cannot be neglected even for the relatively dilute ruby used in maser amplifiers. The theory is, however, a significant step towards a more complete understanding of ruby masers.



## References

- [1] S. Petty, "Introduction to Microwave Devices, Part VIII," in *Low Temperature Electronics*, edited by R. Kirschman, New York: IEEE Press, pp. 358–363, 1986.
- [2] A. Siegman, *Microwave Solid-State Masers*, New York: McGraw-Hill, 1964.
- [3] R. B. Hemphill, P. L. Donoho, and E. D. McDonald, "Spin-Lattice Interaction in Ruby Measured by Electron Spin Resonance in Uniaxially Stressed Crystals," *Physical Review*, vol. 146, pp. 329–335, 1966.
- [4] P. L. Donoho, "Spin Lattice Relaxation in Ruby," *Physical Review*, vol. 133, pp. A1080–A1084, 1964.
- [5] K. J. Standley and R. A. Vaughan, *Electron Resonance Phenomena in Solids*, London: Adam Hilger LTD, 1969.
- [6] B. Bleaney and W. Stevens, "Paramagnetic Resonance," *Reports Prog. Phys.*, vol. 16, pp. 108–159, 1953.
- [7] P. W. Atkins, *Molecular Quantum Mechanics*, 2nd ed., Oxford: Oxford University Press, 1983.
- [8] R. Berwin, "Paramagnetic Energy Levels of the Ground State of  $\text{Cr}^{3+}$  in  $\text{Al}_2\text{O}_3$  (Ruby)," *Technical Memorandum 33-440*, Jet Propulsion Laboratory, Pasadena, California, January 1970.
- [9] J. R. Lyons, "Theoretical Comparison of Maser Materials for a 32-GHz Maser Amplifier," *TDA Progress Report, 42-95*, vol. July–September 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 58–70, November 15, 1988.
- [10] E. Schulz-Du Bois, "Paramagnetic Spectra of Substituted Sapphires—Part 1: Ruby," *Bell Syst. Tech. J.*, vol. 38, pp. 271–290, 1959.
- [11] L. Schiff, *Quantum Mechanics*, 3rd ed., New York: McGraw-Hill, 1968.
- [12] C. T. Stelzreid, *The Deep Space Network—Noise Temperature Concepts, Measurements, and Performance*, JPL Publication 82-33, Jet Propulsion Laboratory, Pasadena, California, 1982.
- [13] A. Abragam and B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions*, New York: Dover, 1986.
- [14] R. D. Mattuck and M. W. P. Strandberg, "Spin-Phonon Interaction in Paramagnetic Crystals," *Phys. Rev.*, vol. 119, pp. 1204–1217, 1960.
- [15] E. B. Tucker, "Interaction of Phonons with Iron-Group Ions," *Proc. of IEEE*, vol. 53, pp. 1547–1573, October 1965.
- [16] J. H. Van Vleck, "Paramagnetic Relaxation Times for Titanium and Chrome Alum," *Physical Review*, vol. 57, pp. 426–447, 1940.
- [17] F. Fumi, "The Direct-Inspection Method in Systems with a Principal Axis of Symmetry," *Acta Cryst.*, vol. 5, pp. 691–694, 1952.
- [18] C. Kittel, *Introduction to Solid State Physics*, 6th ed., New York: John Wiley & Sons, 1986.

- [19] I. Manenkov and R. Orbach, *Spin-Lattice Relaxation in Ionic Solids*, New York: Harper and Row, 1966.
- [20] *CRC Handbook of Chemistry and Physics*, edited by R. C. Weast, Cleveland: The Chemical Rubber Co., 1970.
- [21] H. B. Huntington, "Elastic Constants of Crystals," in *Solid State Physics*, vol. 7, Seitz and Turnbull, eds., New York: Academic Press, 1958.
- [22] J. B. Wachtman, Jr., W. E. Tefft, D. G. Lam, Jr., and R. P. Stinchfield, "Elastic Constants of Synthetic Single Crystal Corundum at Room Temperature," *J. of Res. of N.B.S.*, vol. 64A, pp. 213-228, 1960.
- [23] J. N. Franklin, *Matrix Theory*, Englewood Cliffs, New Jersey: Prentice-Hall, 1968.
- [24] K. J. Standley and R. A. Vaughan, "Effect of Crystal-Growth Method on Electron Spin Relaxation in Ruby," *Phys. Rev.*, vol. 139, pp. A1275-A1280, 1965.
- [25] R. Clauss, "RF Techniques Research: System Studies for Frequencies Above S-Band for Space Communications," *JPL Space Programs Summary 37-61*, vol. III, Jet Propulsion Laboratory, Pasadena, California, pp. 90-93, February 1970.
- [26] R. Berwin, R. Clauss, and E. Wiebe, "Low-Noise Receivers: Microwave Maser Development," *JPL Space Programs Summary 37-56*, vol. II, Jet Propulsion Laboratory, Pasadena, California, pp. 101-105, March 1969.
- [27] C. Moore and D. Neff, "Experimental Evaluation of a Ruby Maser at 43 GHz," *IEEE Trans. MTT*, vol. 30, pp. 2013-2015, 1982.
- [28] J. Shell and D. Neff, "A 32-GHz Reflected-Wave Maser Amplifier with Wide Instantaneous Bandwidth," *TDA Progress Report 42-94*, vol. April-June 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 145-162, August 15, 1988.
- [29] S. Ramo, J. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*, New York: John Wiley & Sons, 1965.
- [30] J. Owen and E. A. Harris, "Pair Spectra and Exchange Interaction," Chap. 6, in *Electron Paramagnetic Resonance*, ed. S. Geschwind, Plenum Press, 1972.

**Table 1. Comparison of Donoho's [4] calculated values of  $A_i$  and  $T_i$  with those of the author for  $\theta = 54.7$  deg,  $f = 9.3$  GHz, and  $T = 4.2$  K**

$\theta = 54.7$ deg, transition = 2-3		
	Donoho's Calculation [4]	Author's Calculation
$T_1$	0.38	0.20
$T_2$	0.22	0.11
$T_3$	0.13	0.07
$A_1$	0.99	1.00
$A_2$	0	0
$A_3$	0	0
$\theta = 54.7$ deg, transition = 3-4		
	Donoho's Calculation [4]	Author's Calculation
$T_1$	0.97	0.51
$T_2$	0.60	0.30
$T_3$	0.22	0.11
$A_1$	0.50	0.26
$A_2$	0.41	0.49
$A_3$	0.09	0.25

**Table 2. Comparison of measured inversion ratio with theoretical predictions for 0.05 percent Cr ruby at 4.5 K**

$\theta$ , deg	$H_0$ , kG	Transition(s) and Frequency, GHz		$I^{opt}$	$I^{meas}$	$I^{theory}$
		Signal	Pump			
54.7	5.72	2-3	1-3, 2-4	3.6	2.8 <sup>a</sup>	2.2
		14.4	32.9			
54.7	5.56	2-3	1-3, 2-4	3.6	3.0 <sup>a,b</sup>	2.3 <sup>b</sup>
		13.9	32.0			
54.7	6.78	2-3	1-3, 2-4	3.4	3.3	2.0
		17.2	28.0			
54.7	3.7	2-3	1-3, 2-4	4.2	2.5	3.0
		8.5	21.9			
54.7	6.6	2-3	1-3, 2-4	3.4	2.5	2.0
		16.8	37.3			
90	5.0	1-2	1-4	4.1	2.5	3.6
		8.5	43.5			
90	5.0	1-2	1-3, 3-4	4.1	3.2	3.6
		8.5	24.2, 19.3			
90	6.8	1-2	1-3, 3-4	3.3	3.0	2.7
		13.6	33.9, 24.6			
90	7.1	1-2	1-3, 3-4	3.2	2.9 <sup>b</sup>	2.4 <sup>b</sup>
		14.2	35.2, 25.2			
90	8.0	1-2	1-4	3.1	-0.1 <sup>c</sup>	2.5
		16.8	68.3			
90	8.0	1-2	1-3, 3-4	3.1	2.9	2.4
		16.8	40.3, 28.0			

<sup>a</sup> 0.075 percent Cr

<sup>b</sup> 1.8 K

<sup>c</sup> Pump not saturated

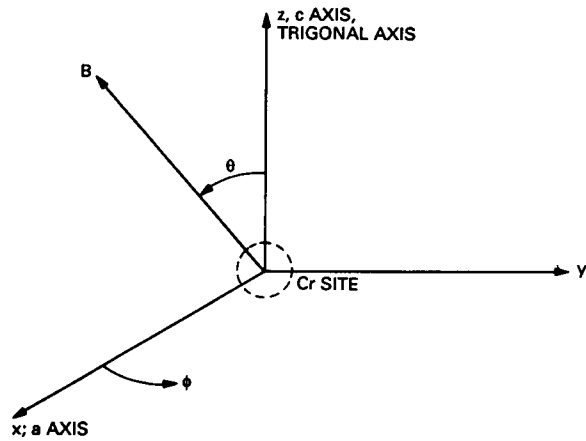


Fig. 1. Geometry of  $\text{Cr}^{3+}$  site.

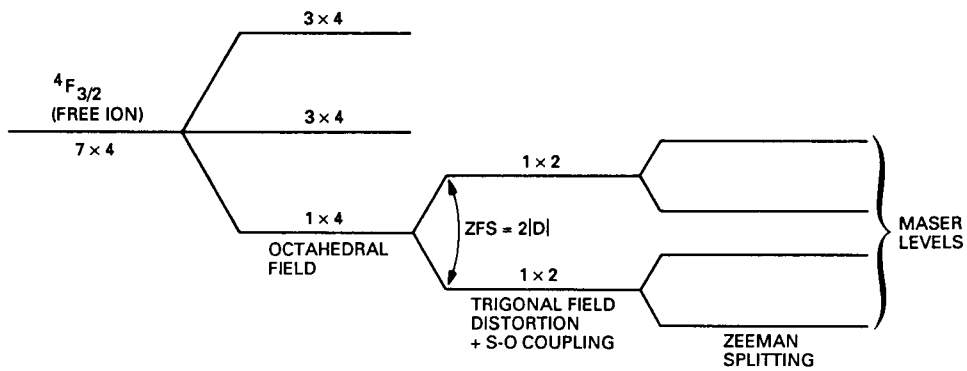


Fig. 2. Simplified energy splitting diagram for Cr in  $\text{Al}_2\text{O}_3$  lattice. Sequence of quantum mechanical perturbations applied is shown.  $m \times n = m$ -fold orbital degeneracy and  $n$ -fold spin degeneracy.

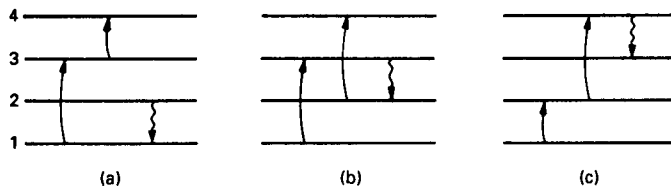


Fig. 3. Double-pumping schemes being considered for ruby for a 32-GHz maser. The downward arrow is the 32-GHz (signal) transition.

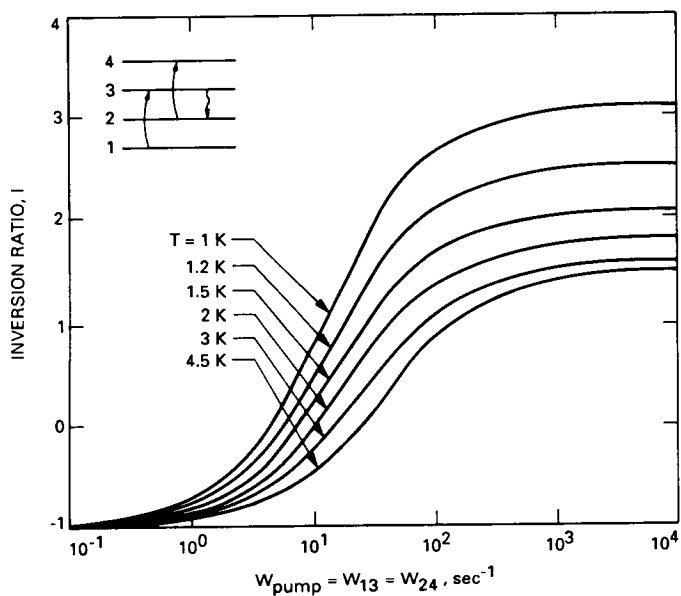


Fig. 4.  $I$  versus  $W_{\text{pump}}$  for  $\theta = 54.7$  deg,  $H = 11.8$  kG, signal = 2-3 transition = 32 GHz.

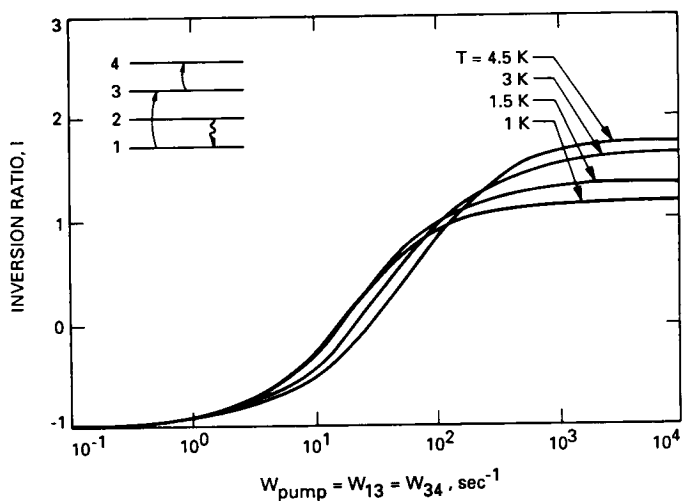


Fig. 5.  $I$  versus  $W_{\text{pump}}$  for  $\theta = 90$  deg,  $H = 13.5$  kG, signal = 1-2 transition = 32 GHz.

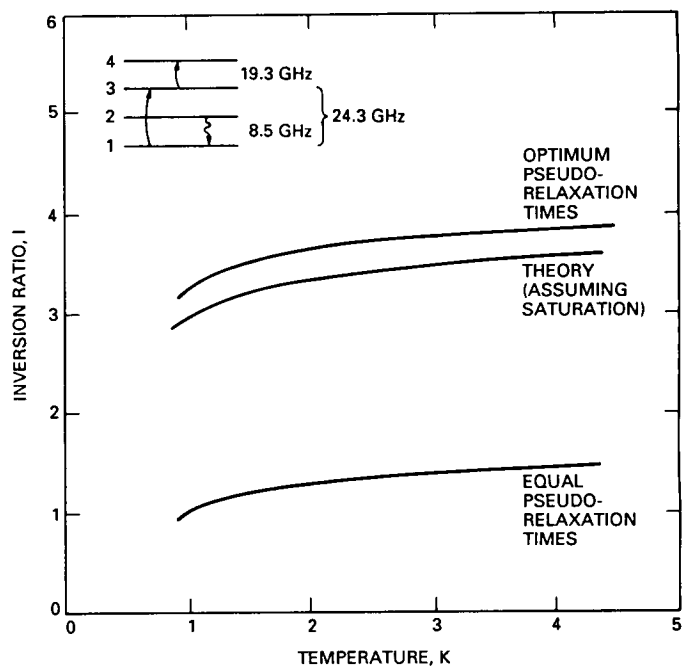


Fig. 6.  $I$  versus  $T$  assuming saturation for  $\theta = 90$  deg,  $H = 5$  kG, signal = 1-2 transition = 8.5 GHz.

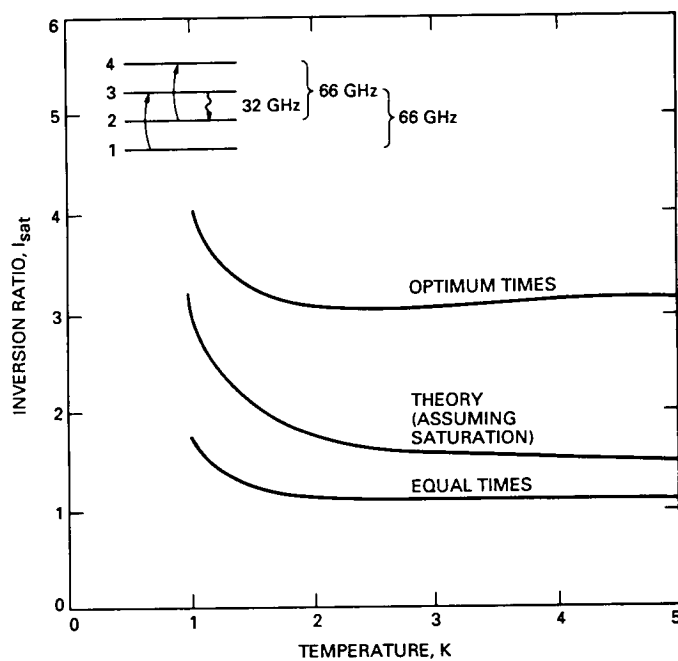


Fig. 7.  $I$  versus  $T$  assuming saturation for  $\theta = 54.7$  deg,  $H = 11.8$  kG, signal = 2-3 transition = 32 GHz.

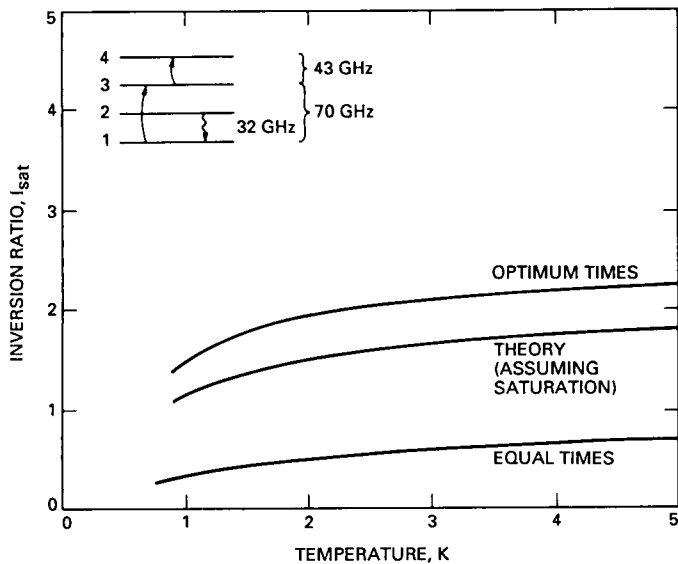


Fig. 8.  $I$  versus  $T$  assuming saturation for  $\theta = 90$  deg,  $H = 13.5$  kG, signal = 1-2 transition = 32 GHz.

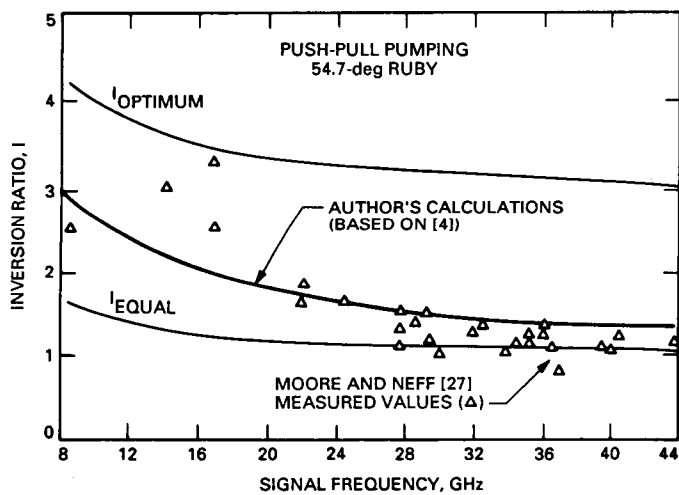


Fig. 9. Measured and theoretical values of  $I$  versus signal frequency for ruby,  $\theta = 54.7$  deg,  $T = 4.5$  K.

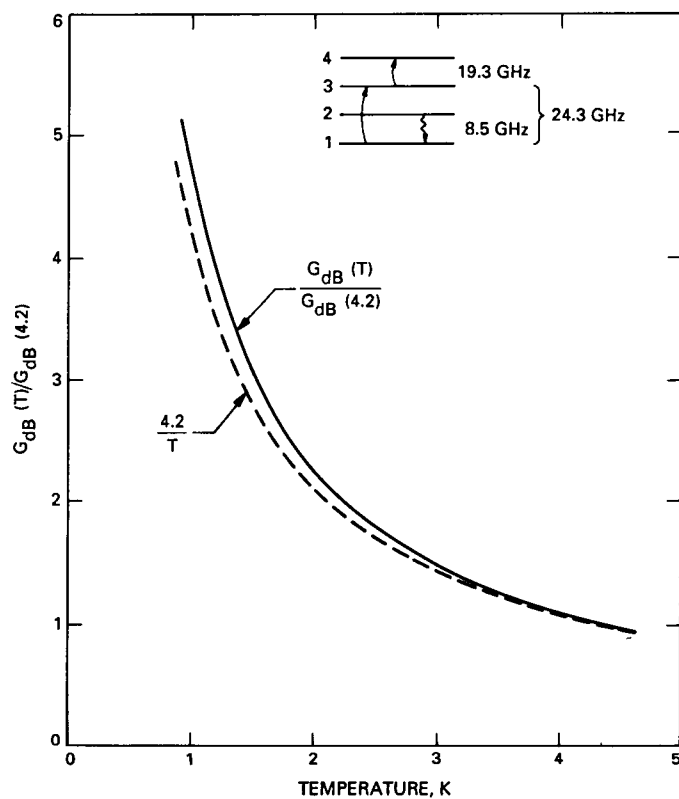


Fig. 10. The normalized gain versus  $T$  (assuming saturation) for  $\theta = 90$  deg,  $H = 5$  kG, and signal = 1-2 transition = 8.5 GHz.

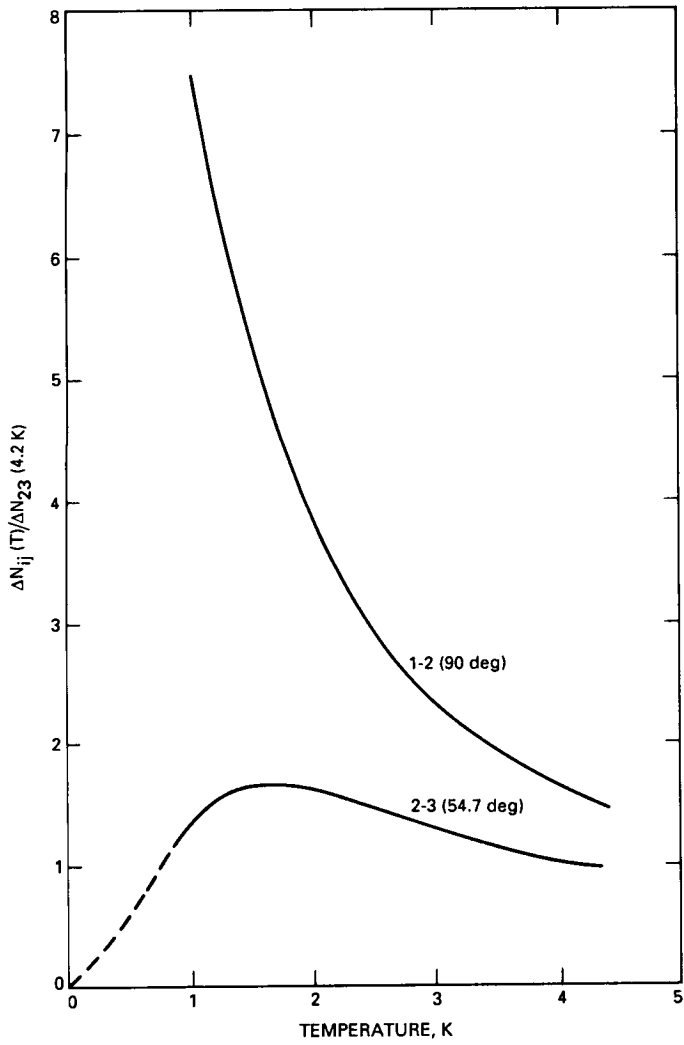


Fig. 11. Thermal equilibrium population difference for two 32-GHz pumping scenarios (normalized to population difference of the 54.7-deg scenario at 4.2 K).

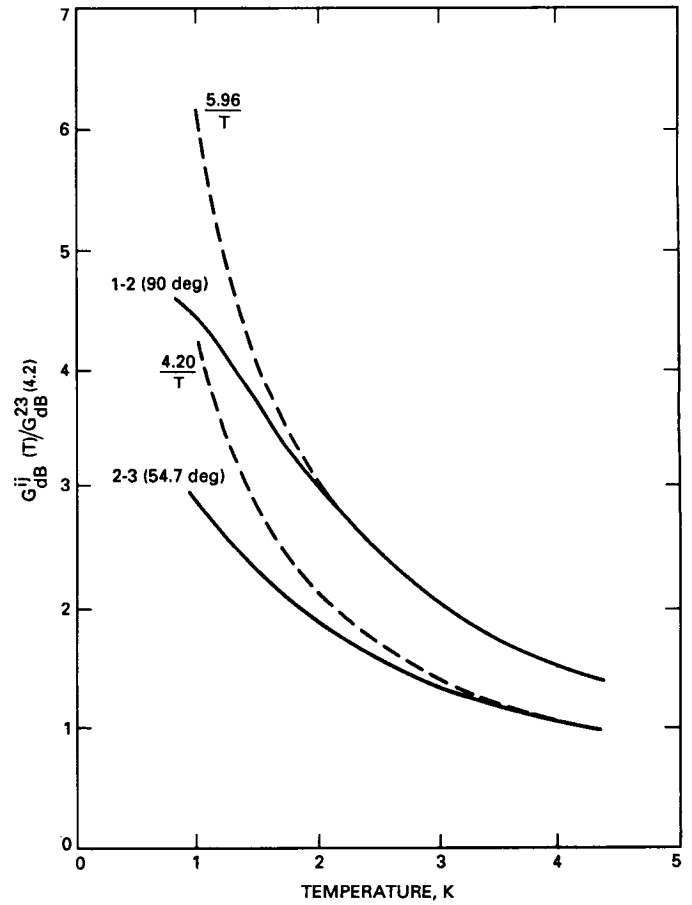
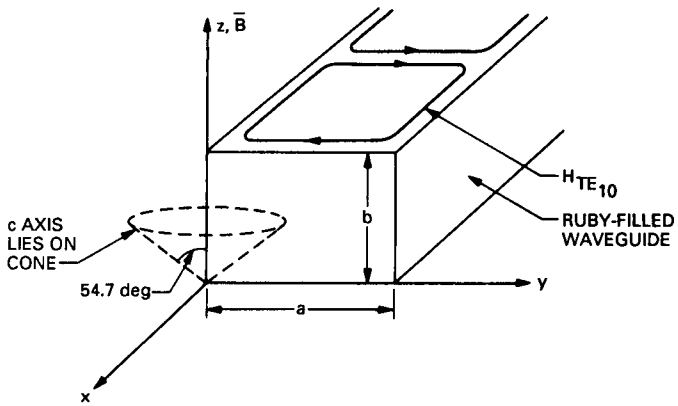
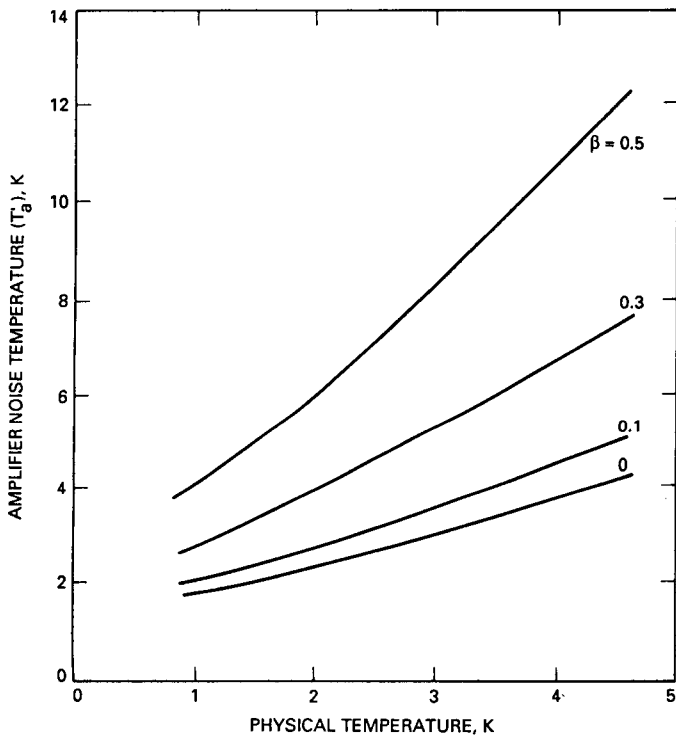


Fig. 12.  $G_{dB}$  versus  $T$  for 54.7-deg and 90-deg pumping schemes (assuming saturation).

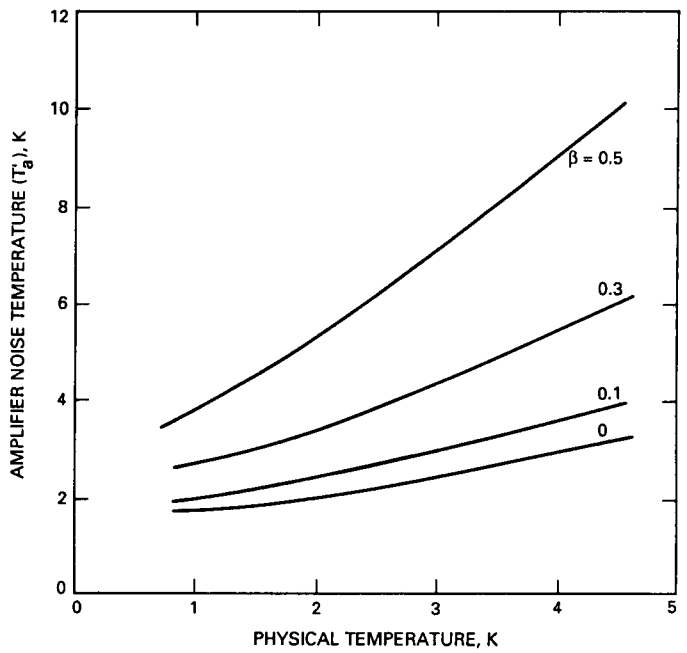




**Fig. 13. Geometry of a single channel of ruby-filled waveguide for 32-GHz RWM.**



**Fig. 14.  $T'_a$  versus temperature for 32-GHz ruby maser,  $\theta = 54.735$  deg, signal = 2-3 transition = 32 GHz.**



**Fig. 15.  $T'_a$  versus temperature for 32-GHz ruby maser,  $\theta = 90$  deg, signal = 1-2 transition = 32 GHz.**

55-32

229615

49.

TDA Progress Report 42-98

547639 P10

N90-12791

August 15, 1989

# Performance Effects of Tie-Truss Modifications for a 70-Meter Centerline Beam Waveguide Antenna

J. J. Cucchisi

Ground Antennas and Facilities Engineering

*The elevation-axis tie truss of the 70-m antennas would have to be modified to accommodate a centerline beam waveguide. To accomplish this, the center section of the tie truss has to be altered, causing a change in the tie-truss compliance and affecting structural performance. Even with the center section completely removed, the worst-case rms pathlength error due to gravity load is increased from 0.025 to only 0.030 inches. Using a simple postprocessor technique, the effects of modifying the compliance can be predicted without resorting to multiple and costly reanalyses of large finite-element models on a mainframe computer.*

## I. Introduction

With the imminent implementation of beam waveguide technology in the Deep Space Network, a study was initiated to predict the performance degradation due to retrofitting existing antennas in the 70-m subnet with a centerline or on-axis beam waveguide. Whereas the 70-m antenna structure was designed for a dish-mounted feedcone (Fig. 1) and no allowance made for an on-axis beam waveguide, some structural components of the main reflector and elevation wheel would obstruct the path of microwave energy of a beam waveguide system. The first significant obstruction is a 56-foot tie truss (Figs. 2 and 3). It is a space truss of triangular cross section that is integral with the reflector backup and elevation wheel substructures and that runs parallel to the elevation axis and connects the two elevation bearing castings. To allow a

beam waveguide shroud to pass through this region, the center section would have to be modified or removed.

The purpose of this study was to determine the effect of tie-truss stiffness on the main reflector surface distortion under gravity load. Specifically, the analyses quantified the sensitivity of the gravity rms half-pathlength error to the stiffness of the tie-truss center section. Supplemental analyses also examined changes to the lower frequency modes when the center section is removed completely.

The sensitivity analyses described herein were performed using the methods of structural modification reanalysis [1, 2] and correlation analysis [2]. The specific procedures used are explained briefly; more rigorous descriptions of the theory and further applications can be found in the references.

## II. Methodology

The 70-m tipping structure comprises the main reflector, elevation wheel, subreflector, and quadripod. A planar-symmetric finite-element half model of this structure includes 6,776 finite elements (primarily axial force bars), 1,994 nodes, and 5,982 translational degrees of freedom. Finite-element analysis and microwave-pathlength analysis is accomplished via the JPL-IDEAS finite-element analysis and design optimization program [3] on the UNIVAC. Sensitivity coefficients, derived from the large mainframe analysis model, are incorporated into PC-based postprocessing software that uses a combination of structural modification reanalysis and correlation analysis to assess the effect on performance when the cross-sectional areas of selected structural members are reduced or when the members are removed completely. Some postprocessor results were verified via mainframe finite-element processing.

The model was analyzed first for gravity load and the forces in the members in the central region of the tie truss were examined (Fig. 3). Those members that carried at least one percent of the gravity load and would obstruct a centerline beam waveguide path were selected for modification. For this study, nine bars were selected.

### A. Structural Modification Reanalysis

To model the effects of modifying these selected members, sets of self-equilibrating unit virtual loading pairs [1], called indicator loads, are applied to the finite-element model at the terminal nodes of each of the members to be altered, one set per member. Static analysis is performed for each set separately and also for the  $y$  and  $z$  components of the gravity load. The JPL-IDEAS program provides: (1) the forces due to  $y$  and  $z$  gravity, and (2) the forces due to the indicator loadings, for each of the members to be altered. The following matrices then can be constructed:

$\mathbf{P}_I$   $2 \times N$  matrix of original forces for  $N$  altered members,  $z$  and  $y$  gravity rows

$\mathbf{P}_S$   $N \times N$  matrix of forces in  $N$  altered members for indicator loadings

$\mathbf{F}$   $N \times N$  diagonal matrix of original flexibilities of  $N$  altered members

$\mathbf{SF}$   $N \times N$  diagonal matrix of modified flexibilities of  $N$  altered members

Using the calculated member distortions

$$\mathbf{E}_I = \mathbf{F} \mathbf{P}_I^T \quad (1)$$

$$\mathbf{E}_S = \mathbf{F} \mathbf{P}_S \quad (2)$$

$$\mathbf{E}_O = \mathbf{SF} = \text{factor} * \mathbf{F} \quad (3)$$

and enforcing compatibility, the  $N \times 2$  modifier matrix  $\mathbf{R}$  is found by solving

$$(\mathbf{E}_O - \mathbf{E}_S)\mathbf{R} = \mathbf{E}_I \quad (4)$$

A more detailed discussion of reanalysis is presented in [1].

### B. Correlation Analysis

After determining the  $\mathbf{R}$  modifier matrix, the gravity loading pathlength error analysis for the modified structure can proceed. The pathlength error vectors for  $z$ - and  $y$ -gravity loads after modification can be assembled from the independent pathlength error vectors for  $z$  and  $y$  gravity before modification and the indicator loads as follows:

$$\begin{bmatrix} \vec{\rho}_z & \vec{\rho}_y \end{bmatrix} = \begin{bmatrix} \vec{\rho}_{z_0} & \vec{\rho}_{y_0} \end{bmatrix} + \begin{bmatrix} \vec{\rho}_1 & \vec{\rho}_2 & \dots & \vec{\rho}_N \end{bmatrix} \mathbf{R} \quad (5)$$

where  $\vec{\rho}_k$  = best-fit pathlength error vector for load  $k$  and subscripts  $z_0$ ,  $y_0$  refer to  $z$  and  $y$  gravity response before modification, and  $1, 2, 3, \dots, N$  refer to indicator loads.

At any elevation angle  $\alpha$ , the gravity pathlength error vector is

$$\vec{\rho}_G = \begin{bmatrix} \vec{\rho}_z & \vec{\rho}_y \end{bmatrix} \mathbf{c} \quad (6)$$

where

$$\mathbf{c} = \begin{bmatrix} \zeta \\ \eta \end{bmatrix}$$

and  $\zeta = \sin \gamma - \sin \alpha$ ,  $\eta = \cos \gamma - \cos \alpha$ , and  $\gamma$  = rigging angle (usually 45 deg),  $\alpha$  = any elevation angle,  $0 \leq \alpha \leq 90$  deg.

So, for the modified structure

$$\vec{\rho}_G = \begin{bmatrix} \vec{\rho}_{z_0} & \vec{\rho}_{y_0} \end{bmatrix} \mathbf{c} + \begin{bmatrix} \vec{\rho}_1 & \vec{\rho}_2 & \dots & \vec{\rho}_N \end{bmatrix} \mathbf{R} \mathbf{c} \quad (7)$$

Forming the sums of squares:

$$\begin{aligned} SS = \vec{\rho}_G^T \vec{\rho}_G &= \mathbf{c}^T \mathbf{P}_{G_0}^T \mathbf{P}_{G_0} \mathbf{c} \\ &+ \mathbf{c}^T \mathbf{P}_{G_0}^T \mathbf{P}_N \mathbf{R} \mathbf{c} \\ &+ \mathbf{c}^T \left( \mathbf{P}_{G_0}^T \mathbf{P}_N \mathbf{R} \right)^T \mathbf{c} \\ &+ \mathbf{c}^T \mathbf{R}^T \mathbf{P}_N^T \mathbf{P}_N \mathbf{R} \mathbf{c} \end{aligned} \quad (8)$$

where  $\mathbf{P}_{G_0} = [\vec{\rho}_{z_0} \ \vec{\rho}_{y_0}]$  and  $\mathbf{P}_N = [\vec{\rho}_1 \ \vec{\rho}_2 \ \dots \ \vec{\rho}_N]$ .

Using the notation  $C_{ij}$  = correlation coefficient for pathlength error vectors  $\vec{\rho}_i$  and  $\vec{\rho}_j$ , and  $RMS_k$  = the best-fit rms pathlength error for vector  $\vec{\rho}_k$ , and noting that  $C_{ij} = \vec{\rho}_i^T \vec{\rho}_j / (RMS_i * RMS_j)$ , the first term in Eq. (8) can be shown to be:

$$A = \mathbf{c}^T \mathbf{R} \mathbf{M}_{zy} \mathbf{C}_{zy} \mathbf{R} \mathbf{M}_{zy} \mathbf{c}$$

where

$$\mathbf{R} \mathbf{M}_{zy} = \begin{bmatrix} RMS_z & 0 \\ 0 & RMS_y \end{bmatrix}$$

and

$$\mathbf{C}_{zy} = \begin{bmatrix} 1 & C_{zy} \\ C_{zy} & 1 \end{bmatrix}$$

the second and third terms can be shown to be:

$$B_1 = \mathbf{c}^T \mathbf{R} \mathbf{M}_{zy} \mathbf{C}_{GN} \mathbf{R} \mathbf{M}_N \mathbf{R} \mathbf{c}$$

$$B_2 = \mathbf{c}^T \left( \mathbf{R} \mathbf{M}_{zy} \mathbf{C}_{GN} \mathbf{R} \mathbf{M}_N \mathbf{R} \right)^T \mathbf{c}$$

where

$$\mathbf{C}_{GN} = \begin{bmatrix} C_{z1} & C_{z2} & \dots & C_{zN} \\ C_{y1} & C_{y2} & \dots & C_{yN} \end{bmatrix}$$

and

$$\mathbf{R} \mathbf{M}_N = \begin{bmatrix} RMS_1 & \dots & \dots & 0 \\ \vdots & RMS_2 & & \\ \vdots & & \ddots & \\ 0 & \dots & \dots & RMS_N \end{bmatrix}$$

for  $N$  altered members (indicator loads), and the fourth term can be shown to be:

$$C = \mathbf{c}^T \mathbf{R}^T \mathbf{R} \mathbf{M}_N \mathbf{C}_{IJ} \mathbf{R} \mathbf{M}_N \mathbf{R} \mathbf{c}$$

where

$$\mathbf{C}_{IJ} = \begin{bmatrix} 1 & C_{12} & C_{13} & \dots & C_{1N} \\ C_{12} & 1 & C_{23} & \dots & C_{2N} \\ C_{13} & C_{23} & 1 & \dots & C_{3N} \\ \vdots & & & \ddots & \\ C_{1N} & C_{2N} & C_{3N} & \dots & 1 \end{bmatrix}$$

The first term is the contribution due to  $y$ -gravity and  $z$ -gravity loads and is assumed invariant. The second and third terms are the contributions of each indicator load correlated with both  $y$ - and  $z$ -gravity. The fourth term is the contribution of each indicator load correlated with the other indicator loads. Note that all but the first term require the modifier matrix  $\mathbf{R}$  calculated earlier using structural modification reanalysis.

The JPL-IDEAS program provides the best-fit rms pathlength error ( $RMS_k$ ) for each of the loads and the correlation coefficient ( $C_{ij}$ ) for all pathlength-error vector pairs. By substituting in these values each term is readily calculated. Summing the four terms and taking the square root yields the gravity rms pathlength error for a particular modification of the  $N$  selected members.

$$rms = (SS)^{1/2} = (A + B_1 + B_2 + C)^{1/2} \quad (9)$$

For other changes to the  $N$  selected members only a new  $\mathbf{R}$  matrix need be calculated. Response over the antenna elevation range is computed by varying  $\alpha$  from 0 to 90 deg.

### III. Natural Frequency Analysis

Since the gravity rms pathlength error was not severely degraded by removing the center section of the tie truss, the normal modes were examined and compared before and after removal. Again, the JPL-IDEAS program was used to perform the natural frequency analyses. For a half-model structure, two stiffness matrix decompositions are required, one with symmetric and one with anti-symmetric boundary restraints, to recover all allowable modes for the full structure.

Instead of being rigidly fixed at the elevation bearings, the stand-alone tipping-structure model is connected to linear springs simulating the lateral flexibility of the alidade at the elevation bearing locations. This modeling provides a more realistic representation of the alidade and tipping structure interaction. Furthermore, since only half the structure is modeled, the alidade flexibility is represented by two springs, one for symmetric and one for anti-symmetric boundary conditions. Independent static analyses of the alidade were performed to determine the lateral flexibility of the alidade at the bearings. These compliances were then incorporated into the tipping-structure model for normal modes analysis.

### IV. Numerical Results

The graph in Fig. 4 shows the change in the gravity rms pathlength error over the antenna elevation range for a series of bar-area reductions. Each rms achieves a maximum value at the extreme elevation angle of 0 deg; the rms is zero, by definition, at the prescribed rigging angle of 45 deg. There is little discernible difference in antenna performance between the original model and one where the tie-truss stiffness was reduced by 80 percent. Even with the center section removed, the worst case rms is 0.030 in.

The extreme values plotted in Fig. 1 are listed in Table 1 under combined gravity rms at 0 and 90 deg, for an antenna rigged at 45 deg. Also included are the rms values for a 1-g  $y$ -gravity load, a 1-g  $z$ -gravity load, and the correlation coefficient for these two loads. The equation that accompanies the table shows a simple way to calculate the combined gravity response from the  $y$  and  $z$  components. These results reflect only the effect of uniformly downsizing

each of the nine selected bars in the tie-truss center section by a percentage of their original area. No allowance is made for the reduction in gravity load due to reducing member areas because the weight of these members represents only 0.5 percent of the total gravity (dead) load. Also, the solutions assume stiffness reduction only with no further optimization of the reflector backup or tie-truss structures to compensate for the rms increase.

When spot-checked with mainframe finite-element model analyses, the postprocessor made accurate predictions as the bar areas were reduced. As the areas approached zero, however, the predictions became unreliable; the  $(\mathbf{E}_O - \mathbf{E}_S)$  matrix became singular, correctly indicating unstable nodes in the finite-element model. To accurately analyze this case, the bars and any extraneous nodes had to be removed from the model, and a mainframe finite-element analysis was performed. Those results for the center section removed are tabulated as the 100-percent area-reduction case.

In Table 2 is a comparison of the lowest anti-symmetric modes, showing the effects of including accurate representations of the alidade compliance across the elevation bearings. The percentage of the total inertia about the  $y$  (roll) axis and  $z$  (yaw) axis contributed by each mode is listed in the table under effective modal inertia. The anti-symmetric modes above the first were changed moderately, as indicated by the frequency shifts and the redistribution of the effective modal inertias. However, the first anti-symmetric mode, which is essentially torsion of the quadripod at 1.28 Hz, was unchanged. This frequency is of particular importance because it is the lowest natural frequency of the 70-m antenna and, as such, is a critical performance constraint. To avoid excitation of the antenna modes, the lowest frequency must be outside the position loop bandwidth of the antenna drive servo system.

When expressed as a percentage of the total inertia about each axis, effective modal inertias provide general information about the mode shape. In the coordinate system used for the tipping structure model (see Fig. 1), a value for  $\theta_x$  indicates a pitch mode, for  $\theta_y$  a roll mode, and for  $\theta_z$  a yaw mode. For example, in Table 3 anti-symmetric mode 2 after removal is a mixture of roll and yaw, representing 12.1 percent of the inertia about the  $y$  axis and 5.6 percent of the inertia about the  $z$  axis. Mode 3 is also a mixed mode, although predominantly yaw and representing 23.9 percent of the inertia about the  $z$  axis. Highly localized vibrations or modes with a small fraction of the rotational inertia, such as torsion of the quadripod, will appear as small values in these tables; examination of the

eigenvector is required to determine the character or mode shape of these vibrations.

Tables 3 and 4 compare the natural frequency performance of the antenna before and after the complete removal of the tie-truss center section. The analyses incorporated appropriate alidade springs with different symmetric and anti-symmetric mode properties. Both the anti-symmetric (antenna roll and yaw) and the symmetric (antenna pitch) modes were unaffected by the removal.

## V. Summary

Analysis indicates that the center section of the elevation tie truss can be removed with a predicted degra-

ation in gravity pathlength error of 0.005 in. rms. In addition, the removal does not compromise the natural frequency performance. In the model, an alternate load path through an adjacent structure must have been in effect to compensate for the tie-truss removal. Before any modification of the existing tie truss is recommended, this load path must be clearly defined and the integrity of the structural elements composing this path must be verified.

The postprocessor techniques used in the study simplified the analysis and provided accurate and reliable results when compared to the mainframe finite-element analyses. They are also inherently self-checking. When the predictions became unreliable, the postprocessor correctly indicated that an instability existed in the finite-element model.

## References

- [1] R. Levy, "A NASTRAN Postprocessor for Structural Modification Reanalysis," *NASA TM X-2378, Vol. II*, pp. 737-747, September 1971.
- [2] R. Levy, "Reanalysis, Compatibility, and Correlation in Analysis of Modified Antenna Structures," *TDA Progress Report 42-97*, vol. January-March 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 367-381, May 15, 1989.
- [3] R. Levy and D. Strain, "JPL-IDEAS Iterative Design of Antenna Structures," Computer Program C-7282, submitted to *Cosmic*, November 1988.

**Table 1. Stiffness reduction effects on RMS pathlength error**

Area or stiffness reduction, percent	1-g y-gravity RMS, in. (cm)	1-g z-gravity RMS, in. (cm)	Correlation coefficient, $C_{zy}$	Combined gravity <sup>a</sup> RMS, in. (cm)	
				EL 0 deg	EL 90 deg
0	0.0274 (0.0696)	0.0354 (0.0899)	0.1095	0.0254 (0.0645)	0.0210 (0.0533)
20	0.0275 (0.0699)	0.0356 (0.0904)	0.1146	0.0255 (0.0648)	0.0210 (0.0533)
40	0.0276 (0.0701)	0.0359 (0.0912)	0.1214	0.0257 (0.0653)	0.0210 (0.0533)
60	0.0281 (0.0714)	0.0365 (0.0927)	0.1303	0.0260 (0.0660)	0.0213 (0.0541)
80	0.0292 (0.0742)	0.0375 (0.0953)	0.1404	0.0267 (0.0678)	0.0220 (0.0559)
100	0.0315 (0.0800)	0.0412 (0.1046)	0.1291	0.0294 (0.0747)	0.0239 (0.0607)

$${}^a RMS_{\alpha} = \left( \eta^2 RMS_y^2 + \zeta^2 RMS_z^2 + 2\eta\zeta RMS_y RMS_z C_{zy} \right)^{1/2}$$

for  $\alpha = 0$  deg,  $\gamma = 45$  deg

$$\eta = -0.2929$$

$$\zeta = 0.7071$$

and for  $\alpha = 90$  deg,  $\gamma = 45$  deg

$$\eta = 0.7071$$

$$\zeta = -0.2929$$

**Table 2. Comparison of anti-symmetric modes with and without simulated alidade compliance**

Mode	With alidade compliance			Without alidade compliance		
	Frequency, Hz	Effective modal inertia, percent of total		Frequency, Hz	Effective modal inertia, percent of total	
		$\theta_y$	$\theta_z$		$\theta_y$	$\theta_z$
1	1.281	-	1.1	1.281	-	1.0
2	1.489	12.1	5.6	1.594	24.8	2.7
3	1.719	13.8	23.9	1.769	2.2	60.8

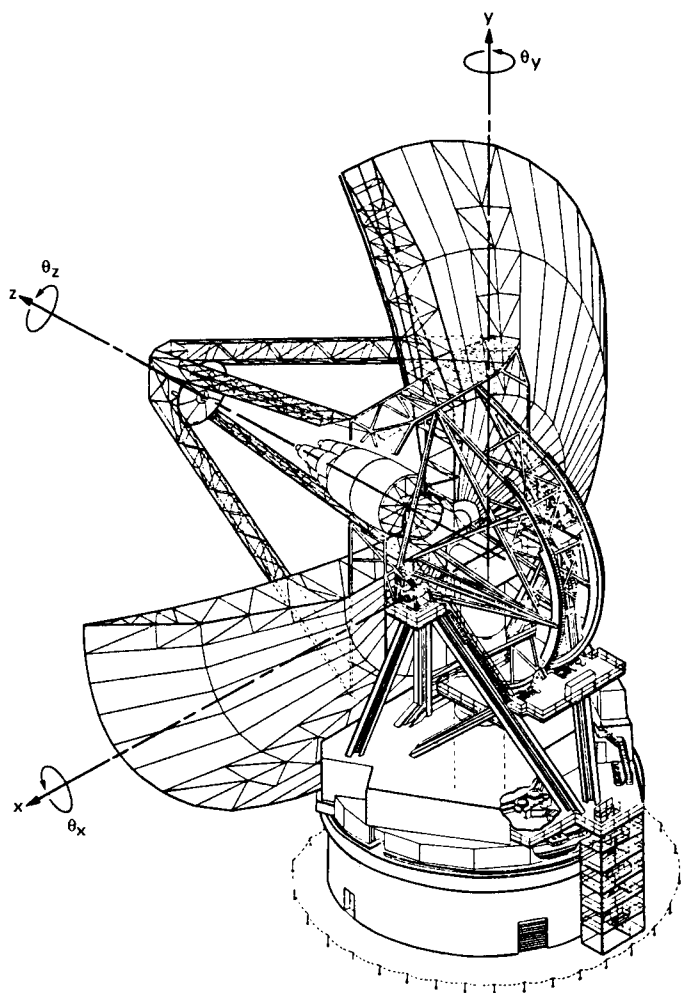
**Table 3. Comparison of anti-symmetric modes before and after removal of tie-truss center section**

Mode	Before removal			After removal			Mode shape
	Frequency, Hz	Effective modal inertia, percent of total		Frequency, Hz	Effective modal inertia, percent of total		
		$\theta_y$	$\theta_z$		$\theta_y$	$\theta_z$	
1	1.281	-	1.0	1.281	-	1.1	Yaw (Quad Torsion)
2	1.491	12.7	4.5	1.489	12.1	5.6	Roll + Yaw
3	1.736	13.0	21.5	1.719	13.8	23.9	Yaw + Roll

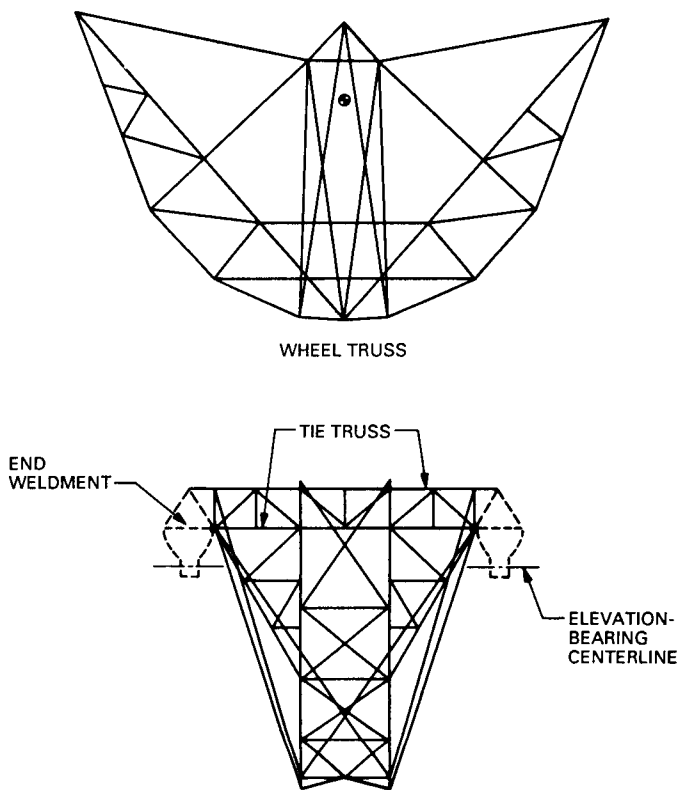
**Table 4. Comparison of symmetric modes before and after removal of tie-truss center section**

Mode	Before removal		After removal		Mode shape
	Frequency, Hz	Effective modal inertia, percent of total $\theta_x$	Frequency, Hz	Effective modal inertia, percent of total $\theta_x$	
1	2.389	41.6	2.381	42.4	Pitch
2	2.853	10.9	2.847	10.5	Pitch
3	3.262	2.3	3.255	2.1	Pitch

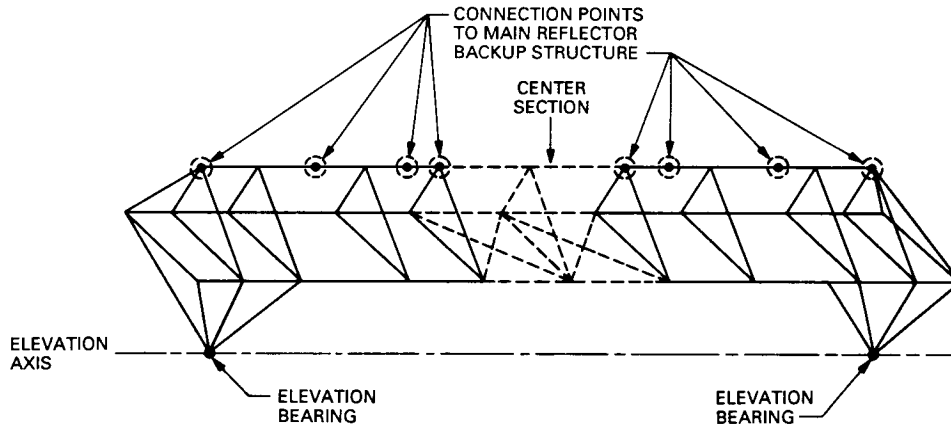




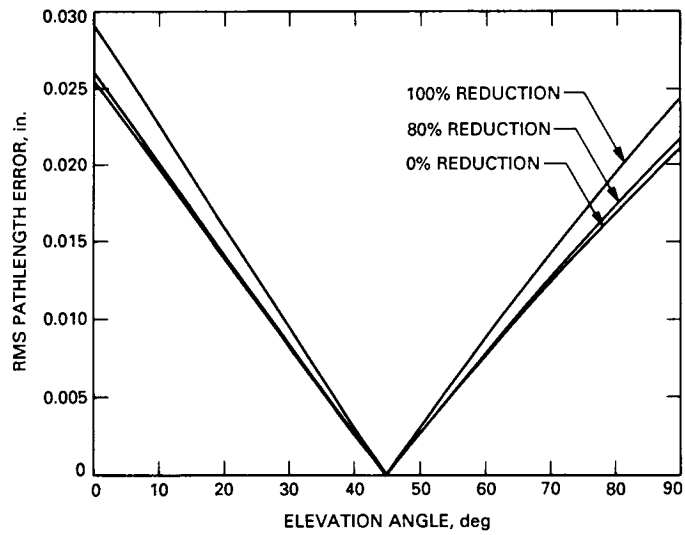
**Fig. 1. Isometric view of the 70-m antenna showing coordinate system of the tipping structure.**



**Fig. 2. 70-m antenna elevation wheel with tie truss.**



**Fig. 3. Simplified view of tie truss showing region to be modified.**



**Fig. 4. Gravity RMS pathlength error for a series of bar-area (stiffness) reductions.**

547641  
P10  
N90-12792  
56-32  
239816

98

# A Performance Comparison Between Block Interleaved and Helically Interleaved Concatenated Coding Systems

K.-M. Cheung and L. Swanson  
Communications Systems Research Section

*This article compares the performance (bit-error rate vs. signal-to-noise ratio) of two different interleaving systems, block interleaving and the newer helical interleaving. Both systems are studied with and without error forecasting. Without error forecasting, the two systems have identical performance. When error forecasting is used with shallow interleaving, helical interleaving gains, but less than 0.05 dB, over block interleaving. For higher interleaving depth, the systems have almost indistinguishable performance.*

## I. Introduction

As NASA wants to receive more data from planetary missions, and as technologies like data compression make projects tighten their error-rate requirements, many missions are turning to concatenated coding schemes with Reed-Solomon (RS) codes as "outer" codes concatenated with "inner" convolutional codes, which have been used in the Deep Space Network for years (Fig. 1). RS decoders for use beginning in the mid-1990s will be installed in the network by February 1, 1992.

When concatenated coding is used, symbols in RS words are generally "interleaved," so that the error cor-

recting ability of one RS word is not used up entirely by one or two long error bursts from the Viterbi (convolutional) decoder. Some missions planning to use RS codes in the Deep Space Network will have "block" interleaving depth five. This means that consecutive symbols in a RS word are separated on the channel by exactly four other RS symbols, one from each of the other four words in the block (Fig. 2).

Recently, other interleaving schemes have been suggested, such as E. Berlekamp's "helical interleaving" [1, 2]. In this article, the initial analysis of this scheme is presented. In helical interleaving, words go into the interleaver/deinterleaver in the staggered way shown in Fig 3.

(Notice that if a block-interleaved code block were pasted onto a cylinder, the set of symbols consisting of the  $j$ th symbol in each word lies on a circle around the cylinder; if the helically interleaved words were pasted onto a cylinder, the set consisting of the  $j$ th symbol in each word lies on a helix.) Helical interleaving is defined precisely in Section II.

It is worth noticing that for a given interleaving depth, the two schemes yield statistically identical code performance. As far as any one codeword is concerned, it consists of certain RS symbols; its decoding does not depend on which other codewords the intervening symbols belong to. So why this apparently more complicated interleaving scheme? There are two ways in which we might expect helical interleaving to be different from block interleaving. The first is synchronization: if a synchronization marker is placed before or at the beginning of each codeword in a block-interleaved block, then several markers appear consecutively at the beginning of each code block; with helical interleaving, the markers would be uniformly scattered. This should allow faster (though possibly more complicated) synchronization, but we do not address this issue here. The second way in which helical interleaving can be expected to perform differently from block interleaving is in error forecasting. For several years, the DSN's planning for RS decoders has taken into account the slight performance gain available from the fact that RS codes can use quality information in decoding; a symbol that is believed questionable and flagged as an "erasure" costs only half an error [3].

The error-forecasting scheme used in this article is a simple but effective one: that is to create erasures in those words that have too many symbol errors [4] by assuming that errors in adjacent words continue in the undecodable word. Based on the above idea, there are two ways to flag an erasure; that is, either flag a symbol of an undecodable word as an erasure if either one of the code symbols next to it on the channel is detected as erroneous (single-sided forecasting), or flag a symbol as an erasure only if the code symbols on both sides of it on the channel are detected as erroneous (double-sided forecasting). Simulations show that in all cases the more aggressive single-sided forecasting scheme performs better than the conservative double-sided forecasting scheme. Thus we choose to use the single-sided forecasting scheme in all our simulations in this article. Since the Viterbi decoder, which decodes the inner convolutional code, creates errors in bursts, previous analysis [5] shows that in the case of block interleaving, using this simple error-forecasting strategy allows a gain of about 0.04 dB.

For block interleaved data, words go into the interleaver as shown in Fig. 2. Each interleaving block is separated from others, and all possible error forecasting can be done within the same block. Thus a finite buffer can do all error forecasting. In the light of this finite buffer feature, error forecasting with redecoding to each interleaving block was selected, disregarding the order of decoding of codewords in each block. That is, when a codeword in a block is decoded successfully, erasure information is generated for both of its adjacent words in the block. After the first round of decoding in a block, the previously undecoded codewords are then redecoded using the additional erasure information. This process goes on until no more undecodable words can be recovered. For helically interleaved data, words go into the interleaver in the staggered way shown in Fig. 3. Unlike the block interleaving scheme in which each codeword is "related" only to words in the same block, each codeword in the helical interleaving scheme is theoretically "related" to all words before and after it, and this makes error forecasting with redecoding impractical, if not impossible, since it would require a buffer with infinite memory. Thus in helical interleaving one does error forecasting without redecoding. That is, when a codeword is successfully decoded, erasure information is generated to those symbols (of other not yet decoded codewords) that are adjacent to the previously erroneous symbols of the decoded codeword. Each codeword is then decoded in order, first without using the erasure information. If the codeword is decodable, the decoder goes on to decode the next codeword. If not, the codeword is decoded once again using the erasure information generated from previously decoded words. Whether the codeword is decodable or not (with erasure information), the decoder goes on to decode the next codeword.

For deep interleaving, both block and helical interleaving give nearly identical error forecasting capability because both schemes are essentially "ideal interleaving" schemes which allow nearly statistically independent RS symbols. But for shallow helical interleaving (e.g., depth 2, which is used by Galileo), a word with too many symbol errors to decode is preceded by a part of one word and part of another. On the one hand, this means that a Viterbi decoder error burst is less likely to keep both different codewords from decoding, and so the error forecasting capability should be enhanced. On the other hand, only some of the symbols in a given codeword are preceded by symbols in a codeword that is decoded before it. For the exact schemes we compare in Section III, the trade-off between these effects depends on the interleaving depths and the inner convolutional codes. For interleaving depth 2, helical interleaving gives a slight performance gain over block interleaving with both the  $(7,1/2)$  and  $(15,1/4)$  inner

convolutional codes. For interleaving depth 4 or 8, helical interleaving is slightly better or slightly worse than block interleaving, depending on whether the (15,1/4) code or the (7,1/2) code is used as the inner code.

## II. Helical Interleaving

In this section, mathematical definitions of the interleaving schemes are included for completeness, but the reader is advised to consider Figs. 2 and 3 to illustrate these definitions.

**Definition 1.** A block code with  $n$  code symbols per codeword is block interleaved to depth  $d$  if the code symbols sent on the channel are packed into codewords in such a way that the code symbol that follows (on the channel) the  $j$ th symbol of the  $m$ th codeword,  $m < d$ , is the  $j$ th code symbol of the  $(m + 1)$ th codeword, and the code symbol that follows the  $j$ th code symbol of the  $d$ th codeword,  $j < n$ , is the  $(j + 1)$ th code symbol of the first codeword, and the code symbol following the  $n$ th code symbol of the  $d$ th codeword is the first symbol of the first codeword in the next block.

**Definition 2.** A block code with  $n$  code symbols per codeword is helically interleaved to depth  $d$ , where  $dr = n - 1$  and  $r$  is any integer, if code symbols are packed into codewords in such a way that the code symbol that follows (on the channel) the  $j$ th code symbol of the  $m$ th word,  $m < d$ , is the  $(j - r)$ th code symbol of the  $(m + 1)$ th codeword; if  $j - r \leq 0$ , then this is the  $(j - r + n)$ th code symbol of the  $m$ th codeword of the "previous" group. The  $j$ th code symbol of the  $d$ th codeword is followed by the  $(j + (d - 1)r)$ th code symbol of the first codeword of this group if  $j \leq r$  and by the  $(j - r)$ th code symbol of the first codeword of the next group if  $j > r$ .

Helical interleaving can be defined for codeword length  $n$  and interleaving depth  $d$  as long as  $n$  and  $d$  are relatively prime. While this more complicated definition for interleaving depths 4 and 8 was needed in the simulations, technical definitions will not be written here. The interested reader can find them in [1].

Helical interleaving was introduced in 1982 [1, 2] by E. Berlekamp. It has since been successfully implemented in several hardware decoders. Figure 3 depicts a helical interleaver for a code of length 4 interleaved to depth 3. As discussed in Section I, use of a helical interleaving scheme

in general does not enhance the gain performance of a concatenated coding system.

For many applications, it is better than block interleaving because of synchronization and interleaving delay. Helical interleaving facilitates synchronization in the following way: in block interleaving, the  $j$ th symbols of two adjacent codewords in a block are adjacent to each other in the channel; in helical interleaving, the  $j$ th symbols of two adjacent codewords are separated from each other in the channel by  $n - 1$  symbols in a block-interleaved system. If a synchronization marker is placed before or at the beginning of each codeword in a block-interleaved block, several markers appear consecutively at the beginning of each code block in the channel symbol streams, and the receiver can acquire only synchronization modulo  $nd$ . With helical interleaving, the markers are uniformly scattered, and this allows synchronization modulo  $n$ . The difference in interleaving delay is that the end-to-end delay of a block interleaver is  $2nd$ , exclusive of the channel delay, whereas the delay of a helical interleaver is only  $nd$ .

## III. Simulation

The performance of the concatenated coding systems described in this section were obtained by simulation. The inner codes used were the standard (7,1/2) convolutional code, which was simulated with a software Viterbi decoder, and Galileo's experimental (15,1/4) code, which was decoded using the Communications Systems Research Section's long constraint length "Little Viterbi Decoder" built for Advanced Systems. The outer codes used were the standard (255,223) RS code, which has been adopted for use by a number of present and future missions, and the proposed (1023,959) RS code for deep-space missions of the future.

Figures 4 through 11 give the SNR performance comparisons of various concatenated schemes using block interleaving and helical interleaving. Figures 4 and 5 show the performance curves of the concatenated coding schemes using the (7,1/2) convolutional code as inner code, and the (255,223) RS code as outer code, without and with error forecasting respectively. Similarly, Figs. 6 and 7 show the performance curves of the concatenated coding schemes with the (7,1/2) convolutional code and the (1023,959) RS code. Figures 8 and 9 show the performance curves of the concatenated schemes using the (15,1/4) convolutional code and the (255,223) RS code. And finally, Figs. 10 and 11 show the performance curves of the concatenated coding schemes using the (15,1/4) convolutional code and the (1023,959) RS code.

## IV. Conclusion

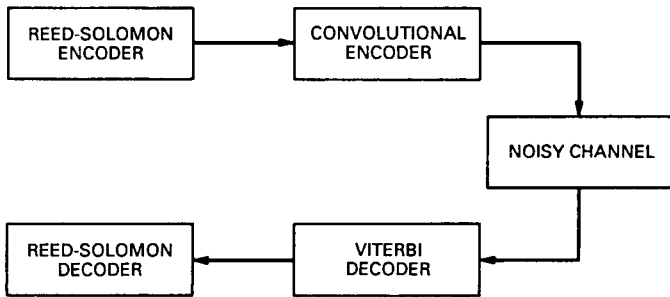
It is observed from Figs. 4 through 11 that, as pointed out in Section I, when no error-forecasting strategy is used, both helical interleaving and block interleaving perform identically for all interleaving depths. When error forecasting is used, helical interleaving gains less than 0.05 dB at shallow interleaving. For deeper interleaving (e.g., depth 4 or depth 8), helical interleaving is slightly worse or slightly better than block interleaving, depending on whether the (7,1/2) convolutional code or the (15,1/4) convolutional

code is used as inner code. The difference in performance between helical and block interleaving for higher depths might be attributed to the fact that the (15,1/4) code produces longer bursts than the (7,1/2) code.

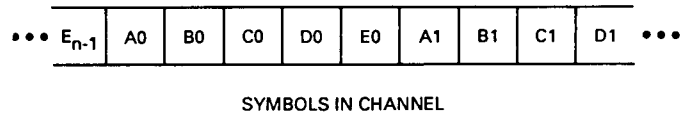
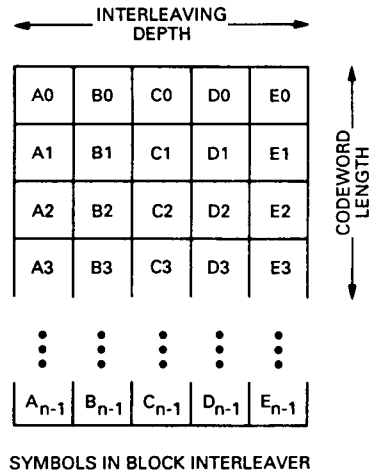
On the basis of these results, it is recommended that helical interleaving be considered for possible use only in particular situations, and not for general use. As usual, the real moral seems to be that one should allow greater interleaving depth, whatever interleaving scheme is used.

## References

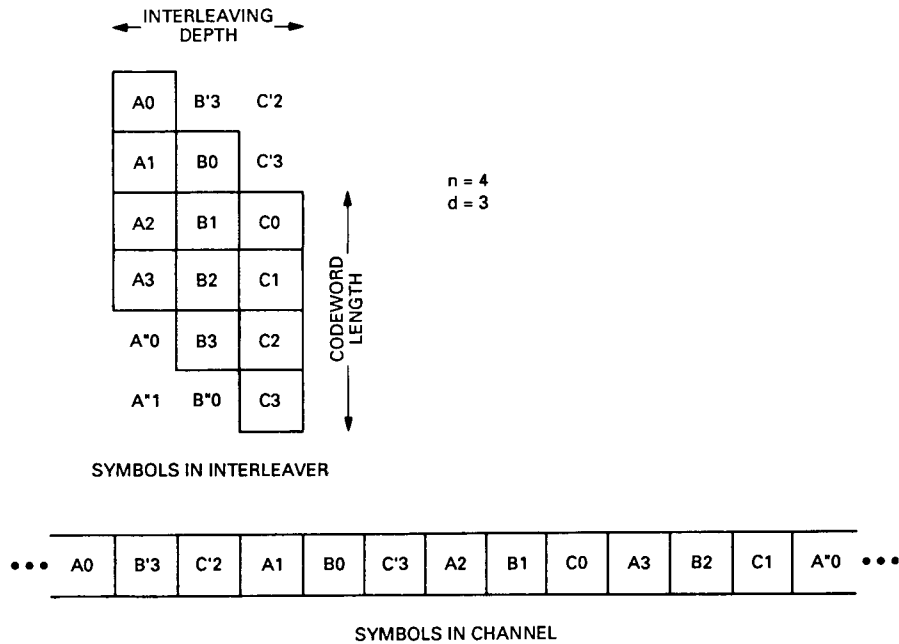
- [1] Cyclotomics Corporation, *Interleaved Coding for Bursty Channels*, Final Report on Phase I, Small Business Innovation Research, NSF Grant No. ECS-8260180, Berkeley, California, April 12, 1983.
- [2] E. R. Berlekamp and P. Tong, *Improved Interleavers for Digital Communications*, United States Patent Number 4559625 (granted December 17, 1985).
- [3] R. E. Blahut, *Theory and Practice of Error Control Codes*, Reading, Massachusetts: Addison Wesley, 1983.
- [4] R. J. McEliece and L. Swanson, "On the Decoder Error Probability for Reed-Solomon Codes," *IEEE Trans. Inform. Theory*, IT-32, pp. 701-703, September 1986.
- [5] G. H. Pitt, III, and L. Swanson, "Erasure Information for a Reed-Solomon Decoder," *TDA Progress Report 42-83*, Jet Propulsion Laboratory, Pasadena, California, pp. 39-44, November 15, 1985.



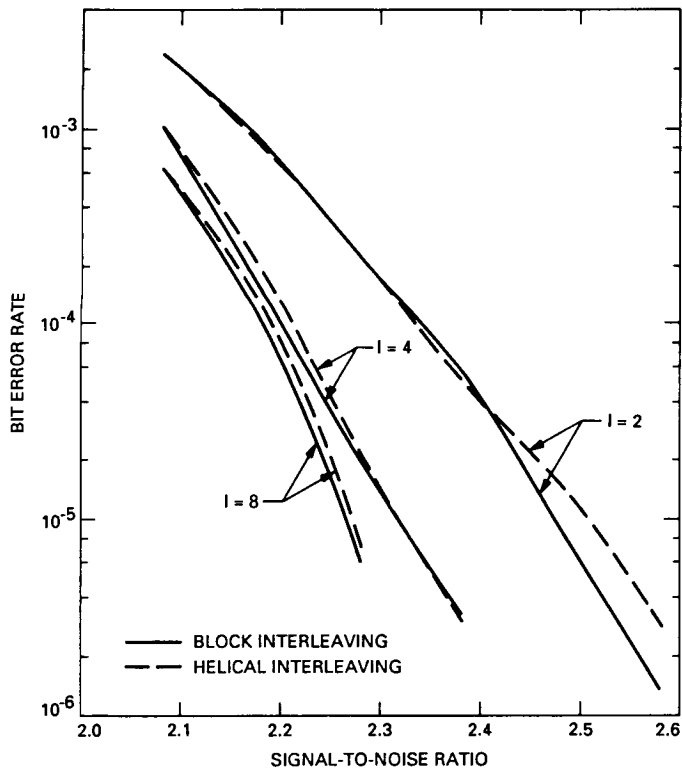
**Fig. 1. Concatenated coding system.**



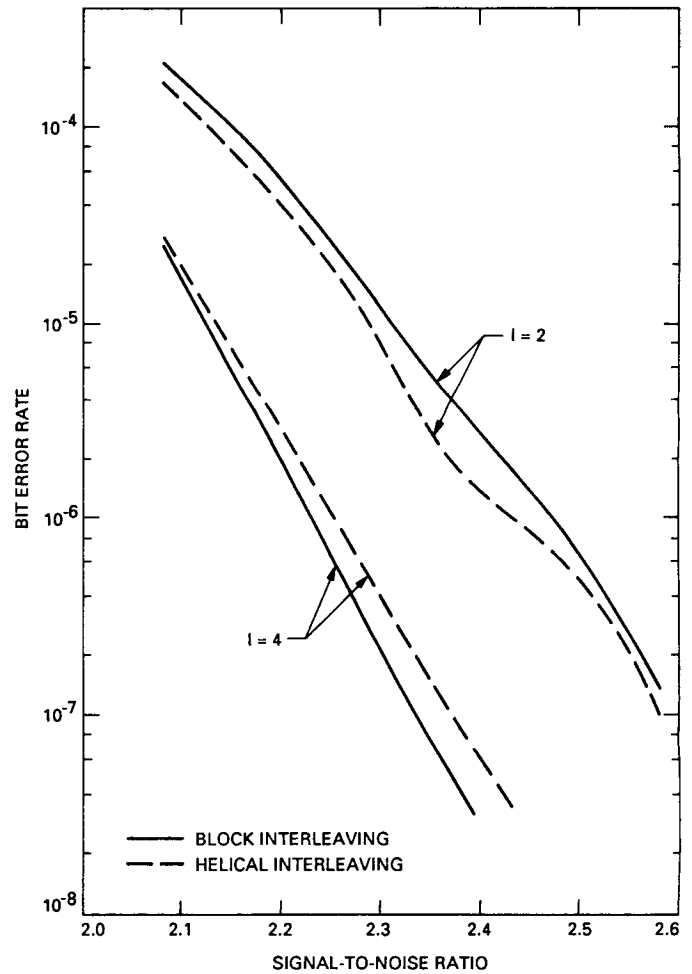
**Fig. 2. Example of block interleaving.**



**Fig. 3. Example of helical interleaving.**

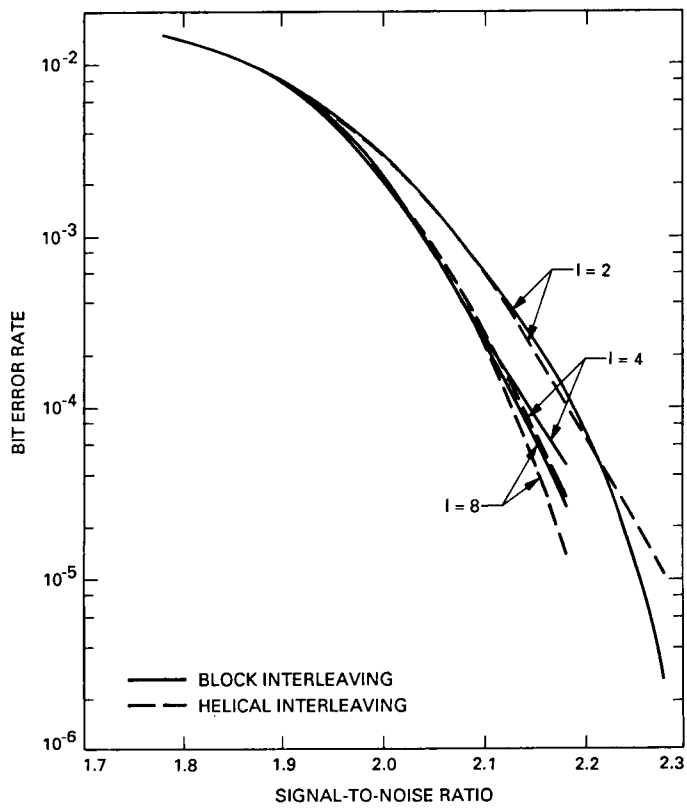


**Fig. 4. (7,1/2) convolutional code and (255,223) Reed-Solomon code; no error forecasting.**

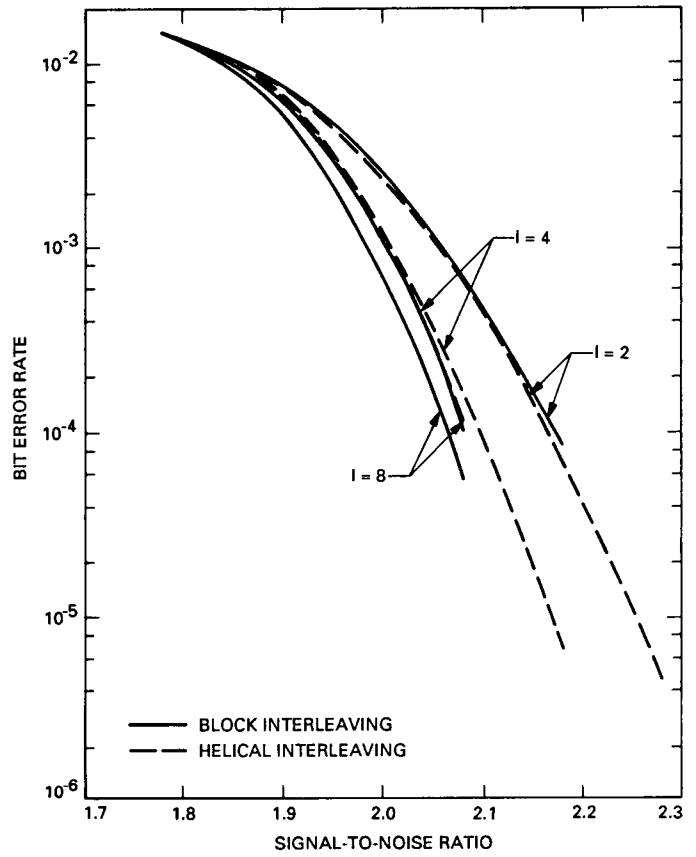


**Fig. 5. (7,1/2) convolutional code and (255,223) Reed-Solomon code with error forecasting.**





**Fig. 6. (7,1/2) convolutional code and (1023,959) Reed-Solomon code; no error forecasting.**



**Fig. 7. (7,1/2) convolutional code and (1023,959) Reed-Solomon code with error forecasting.**

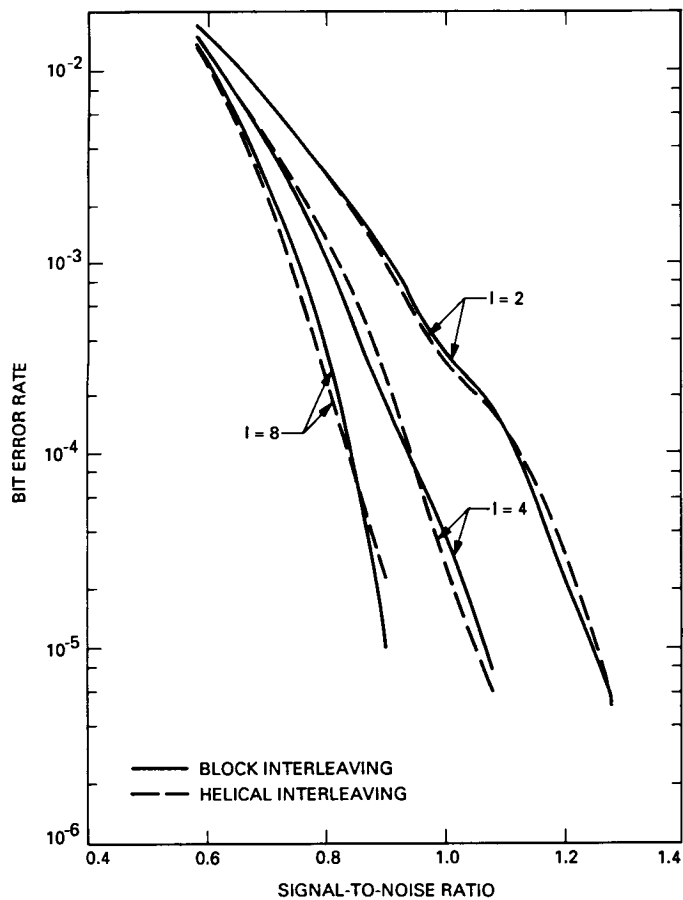


Fig. 8. (15,1/4) convolutional code and (255,223) Reed-Solomon code; no error forecasting.

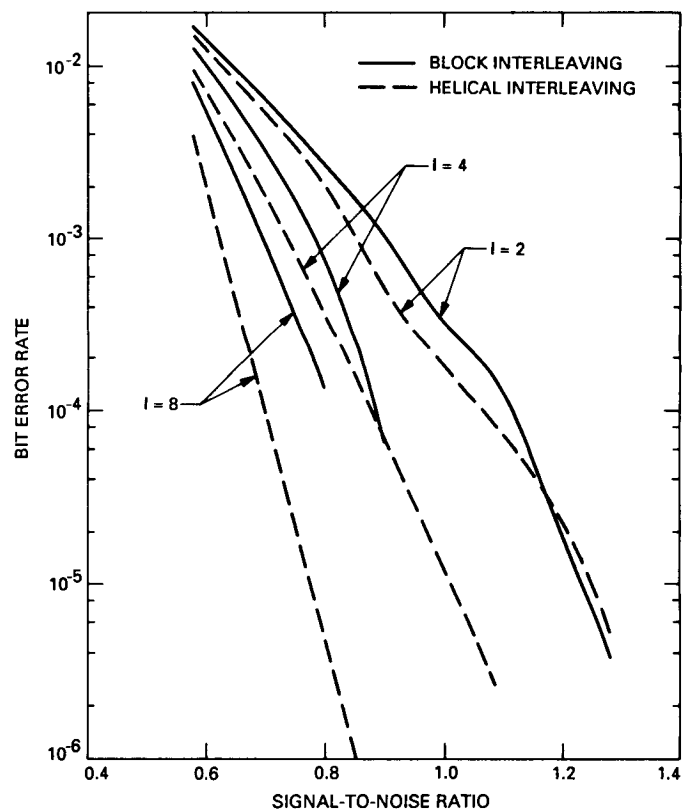
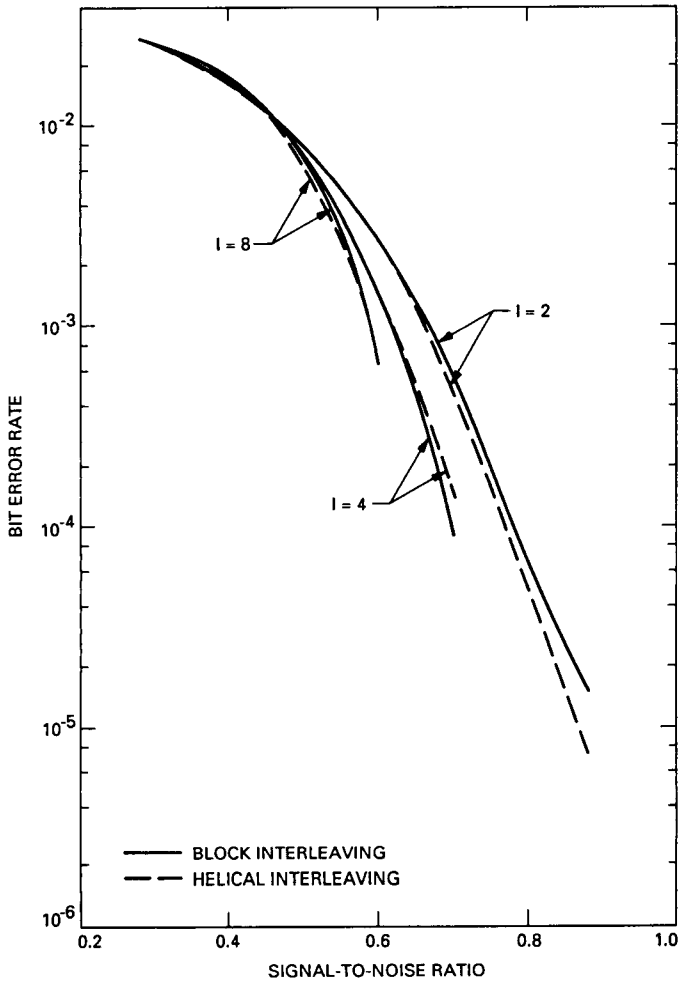
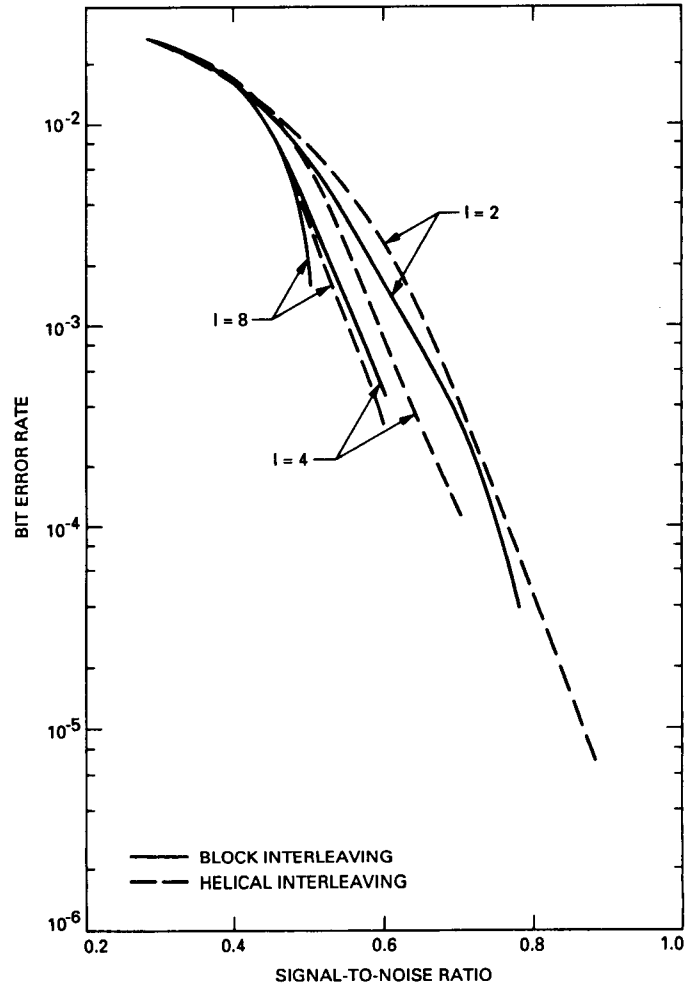


Fig. 9. (15,1/4) convolutional code (255,223) Reed-Solomon code with error forecasting.



**Fig. 10. (15,1/4) convolutional code (1023,959) Reed-Solomon code; no error forecasting.**



**Fig. 11. (15,1/4) convolutional code (1023,959) Reed-Solomon code with error forecasting.**

57-32

239817

547655p5

N90-12793

68

# On the Decoder Error Probability of Linear Codes

K.-M. Cheung

Communications Systems Research Section

*In this article, by using coding and combinatorial techniques, an approximate formula for the weight distribution of decodable words of most linear block codes is evaluated. This formula is then used to give an approximate expression for the decoder error probability  $P_E(u)$  of linear block codes, given that an error pattern of weight  $u$  has occurred. It is shown that  $P_E(u)$  approaches the constant  $Q$  as  $u$  gets large, where  $Q$  is the probability that a completely random error pattern will cause decoder error.*

## I. Introduction

Coding is used in a digital communication system to detect and correct errors introduced in the data stream by channel noise. An important parameter to evaluate the performance of a code is its decoder error probability. Let  $C$  be a linear  $(n, k, d)$  code over  $GF(q)$ , and  $C^\perp$  be its  $(n, n-k, d^\perp)$  dual code. Let  $t$  be the number of errors the code is designed to correct. Let  $G$  be the generator matrix of  $C$ . Let  $A_u$  denote the number of codewords of weight  $u$ , and  $D_u$  denote the number of decodable words of weight  $u$ . Decodable words are defined as all words lying within distance  $t$  from a codeword. If the decoder is assumed to be a bounded distance decoder, then the weight distribution for the decodable words can be used to find the decoder error probability of the code.

When a codeword  $\underline{c} \in C$  is transmitted over a communication channel, channel noise may corrupt the trans-

mitted signals. As a result, the receiver gets a corrupted version of the transmitted codeword  $\underline{c} + \underline{e}$ , where  $\underline{e}$  is an error pattern of some weight  $u$ . If  $u \leq t$ , then a bounded distance decoder on the receiver's end detects and corrects the error  $\underline{e}$ , and recovers  $\underline{c}$ . If  $u > t$ , the decoder fails, and it either

- (1) Detects the presence of the error pattern  $\underline{e}$ , but is unable to correct it, or
- (2) Misinterprets the received pattern  $\underline{c} + \underline{e}$  for some other codeword  $\underline{c}'$  if the received pattern falls into the radius  $t$  of the Hamming sphere around  $\underline{c}'$ .

Case (2) is, in most cases, more serious than case (1). This can occur when an error pattern  $\underline{e}$  is of weight  $u \geq d - t$ . As pointed out in [1] and [2], if all error patterns of weight  $u$  are equally probable, the decoder error probabil-

ity given that an error pattern of weight  $u$  occurs, denoted by  $P_E(u)$ , is given by the following expression:

$$P_E(u) = \frac{D_u}{\binom{n}{u}(q-1)^u} \quad d-t \leq u \leq n \quad (1)$$

In this article, by using combinatorial and coding techniques, an approximate formula for the weight distribution of decodable words for most linear block codes is evaluated. This formula together with Eq. (1) gives an approximate formula for the decoder error probability  $P_E(u)$  for most linear block codes. It is also shown that

$$P_E(u) \xrightarrow{\text{large } u} Q$$

where  $Q$  is the probability that a completely random error pattern will cause decoder error. That is,

$$Q = \frac{(q^k - 1)V_n(t)}{q^n} \approx q^{-r} V_n(t) \quad (2)$$

where  $r = n - k$  is the code's redundancy and  $V_n(t) = \sum_{i=0}^t \binom{n}{i}(q-1)^i$  is the volume of a Hamming sphere of radius  $t$ .

## II. Mathematical Preliminaries

In this section, combinatorial and coding techniques required to derive the results in later sections are introduced. These techniques are similar to those used in [6] to obtain the weight distribution of linear block codes.

### A. Principle of Inclusion and Exclusion

Let  $\chi$  be a set of  $N$  objects, and  $P(1), P(2), \dots, P(u)$  be a set of  $u$  properties. Let  $N(i_1, i_2, \dots, i_r)$  be the number of objects with properties  $P(i_1), P(i_2), \dots, P(i_r)$ . The number of objects  $N(\emptyset)$  with none of the properties is given by [3]:

$$\begin{aligned} N(\emptyset) = & N - \sum_i N(i) + \sum_{i_1 < i_2} N(i_1, i_2) + \dots + (-1)^r \\ & \times \sum_{i_1 < i_2 < \dots < i_r} N(i_1, i_2, \dots, i_r) + \dots \\ & + (-1)^u N(1, 2, 3, \dots, u) \end{aligned} \quad (3)$$

There are  $u + 1$  terms on the right-hand side of Eq. (3), with the 0th term representing the total number of objects in  $\chi$ . If all terms beyond the  $r$ th term on the right-hand side of Eq. (3) are ignored, then the resulting truncated sum is an upper bound when  $r$  is even, or a lower bound if  $r$  is odd. Thus the maximum error magnitude introduced in the inclusion and exclusion formula by ignoring all terms beyond the  $r$ th term does not exceed the magnitude of the  $(r+1)$ th term. This fact will be used later to upper bound the magnitude of the errors of the approximate weight distribution formula.

### B. Facts on Coding Theory

A linear  $(n, k, d)$  code over  $GF(q)$  can be generated by a  $k \times n$  generator matrix  $G$ , not necessarily unique and such that  $\text{rank}(G) = k$ . Let  $l$  be the maximum number such that no  $l$  or fewer columns of  $G$  add to zero. Then

$$l \leq k \quad (4)$$

Equality in Eq. (4) is achieved in the case of *maximum distance separable* (MDS) codes. Since  $G$  is the parity-check matrix of  $C^\perp$ ,  $l = d^\perp - 1$ . Let  $\text{col}_{i_1}, \text{col}_{i_2}, \dots, \text{col}_{i_j}$  be any  $j$  particular columns of  $G$ ,  $j \leq l \leq k$ . It is obvious that there exists a  $k \times n$  generator matrix  $G'$  of  $C$  and a  $k \times k$  nonsingular matrix  $K$  such that

$$G' = KG \quad (5)$$

and  $\text{col}_{i_1}, \text{col}_{i_2}, \dots, \text{col}_{i_j}$  of  $G'$  form a  $k \times j$  submatrix of the form  $\begin{pmatrix} I \\ \dots \\ 0 \end{pmatrix}$ . This fact guarantees that given any pattern of  $j$  symbols on the  $i_1$ th,  $i_2$ th,  $\dots$ ,  $i_j$ th coordinates, the number of codewords with the  $j$ -symbol pattern on the  $i_1$ th,  $i_2$ th,  $\dots$ ,  $i_j$ th coordinates equals  $q^{k-j}$  for  $j \leq l$ . This fact is important in the next section to evaluate the cardinalities of some sets of decodable words.

## III. Derivation of Formulae

Let  $D$  be the set of decodable words of  $C$ . Let  $\underline{d}$  be a decodable word with Hamming weight  $u$ ,  $u \geq n-l$ . Let the coordinates of  $\underline{d}$  be indexed by  $\{0, 1, \dots, n-1\}$ . Then  $\underline{d}$  has  $v$  zeros ( $v \leq l$ ), where  $v = n - u$ . Let  $V$  be a set of  $v$  coordinates,  $|V| = v$ . Let  $\{i_1, i_2, \dots, i_j\} \subseteq \{0, 1, \dots, n-1\} - V$  be a set of  $j$  coordinates. Define  $S(i_1, i_2, \dots, i_j) = \{\underline{d} :$

$\underline{d} \in D$  and  $\underline{d}$  has zeros in  $V \cup \{i_1, i_2, \dots, i_j\}$ . A decodable word  $\underline{d} \in S(i_1, i_2, \dots, i_j)$  always has at least  $v + j$  zeros. For  $0 \leq j \leq l - v$ , the number of zeros in the decodable words of  $S(i_1, i_2, \dots, i_j)$  is less than or equal to  $l$ . Now, since all words lying within the Hamming spheres (with volume  $V_n(t)$ ) that surround codewords are decodable words, there are  $V_n(t)$  disjoint cosets that contain decodable words. Each coset can be constructed by adding a coset leader  $\underline{a}$  (Hamming weight of  $\underline{a} \leq t$ ) to each codeword in  $C$ . Thus from the discussion in Section II.B, for each of the  $V_n(t)$  different coset leaders (each corresponding to a coset), there are  $q^{k-v-j}$  codewords in  $C$  which, when added to the coset leader, give decodable words with zeros in the  $v + j$  coordinates. The number of decodable words in  $S(i_1, i_2, \dots, i_j)$  is then given by

$$|S(i_1, i_2, \dots, i_j)| = q^{k-v-j} V_n(t) \quad \text{for } 0 \leq j \leq l - v \quad (6)$$

For  $l - v + 1 \leq j \leq n - v - d + t$ , the number of zeros in the decodable words of  $S(i_1, i_2, \dots, i_j)$  exceeds  $l$ , and apparently there is no simple expression for  $|S(i_1, i_2, \dots, i_j)|$ . For  $n - v - d + t + 1 \leq j \leq n - v - t$ , the number of zeros in a decodable word is greater than or equal to  $n - d + t + 1$ , but less than or equal to  $n - t$ . Thus any decodable word in  $S(i_1, i_2, \dots, i_j)$  has weight less than or equal to  $d - t - 1$ . It is not hard to see that the elements of  $S(i_1, i_2, \dots, i_j)$  cannot be decoded into a codeword of weight other than 0. Therefore,  $S(i_1, i_2, \dots, i_j)$  contains all words of weight less than or equal to  $t$  in the coordinates  $\{0, 1, \dots, n - 1\} - (V \cup \{i_1, i_2, \dots, i_j\})$ . Thus,

$$|S(i_1, i_2, \dots, i_j)| = \sum_{i=0}^t \binom{u-j}{i} (q-1)^i \quad \text{for } n - v - d + t + 1 \leq j \leq n - v - t \quad (7)$$

For  $n - v - t + 1 \leq j \leq n - v$ , since  $j$  is greater than or equal to  $n - v - t + 1$ , the number of zeros  $v + j$  is greater than or equal to  $n - t + 1$ . Therefore, the number of nonzero components is less than  $t$ . Thus, all words with zeros on  $V \cup \{i_1, i_2, \dots, i_j\}$  are decodable and thus

$$|S(i_1, i_2, \dots, i_j)| = q^{u-j} \quad \text{for } n - v - t + 1 \leq j \leq n - v \quad (8)$$

In the cases for  $0 \leq j \leq l - v$ ,  $n - v - d + t + 1 \leq j \leq n - v - t$ , and  $n - v - t + 1 \leq j \leq n - v$ , the set  $i_1, i_2, \dots, i_j$  can be chosen arbitrarily from a set of  $u = n - v$  coordinates. Thus for every choice of  $j$ , there are  $\binom{u}{j}$  sets  $S(i_1, i_2, \dots, i_j)$ . By the principle of inclusion and exclusion, the number of decodable words with exactly  $v$  zeros in  $V$ , which is denoted by  $D'_V$ , is:

$$\begin{aligned} D'_V &= |S(\emptyset)| + (-1) \sum_{i_1} |S(i_1)| + \dots + (-1)^r \\ &\quad \times \sum_{i_1 < i_2 < \dots < i_r} |S(i_1, i_2, \dots, i_r)| \\ &\quad + \dots + (-1)^{n-v} S(i_1, i_2, \dots, i_{n-v}) \\ &= \sum_{j=0}^{l-v} (-1)^j \binom{u}{j} q^{k-v-j} V_n(t) + \sum_{j=t-v+1}^{n-v-d+t} (-1)^j \\ &\quad \times \sum_{i_1 < i_2 < \dots < i_j} |S(i_1, i_2, \dots, i_j)| \\ &\quad + \sum_{j=n-v-d+t+1}^{n-v-t} (-1)^j \binom{u}{j} \sum_{i=0}^t \binom{n-v-j}{i} (q-1)^i \\ &\quad + \sum_{j=n-v-t+1}^{n-v} (-1)^j \binom{u}{j} q^{u-j} \end{aligned} \quad (9)$$

If all the terms beyond the  $l - v - 1$  terms are ignored in the above inclusion and exclusion formula, Eq. (9) is reduced to

$$D'_V = \sum_{j=0}^{l-v-1} (-1)^j \binom{u}{j} q^{k-v-j} V_n(t) + E_1 \quad (10)$$

where

$$E_1 = (-1)^{l-v} \binom{u}{l-v} q^{k-l} V_n(t) + \sum_{j=l-v+1}^{n-v} \sum_{i_1 < \dots < i_j} (-1)^j |S(i_1, \dots, i_j)|$$

and  $|E_1| \leq \binom{u}{l-v} q^{k-l} V_n(t)$  (from the discussion in Section II.A). If  $E_2 = \sum_{j=l-v}^u (-1)^j \binom{u}{j} q^{k-v-j} V_n(t)$  is added and subtracted from Eq. (10), one has

$$D'_V = \frac{(q-1)^u}{q^{n-k}} V_n(t) + E_1 + E_2 \quad (11)$$

If  $\binom{u}{l-v} q \geq \binom{u}{l-v+1}$ , that is, if  $u \geq \frac{q+1}{q}(n-l) - 1$ ,  $E_2$  is a sum of terms with alternate signs and descending magnitude. Then  $|E_2| \leq \binom{u}{l-v} q^{k-l} V_n(t)$ . Thus

$$D'_V = \frac{(q-1)^u}{q^{n-k}} V_n(t) + E \quad (12)$$

where  $E = E_1 + E_2$  and  $|E| \leq 2 \binom{u}{l-v} q^{k-l} V_n(t)$ .  $D'_V$  can thus be approximated by  $\frac{(q-1)^u}{q^{n-k}} V_n(t)$ , and the goodness of approximation depends on how small the ratio  $R = E / [(q-1)^u q^{-(n-k)} V_n(t)]$  is. By using the upper bound on  $|E|$ , an upper bound on this ratio is given by

$$R \leq \frac{2 \binom{u}{n-l} q^{k-l}}{(q-1)^u} \quad (13)$$

Since  $v \leq l$ , there are  $\binom{n}{v} = \binom{n}{n-v}$  ways to choose  $v$  zeros from  $\{0, 1, \dots, n-1\}$ . Then  $D_u$  can be approximated by the following expression:

$$D_u = \sum_{|V|=n-u} D'_V \approx q^{-(n-k)} \binom{n}{u} (q-1)^u V_n(t) \quad (14)$$

for  $u \geq \max\{n-l, \frac{q+1}{q}(n-l) - 1\}$ .

Strictly speaking, the derivation of Eq. (14) is valid only for  $u \geq \max\{n-l, \frac{q+1}{q}(n-l) - 1\}$ . However, it is observed that in most cases Eq. (14) is also a close approximation to  $D_u$  for  $u$  considerably smaller than  $n-l$  (as in the case of Reed-Solomon codes). The upper bound of  $R$  derived above has a denominator term  $(q-1)^u$  and this indicates that this approximation formula is good for nonbinary linear codes, and is not useful for binary linear codes. The looseness of this approximation for binary linear codes is best illustrated by extended binary codes that have only even weights. In the case of binary primitive codes, Kasami et al. [4] generalized Sidel'nikov's approach [5] and showed that the weights of most binary primitive codes have approximate binomial distribution.

Cheung [6] later showed this is also true for nonbinary codes. It is conjectured in this article that the approximate Eq. (14) for the weight distribution of decodable words is also good for binary primitive codes. For nonbinary linear codes, the upper bound on  $R$  shows that the approximation in Eq. (14) is particularly good for codes with large alphabet sets. The upper bound on  $R$  for the (31,15,17) Reed-Solomon code over  $GF(32)$  is given in Table 1. The weight distribution of decodable words and its approximation (using Eq. 14) of the (31,15,17) Reed-Solomon code are given in Table 2.

Given the approximate formula of  $D_u$ , an approximate decoder error probability  $P_E(u)$  is obtained by substituting Eq. (14) into Eq. (1). It is observed that  $P_E(u)$  approximates the constant  $Q = q^{-r} V_n(t)$  as  $u$  gets large, where  $Q$  is the probability that a completely random error pattern will cause decoder error. An upper bound of  $R$  given by Eq. (13) shows that  $P_E(u)$  approaches  $Q$  "nearly exponentially" (for nonbinary codes) as  $u$  increases.

## IV. Conclusion

In this article, by using the inclusion and exclusion principle, an approximate formula for the weight distribution of decodable words of most linear block codes is derived. The decoder error probability  $P_E(u)$ , which is a function of  $D_u$ , is then shown to approach the constant  $Q$  as  $u$  gets large, where  $Q$  is the probability that a completely random error pattern will cause decoder error.

## References

- [1] R. J. McEliece and L. Swanson, "On the Decoder Error Probability for Reed-Solomon Codes," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 701-703, 1986.
- [2] K. Cheung, "More on the Decoder Error Probability of Reed-Solomon Codes," to appear in *IEEE Trans. Inform. Theory*.
- [3] R. Stanley, *Enumerative Combinatorics, Vol. I*, Monterey, California: Wadsworth and Brooks-Cole, 1986.
- [4] T. Kasami, T. Fujiwara, and S. Lin, "An Approximation to the Weight Distribution of Binary Linear Codes," *IEEE Trans. Inform. Theory*, vol. IT-31, no. 6, pp. 769-780, November 1985.
- [5] V. M. Sidel'nikov, "Weight Spectrum of Binary Bose-Chaudhuri-Hocquenghem Codes," *Probl. Peredachi Inform.*, vol. 7, no. 1, pp. 14-22, January-March, 1971.
- [6] K. Cheung, "The Weight Distribution and Randomness of Linear Codes," *TDA Progress Report 42-97*, vol. January-March 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 208-215, May 15, 1989.



53-32  
239 818

547657 p12  
N90-12794

TDA Progress Report 42-98

August 15, 1989

116.

# Automated Monitor and Control for Deep Space Network Subsystems

P. Smyth

Communications Systems Research Section

*This article will consider the problem of automating monitor and control loops for Deep Space Network (DSN) subsystems. The purpose of the article is to give an overview of currently available automation techniques. In particular, the article will consider the use of standard numerical models, knowledge-based systems, and neural networks. Among the conclusions is argued that none of these techniques alone possess sufficient generality to deal with the demands imposed by the DSN environment. However, the article will show that schemes that integrate the better aspects of each approach and are referenced to a formal system model show considerable promise—although such an integrated technology is not yet available for implementation. The article should be of benefit to readers interested in any aspects of DSN automation as it highlights the advantages and dangers of currently available approaches to the automation problem. The article frequently refers to the receiver subsystem since this work was largely motivated by experience in developing an automated monitor and control loop for the advanced receiver.*

## I. Introduction

Consider the problem of designing an automated monitor and control loop for a DSN subsystem. This article will refer to the "system" as the system being controlled and the "loop" as the hardware or software (or human) component which is used to monitor the system and control its behavior. For example, a receiver is a subsystem which is normally controlled in a relatively manual manner by human operators. This article will argue that reliance on purely manual control techniques will not be sufficient to cope with future subsystem technologies

and an increasingly complex DSN ground station environment. Autonomous or semiautonomous control loops will be necessary. This article will address some basic issues regarding automated monitor and control loops: how should such loops be designed and what general principles should be used? Specifically, the problem is defined in a very general manner using a state-space model and evaluating current technologies such as knowledge-based systems and neural networks in this context. The monitor and control problem can be decomposed into four basic subfunctions: obtain sensor data, estimate system parameters from the data, make decisions based on the estimated

**Table 1. Upper bound on R for the (31,15,17) RS code**

<i>u</i>	<i>R</i>
16	2.74943144e-024
17	1.50775270e-024
18	4.37734673e-025
19	8.94296586e-026
20	1.44241389e-026
21	1.95423782e-027
22	2.31146419e-028
23	2.44993596e-029
24	2.37090920e-030
25	2.12446863e-031
26	1.78181347e-032
27	1.41082139e-033
28	1.06190837e-034
29	7.64152968e-036
30	5.28215447e-037
31	3.52143591e-038

**Table 2. Weight distribution and its approximation for the number of decodable words of the (31,15,17) RS code over GF(32)**

<i>u</i>	<i>D<sub>u</sub></i> (exact)	<i>D'<sub>u</sub></i> (approximate)
9	1.998e+014	2.999e+015
10	5.232e+016	2.045e+017
11	6.348e+018	1.210e+019
12	4.817e+020	6.253e+020
13	2.613e+022	2.833e+022
14	1.113e+024	1.129e+024
15	3.970e+025	3.968e+025
16	1.231e+027	1.230e+027
17	3.364e+028	3.364e+028
18	8.111e+029	8.111e+029
19	1.721e+031	1.721e+031
20	3.200e+032	3.200e+032
21	5.196e+033	5.196e+033
22	7.322e+034	7.322e+034
23	8.822e+035	8.882e+035
24	9.178e+036	9.178e+036
25	7.967e+037	7.967e+037
26	5.699e+038	5.699e+038
27	3.272e+039	3.272e+039
28	1.449e+040	1.449e+040
29	4.647e+040	4.647e+040
30	9.603e+040	9.603e+040
31	9.603e+040	9.603e+040

parameters, and implement the decision by changing system inputs. Previously reported work on station automation [1, 2] has focused on the automation of operational procedures, namely the automation of data collection and control implementation, the first and fourth subfunctions listed above. Traditionally, the estimation and decision-making subfunctions have been performed manually by human operators. These are precisely the subfunctions on which this article focuses, i.e., this is what is referred to as automated monitor and control.

## II. A Statement of the Problem

A simple general model of a DSN subsystem which one wishes to control is shown in Fig. 1.  $I_o(t)$  and  $O_o(t)$  are defined as the observable inputs and outputs respectively, and  $I_{\bar{o}}(t)$  and  $O_{\bar{o}}(t)$  as the nonobservable inputs and outputs, where  $t$  denotes time.  $H$  is the open-loop transfer function of the system, i.e.,  $O(t) = H(I(t))$ .

In addition, there is possibly an a priori system model  $M$ . The rules of the game comprise of being given  $I_o(t)$ ,  $O_o(t)$ , and  $M$  at time  $t$ , and trying to effect a particular input/output transfer function  $H$  at some time  $t + \delta t$  by changing whatever subset of the inputs  $I_o(t)$  one is allowed to control. This is a very general description, and, for example, could equally well describe an implementation of an adaptive control loop using classical control theory, or a human trying to perform real-time fault-diagnosis in order to restore normal operating system behavior. A key point of this article is that there is a need to be concerned with all such levels of control. The need for modeling very different control loops of this kind in an integrated manner will be identified. The function of most DSN subsystems is to deliver or process signals or data subject to some criteria. Hence, the automated loop must be able to perform a variety of functions such as performance optimization and fault detection in order to satisfy overall system performance criteria. The article will return to this point in much greater detail later. This requirement will entail that the loop has a *model*, in some form, of the system it is trying to control.

Figure 2 shows the general feedback control scheme, where  $G$  represents the control-loop function. For example, if the system can be modeled accurately by closed-form analytic solutions, then  $G$  may be implemented as a set of equations such as a phase-locked loop (PLL) or other such receiver structure. However, at the level of interest here (namely, entire subsystems such as the receiver), analytic closed-form solutions are not practical and it is

traditional to close the loop manually, i.e.,  $G$  is a human operator.

For two reasons the usual engineering analysis approach will not work directly on this problem: first, the open-loop system transfer function (call it  $H$ ) is often non-linear and too complex to model accurately. Secondly, the transfer function may vary over time and be a function of the input variables. Remember that this is an entire subsystem involving the interaction of hardware and software components. Hence, the system will have a set of states (possibly infinite). Let the state of the system at time  $t$  be  $S(t)$ . The transfer function  $H$  is then  $H(S(t), I(t))$ . The only time when  $H$  is known exactly may be for a subset of states  $S_n$ , which might be thought of as "normal operating states." This is precisely the case for which little or no external feedback ( $G$  in Fig. 2) is required since internal feedback in the subsystem components (for example in the PLL component) will automatically produce the desired output. One is interested of course in the other cases ( $S(t) \notin S_n$ ), namely, spurious system behavior due to "non-normal" inputs, faulty components, etc.

As a concrete example, consider the receiver subsystem. Receivers are designed to work specifically on a known class of modulated signals at specific predicted frequencies. Hence,  $S_n$  corresponds to the case where the signal is at the correct position in the spectrum, of the correct modulation format, and the noise is bounded within expected characteristics. As is well-known, achieving optimal receiver performance even when all these assumptions are true is a very non-trivial design problem. Hence, system designers can hardly be expected to account for the large class of spurious system states that may arise during field operations, e.g., new noise characteristics, in-band radio-frequency interference (RFI), problems in other station subsystems, or system components which behave erroneously or fail. As previously mentioned, it has been traditional in field environments such as the DSN to have the human operator detect and correct spurious system behavior using his/her knowledge of the system and the environment. A point of this article is that manual control of this nature is neither feasible nor desirable in the long run.

There are two primary considerations: where to place  $G$  relative to existing subsystems and human operators and, secondly, how to implement  $G$ . This article will primarily focus on implementation issues, but will first briefly consider the issue of where to place an automated loop component relative to the existing DSN subsystem environments.

### III. Basic Architectural Considerations

Consider once more Fig. 2. There are a number of possible avenues one could pursue. One could close the loop within the box (i.e., internally in the system as shown in Fig. 3) and let the operators function as they currently do, relating inputs to outputs with little information as to the inner state of the system. This is feasible and reduces interface problems, but may actually introduce a new set of difficulties for the operator: the behavior of the internal control loop may make it very difficult for the operator to understand the behavior of the system, i.e., it is even more difficult to operate than previous "open-loop" systems.

An alternative approach is to replace the operator entirely and close the loop completely autonomously (Fig. 4). This approach is probably not necessary for DSN subsystems, since it is widely acknowledged that in any automated control system operating in uncertain environments, human intervention may always be necessary in emergency situations. A good example is that of autopilot control schemes in commercial airliners, where emergency situations require that the human pilots break the automatic control loop.

This paradigm of *semiautonomous* control (Fig. 5) introduces the important component of *situation assessment*. In airliners, for example, for the pilot to break the automated loop and effectively take over control, he must be able to assess the current state of the system or the context in which the automated control loop was acting. This issue requires that the automated system must in some sense be able to communicate at any time the current state of both the system and the control loop to a human operator. Hence, semiautonomous control entails that the control loop contains a high-level *model* of the system it is trying to control.

Another important practical point is that in order for the control loop  $G$  to be effective it must have access to all of the relevant system data and be able to control system parameters, i.e., a comprehensive sensory and effector system must be in place. This is a fundamental requirement. Subsystems should be designed initially such that future hardware and software links to external automated control loops can be easily and cleanly implemented, e.g., hardware ports, software accessibility, etc. This holds true even if the automated loop is built into the originally designed system, since at the next level of automation these control loops will need to communicate with higher-level systems such as a monitor and control system for an entire ground station.

### IV. A State-Space Model for Automated Monitor and Control

Consider in more detail the nature of the control loop. One can decompose the problem into two parts: determining the state of the system at time  $t$  (*estimation*) and effecting some control action on the system to achieve desired behavior at time  $t + \delta t$  (*decision*). Monitoring the state of the system is a necessary prerequisite to controlling it, since controller actions need to be context dependent, i.e., what is done to control the system depends on the state of the system and the input variables (both observables and unobservables). This is the crux of the problem.

As described earlier one can define  $S(t)$  as the state of the system at any time  $t$ , and  $S_n$  is a set of states defined as "normal operating behavior." Hence one can define the set of states  $S_{\bar{n}}$  as system states indicating spurious behavior and potentially requiring corrective control action, where  $S_{\bar{n}}$  is the complement of  $S_n$  over the entire set of possible system states. The purpose of the monitor/controller  $G$  can be restated as follows:

- (1) At any given time  $t$  obtain a state estimate  $\hat{S}$  of the true state  $S(t)$  of the system, where

$$\hat{S} = f(I_o(t), O_o(t), M, h(\hat{S}, I_o, O_o, t)) \quad (1)$$

where  $I_o(t)$  and  $O_o(t)$  are the observable inputs and outputs,  $M$  is an a priori model of the system, and  $h(\cdot)$  is a finite history of previous estimated states and input and output variables.

- (2) If  $\hat{S} \in S_{\bar{n}}$  choose a control action to achieve a specific set of outputs at time  $t + \delta t$  in the future, by setting

$$I_o(t + \delta t) = g(\hat{S}(t), M, I_o(t), h(\hat{S}, I_o, O_o, t)) \quad (2)$$

While this state estimation and transformation model is not the only possible model, and is somewhat of a simplification, it serves the purpose of giving a good insight as to the basic nature of the problem. The first step, that of estimating the state of the system, is quite challenging in itself since the state-space model will necessarily be quite complex for any realistic system. This step must be solved before proceeding to step (2), since typically control actions are quite sensitive to context information, i.e., the state estimate  $\hat{S}$ . For example, an operator's actions on

a receiver depend directly on what the operator perceives the current state of the receiver to be, e.g., whether or not the PLL is in lock or not.

The state-space model is useful because of its generality. In this framework a phase-locked loop is a very well-specified state estimation and control scheme operating on a signal in real-time. At the other end of the spectrum, the human operator is implementing a much more heuristic and adaptive estimation and control scheme at a much slower rate, what one might term "human real time." One is interested in control loops that operate between these two extremes. The remainder of the article will examine current technologies in the context of this state estimation and control model. The possible approaches can be divided into three broad paradigms (standard numeric algorithms, knowledge-based systems, and neural networks) not because they form a well-defined taxonomy for modeling autonomous control systems, but simply because of the widespread interest in techniques of this nature and their potential applications. A goal of the article is to clarify to which types of problems each approach is best suited, and in the process debunk some common myths.

## V. Standard Analysis and Algorithms

Engineers like to model systems using differential or difference equations and then, by analysis of the mathematical model, design an estimator and controller subject to cost and performance criteria. The quality of the resulting control loop hinges critically on the accuracy of the assumed model, e.g., whether the real noise statistics fit the parametric model or not. By weakening the assumptions (e.g., by using non-parametric statistical models) one can improve the generality of the model but it then becomes correspondingly more difficult to design the control loop.

Some systems are simply impossible to model accurately using closed-form mathematical solutions alone. It is a key point of this article that DSN subsystems fall into this category. The advanced receiver subsystem, for example, will be a complex mixture of interacting hardware and software components. Clearly, at the local *component* level, existing analytical models will be quite accurate, e.g., the transfer function for an accumulator or an automatic gain control (AGC) circuit. However, one is interested in the global picture, how the components interact together to determine system behavior. Direct modeling via equations is hardly practical. For instance, how does one combine information such as " $E_s/N_0 = 9.5$  dB" and "the board for calculating FFTs is down"?

The key point to note is that variables of interest must be abstracted to the appropriate level of representation. The issue of representation is critical. The model should necessarily entail different levels of abstraction, from relatively high-level system configuration concepts (such as a board being down) down to quantitative real-time parameter estimates (such as the value of  $E_s/N_0$ ). One can conclude that exact mathematical models have an important role in control loops but are not sufficiently powerful in terms of representation to meet requirements.

## VI. Knowledge-Based Systems

The term "knowledge-based system" is used rather loosely to describe any system where knowledge about the system is explicitly represented in symbolic form, as opposed to being embedded in a low-level algorithm. There is an increasing realization that any system which intends to communicate with other agents, be they other pieces of software or humans, needs to be able to abstract its information to the appropriate level. In addition, explicit knowledge representation buys the ability to modify the control system, to update it, etc., in an efficient manner. The potential value of portable, easily-modifiable code that can be changed relatively effortlessly is enormous.

### A. Rule-Based Expert Systems

Rule-based expert systems have proven to be very popular in recent years as solutions to the type of control problem discussed so far. Essentially, they attempt to replace or augment the human operator by a program which contains a representation of his expertise as a set of rules. The question will be "Is this a good modeling tool for solving the DSN automated monitor and control problem?"

First consider the basic characteristics of this approach. The rules represent situation-action pairs or symptom-diagnosis pairs based on the expert's experience. Generally this approach works well in domains where experts do well and the domain is not particularly well-understood at a basic theoretical level, e.g., in many medical applications where basic cause and effect relationships are not known. Conversely, a rule-based system degrades rapidly in performance when faced with novel problems (symptom combinations not encoded as rules) which the expert has never encountered before. In artificial intelligence, this type of rule-based representation, based on experiential knowledge alone, is referred to as a "shallow" model, since it is based on simple pattern-matching

without reference to any deeper causal domain theory. It is important to point out that impressive results have been achieved in real applications using rule-based systems and that the technology is available for implementation at present and is well-understood. What this article is trying to do is to identify the limitations of the approach in relation to a particular monitor and control problem.

With reference to Section IV, which defined the monitor and control problem as a two-step procedure based on the estimation function  $f(\ )$  and decision function  $g(\ )$ , a rule-based implementation can be considered as follows: effectively the functionality of  $f(\ )$  and  $g(\ )$  are combined together into single condition-action pairs or rules. Hence the rules define a heuristic mapping from system parameters to control actions. In this sense the state of the system is never formally evaluated. The consequent implications for semiautonomous control and situation assessment are obvious, i.e., it will be very difficult for an external agent to take over system control in an effective manner.

Rule-based systems are not particularly appropriate in domains where any of the following hold true:

- (1) The experts have limited experiential knowledge (are not really experts) or, indeed, there are no experts. This latter problem arose in the advanced receiver expert system project since the system has not been fielded in an operations environment.
- (2) There exists a well-defined domain theory, at least at the component level, i.e., well-defined functional and structural properties in the system exist which are fundamental to problem diagnosis.
- (3) The external environment is subject to change and, hence, novel problems are likely.
- (4) Time and procedural information are critical to reasoning about the domain, e.g., in setting up a link to a spacecraft there is a definite sequence of procedural steps which occur. Rules are very inefficient at representing procedural knowledge.

System designers are *not* the types of experts that rule-based expert systems are intended to model. Their knowledge consists of causal reasoning based on first principles rather than the experiential situation-action pairs of an experienced system operator. While it is true that one can recast basic domain knowledge in the form of rules, it tends to obscure the underlying causal domain theory, i.e., it is not an efficient way to represent this type of knowledge. For a new system like the advanced receiver there

is no experiential or heuristic knowledge available. This problem of designing "expert" systems without experts is quite challenging—this article will show later how the notions of model-based reasoning and learning algorithms may provide useful solutions. A practical approach might be to initially field the system with some form of intelligent model-based control loop which includes a capability for rule-based representation, and, as expertise develops, integrate experiential rules into the loop.

Given the above observations, it is clear that rule-based expert systems are not a very good overall model for the problem discussed here. It is important to keep in mind that rule-based systems were originally developed primarily by the medical community as a *consultative* off-line tool, i.e., a program with which a human interacted, answered queries, and received recommendations. Expert systems were never explicitly designed to operate in active, real-time modes in domains where expertise is scarce and reasoning from first principles is essential.

Nonetheless it is important to counter the common myth that rule-based models require dedicated hardware and software components to work and operate separately from "ordinary" software on special purpose machines. Recent trends in the application of rule-based systems have been very much focused on integrating the rule-based approach with standard software, e.g., rule bases can be called by standard procedural programs, integrated with common databases, etc. At the present time most commercial shells are available in languages such as C. Hence, the paradigm of a standard procedural control loop that can call a function to perform rule-based reasoning (to determine the value of a global context variable, perhaps), which is written entirely in a relatively low-level language such as C, and which can be executed by a single-chip microprocessor may be a much more useful paradigm than relying on the conventional rule-based approach as the basic model for control loops.

## B. Model-Based Reasoning

Researchers in artificial intelligence have recognized the limitations of purely rule-based systems as outlined above and have consequently moved on to investigating more powerful knowledge representation techniques. Atkinson [3] gives a useful overview of recent AI applications for monitor and control in the aerospace domain. This is of interest since, as with the DSN subsystem problem, monitor and control of aerospace systems involves a large amount of temporal, procedural, and first-principles knowledge. The notion of *model-based reasoning* is loosely based on the idea that if a reasoning system knows the

function and structure of each system component (at some appropriate level of representation), and knows how these components are connected together, then by observing certain input and output system parameters it should be able to reason from first-principles to infer system behavior, e.g., detect the most likely cause of a fault. This is an attractive paradigm for the domain discussed here. Research in this area has primarily been domain-dependent, with perhaps the most common domain being that of troubleshooting electrical circuits [4]. A particularly notable application (in a different domain) is the LES system [5] which monitors and controls the critical loading of liquid oxygen onto the space shuttle at Kennedy Space Center prior to launch—the system has been able to perform fault detection and isolation in 10 sec as compared to the 20 min taken by manual methods.

Model-based reasoning is quite different from the rule-based expert system approach and has been associated with the term “deep knowledge” or common-sense reasoning. Unfortunately, from an applications point of view, there are many unsolved problems which remain the subject of ongoing research. For example, the computational requirements can scale exponentially with the number of components under consideration. Another problem is that there has been very little applied research in this area and hence there is very little practical feedback available for the purposes of determining the appropriate level of granularity for such models. For example, for a signal-processing subsystem such as a receiver, should the model entail first-principles knowledge regarding linear systems theory, or is this too detailed? A useful overview on the current state of research in applying model-based reasoning to communication systems is given in [6].

However, despite the unresolved issues, it is important to realize that a priori knowledge of DSN subsystems often falls into this model-based category, i.e., well-defined component models of function and structure. With the advanced receiver, for example, this is very much the case, but with the scarcity of research results and applications in this area (none at all in the communication systems domain to the author’s knowledge) it was not possible to apply these ideas directly to the problem at hand within the resources available. One can conclude that model-based reasoning appears promising as a useful modeling tool for the DSN domain and worthy of further attention.

### C. Artificial Intelligence and Decision Theory

There has recently been a considerable amount of cross-disciplinary work relating AI with decision analysis and decision theory. Decision theory and analysis grew

out of statistical decision theory and has developed a considerable basis of theoretical principles and practical techniques for modeling choice and decision in uncertain environments. Applications of this technique have typically been for decisions in economic and military environments where decision theory can render decision-making a formal statistical modeling problem rather than a subjective procedure, e.g., an oil company may wish to decide whether or not to commit resources to drilling an oil well in a particular location. More recently, this work has been applied to problems which are typically within the AI domain, e.g., fault diagnosis or reasoning with uncertainty. The key shift in focus has been the notion of using decision theory as a theoretical model for decision by an autonomous agent, rather than using it to analyze specific human decision scenarios. In a sense, one might say that this involves decision synthesis rather than decision analysis.

Whether or not decision theory can be applied to the DSN monitor and control problem remains an open question. As with model-based reasoning, research on this topic is still in relative infancy. Theoretically, since the monitor and control loop discussed in this article involves a decision element, there can be no question that decision theory is an appropriate formal model for an autonomous decision agent. There are two key benefits that could be gained from using this approach:

- (1) Decision analysis emphasizes and provides many tools for the initial structuring and modeling of the problem at hand, e.g., the use of influence diagrams, a technique for structuring and identifying causal models. Using these techniques would formalize the initial model-building part of the control loop design process.
- (2) The use of statistical decision theory as the basis for a rational decision agent is particularly appealing. For example, in an advanced receiver it may be possible to combine several fast Fourier transform (FFT) power spectra in a noncoherent manner. Deciding how many spectra to average over involves a tradeoff between the variance of the spectral estimate and the time taken to perform the estimate, i.e., the time delay in the control loop. Clearly this decision is very much *context*-dependent on parameters such as carrier-to-noise ratio and estimated carrier Doppler rate. For example, it is more important to minimize the time-delay when one is not in lock than when one is in lock. Utility is defined as a quantitative measure of the overall benefit or cost which accrues to the decision-maker from a particular action/outcome pair, i.e., if he/she decides on an action  $a_i$ , and event  $e_j$  subsequently occurs,  $u(a_i, e_j)$  is a defined utility for this pair. For the

purposes here, utility might be defined as the variance/time tradeoff associated with each action/event pair. Rational decision-making corresponds to choosing the action  $a_k$  which maximizes one's expected utility with respect to the probability space defined over relevant events, i.e., a probability space involving context-valued variables such as in-lock/not-in-lock.

However, as with standard algorithms and analysis, decision theoretic models alone are not sufficiently rich as a knowledge representation scheme to offer a stand-alone solution. Nonetheless, there is a strong case for their inclusion as a solution component and their integration with more conventional approaches.

## VII. Neural Networks

What is a neural network? One view (from an algorithm designer's perspective) is that they are highly distributed, nonlinear models for implementing function estimation, i.e., given input and output *samples* of a "black box" the neural network can reproduce the functionality of the "black box" *in general*. This view may be somewhat narrow, but generally speaking, optimization, adaptation, pattern recognition, etc., can all be suitably recast into the notion of function estimation. Hence, for example, neural networks can form the basis for computationally powerful, nonparametric statistical estimation tools. Indeed, it seems that early successful applications of this technology may be found in computationally intensive pattern recognition and signal processing problems which have proven to be difficult to model via traditional serial, linear, non-adaptive models, e.g., speech recognition, optical character recognition, nonlinear signal processing, adaptive control, etc. Typically, it seems that the less one knows about the problem solution a priori (e.g., the less one understands about the noise characteristics of the problem), the more favorably a neural solution will compare to standard alternatives such as Markov models, Bayesian classifiers, etc. The availability of dedicated hardware within the next decade will be a major factor in determining the future of neural network technology.

How do neural networks relate to the monitor and control problem in this article? One could certainly envision possible applications in terms of real-time estimation and adaptive control, particularly at the *signal* level. But as with each of the approaches we have considered earlier, a neural network would not be sufficient as an overall system model. In particular, given the current understanding of the neural approach, there is no way to incorporate a priori

knowledge into the network. In terms of the earlier two-step model, the decision function  $g(\ )$  is learned entirely by the network itself, with no reference to the a priori system model  $M$ . Neural networks are essentially "black boxes" in that they learn and implement input/output mappings efficiently—but their representation of this mapping is implicitly embedded in a distributed manner in terms of learned weights. Hence, for example, one cannot incorporate existing functional and structural knowledge about the subsystem discussed here into the network. One can conclude however that as a subcomponent of the overall model, say as a nonparametric signal/RFI classifier, the neural network approach shows significant promise provided the hardware becomes available.

## VIII. Summary

The need for automating monitor and control functionality in DSN subsystems is becoming more apparent as technology (sensors and computational resources) improves to the extent that operators can no longer be expected to assimilate system data and make decisions in real time. This article has shown that in whatever manner one may implement such a monitor and control loop, it clearly needs to include a *model* of the system, whether it is implicitly embedded in equations or rules, or more formally represented in an explicit manner. The argument for a formal representation model revolves around three facts: first, it renders the loop design problem more tractable as it provides the reference point from which to integrate different numeric and symbolic approaches. Secondly, it increases both the life cycle and the flexibility of the designed software (making it easier to transfer prototype systems of this nature from a research laboratory to field implementation). Thirdly, it facilitates communication between the subsystem and higher authorities, be they human or automated.

It is apparent that the currently available paradigms for implementing control loops each have distinct advantages, yet, for the purposes described in this article, none are sufficiently powerful as an overall system model. It may be conjectured from these observations that successful monitor and control systems in the DSN domain will incorporate models which integrate the more appropriate aspects of each approach. One could envision a model-based representation that is run using decision-theoretic principles and that controls and calls lower-level analytic models and neural-based estimation schemes in a context-dependent manner. For example, in the receiver, computational resources could initially be focused on acquisition



(using an FFT to locate peaks in the spectrum [7]), while, once lock is achieved, resources could be allocated to RFI monitoring and parameter optimization.

In the short term, the issue of model-based techniques and their integration with standard numerical models appears worthy of further investigation and application. For example, the use of object-oriented programming techniques appears to have considerable potential for implementing models of this nature. Rule-based expert systems will have their role to play, but the system implementer is advised to be wary if the domain primarily involves temporal, procedural, or first-principles knowledge. Using rule-based systems as a *component* of an overall solution seems like a much more reasonable approach, particularly given that (as discussed earlier) rule-based code can now be written in "low-level" languages such as C and consequently embedded in standard software. Neural network technology may be applicable to specific subcomponents of the estimation and control problem, particularly once dedicated hardware becomes available. Hybrid systems, which combine aspects of each approach appear potentially to be the most promising avenue, but little research and/or applications has been carried out in this arena, and hence they appear to be more of a long-term prospect.

A final point worth noting concerns that of learning algorithms, i.e., systems which can improve their performance over time. A little thought will convince the reader that, in this domain, learning may proceed much more ef-

ficiently and effectively if it is referenced to an initial system model. A very general viewpoint is that in modeling any system one can introduce an a priori bias (e.g., a set of assumptions) into the model, as in Bayesian statistics. A strong bias will only help the model if it is accurate, whereas very weak biases will make the learning problem much more difficult. A case in point for automated monitor and control of man-made systems is that the a priori bias, in the form of a functional and structural component model  $M$ , is necessarily correct and accurate. Hence, in principle, it can only improve the quality of the overall control loop and any learning component therein.

## IX. Conclusion

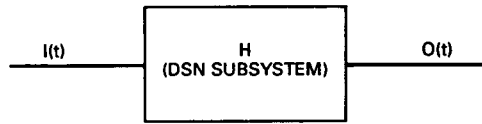
Many real-time monitor and control applications will not be sufficiently well-modeled by rule-bases, neural networks, or standard algorithms *on their own*. This article reviewed the current state-of-the-art in this area and saw that more powerful representational models are in the offing. One concludes that an effective modeling technique would be one which integrated the better aspects of each approach by referencing them to an explicit functional and structural model of the system. Off-the-shelf solutions of this nature are not currently available. Hence, until such techniques are available, system developers should choose their monitor and control models carefully and recognize the limitations of each particular technology.

## Acknowledgment

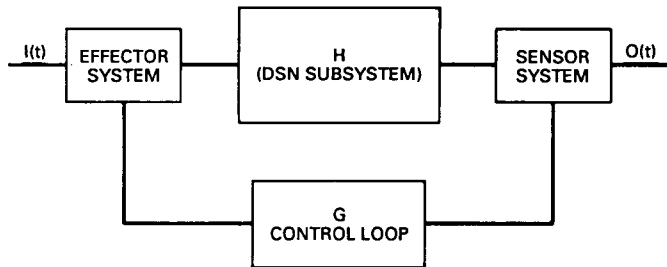
The author gratefully acknowledges the contributions of Gordon Bagby, Joseph Statman, Sami Hinedi, and William Hurd to this work.

## References

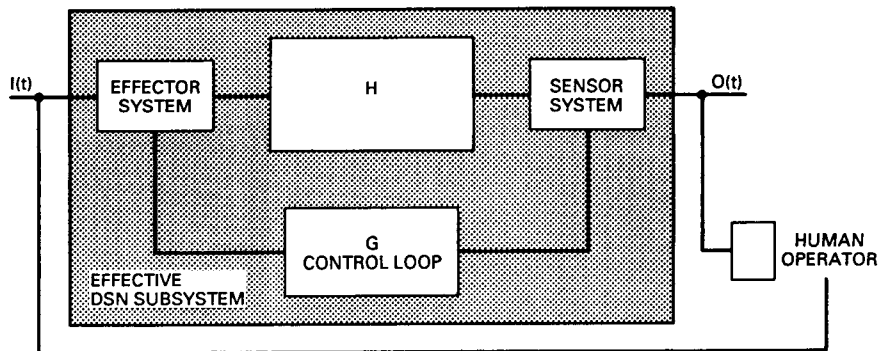
- [1] D. S. Remer and G. Lorden, "Deep Space Station (DSS-13) Automation Demonstration," *TDA Progress Report 42-57*, Jet Propulsion Laboratory, Pasadena, California, pp. 103-119, June 15, 1980.
- [2] C. Foster, "Unattended Deep Space Station Tracking Station Development: Monitor and Control Technology," *TDA Progress Report 42-86*, vol. April-June 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 164-170, August 15, 1986.
- [3] D. J. Atkinson, *Knowledge-Based Diagnosis for Aerospace Systems*, JPL Publication 88-7, Jet Propulsion Laboratory, Pasadena, California, March 15, 1988.
- [4] R. Davis, "Diagnostic Reasoning Based on Structure and Behaviour," *Artificial Intelligence*, vol. 24, nos. 1-3, pp. 347-410, December 1984.
- [5] E. Scarl, et al., "A Fault Detection and Isolation Method Applied to Liquid Oxygen Loading for the Space Shuttle," *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, Los Altos, California: Kaufmann Publishers, pp. 18-23, 1985.
- [6] R. O. Yudkin, "On Testing Communication Networks," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 5, pp. 805-812, June 1988.
- [7] D. H. Brown and W. J. Hurd, "DSN Advanced Receiver: Breadboard Description and Test Results," *TDA Progress Report 42-89*, vol. January-March 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 48-66, May 15, 1987.



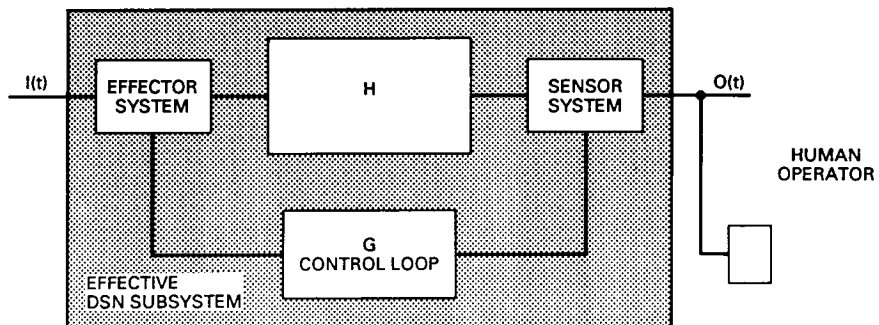
**Fig. 1. Simple system model.**



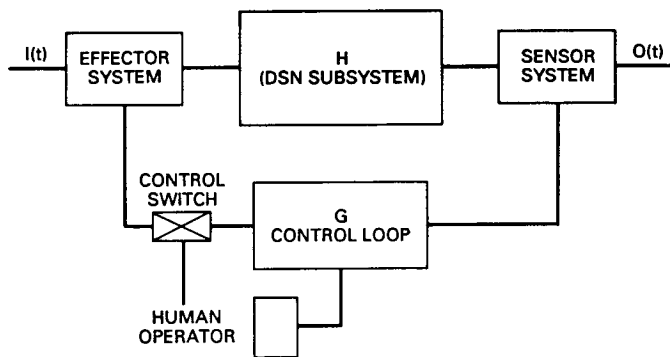
**Fig. 2. System model with feedback control.**



**Fig. 3. Internal control loop with external manual control.**



**Fig. 4. Autonomous control loop.**



**Fig. 5. Semiautonomous control loop.**

# Frame Synchronization Methods Based on Channel Symbol Measurements

S. Dolinar and K.-M. Cheung  
Communications Systems Research Section

*The current DSN frame synchronization procedure is based on monitoring the decoded bit stream for the appearance of a sync marker sequence that is transmitted once every data frame. This article explores the possibility of obtaining frame synchronization by processing the raw received channel symbols rather than the decoded bits. Performance results are derived for three channel symbol sync methods, and these are compared with results for decoded bit sync methods reported elsewhere. It is shown that each class of methods has advantages or disadvantages under different assumptions on the frame length, the global acquisition strategy, and the desired measure of acquisition timeliness.*

*It is shown that the sync statistics based on decoded bits are superior to the statistics based on channel symbols, if the desired operating region utilizes a probability of miss many orders of magnitude higher than the probability of false alarm. This operating point is applicable for very large frame lengths and minimal frame-to-frame verification strategy. On the other hand, the statistics based on channel symbols are superior if the desired operating point has a miss probability only a few orders of magnitude greater than the false alarm probability. This happens for small frames or when frame-to-frame verifications are required. Among the three channel symbol methods examined, the squared-distance statistic offers the best performance in the range of normal signal-to-noise ratios, but it degrades more rapidly than the correlation statistic or the hard-limited symbol discrepancy count statistic when the signal-to-noise ratio is extremely low.*

## I. Introduction

The current DSN frame synchronization procedure is based on transmitting from the spacecraft a convolutionally encoded 32-bit sync marker sequence at the beginning of every data frame consisting of around  $10^4$  bits. On the ground, the decoded bit stream is monitored for a 32-bit window of agreements with the marker, and a likely sync location is identified by comparing the number of disagreements with a preselected threshold. In addition, a frame-to-frame verification strategy is employed to definitively declare sync acquisition or sync loss.

This article explores the possibility of obtaining frame synchronization by processing the raw received channel symbols rather than the decoded bits. Performance results are derived for three channel symbol sync methods, and these are compared with results for decoded bit sync methods reported elsewhere [1, 2]. It is shown that each class of methods has advantages or disadvantages under different assumption on the frame length, the global acquisition strategy, and the desired measure of acquisition timeliness.

## II. Various Possible Synchronization Methods

### A. General Framework for a Wide Class of Synchronization Methods

The DSN's current method of frame synchronization can be viewed as a prototype of a fairly general type of synchronization scheme. There are three basic levels to this scheme. First, a statistic  $x$  is measured at all possible locations of the frame marker, and the observed values of  $x$  are used to identify likely locations of the marker. Second, values of  $x$  obtained at all possible locations within a frame and for different frames are correlated for one or more frames to verify and select the most likely marker location. Third, the statistic  $x$  is monitored from frame to frame at the selected marker location to verify continued sync.

This three-level synchronization model can be completely general if the decisions at all three levels can depend in an arbitrary way on all the observed data. In practice, it is usually assumed for ease of implementation that the three levels of decisions are mostly decoupled. This article

considers only the first two levels, with primary emphasis on the first level.

Ideally, the statistic  $x$  should be a sufficient statistic summarizing all the relevant information in the data. In practice,  $x$  is constrained to be a meaningful, but easy to compute, measure of the data. There are many possible choices for the statistic  $x$ . This statistic may be based directly on observation of the received channel symbols, or it may be calculated from the decoded output of the Viterbi decoder. This article analyzes several different statistics based on channel symbol measurements, and briefly discusses some other statistics based on decoded bits; these statistics are reported elsewhere [1, 2].

Figure 1 is a general system diagram introducing notation for describing the two types of synchronization statistics. The stream of incoming "data" bits  $\{b_i\}$  includes both true data bits and sync marker bits  $\{\lambda_i\}$ . The "data" bit stream is packaged into data frames  $\{b_i, i = 1, \dots, B\}$  of  $B$  bits each, and  $L$  sync marker bits  $\{\lambda_i, i = 1, \dots, L\}$  are included in every data frame. The "data" bit stream is convolutionally encoded by a rate  $1/N$ , constraint length  $K$  convolutional encoder. The encoded channel symbol stream  $\{s_i, i = 1, \dots, S\}$  is likewise partitioned into frames of  $S = NB$  symbols each, and each frame includes a set of  $M = N(L - K + 1)$  sync marker symbols  $\{m_i, i = 1, \dots, M\}$  that are totally determined by the sync marker bits  $\{\lambda_i, i = 1, \dots, L\}$ . The remaining  $N(B - L + K - 1)$  symbols in each frame are dependent solely on the true data bits or else on a combination of true data bits and sync marker bits.

The channel symbols are assumed to have constant magnitude  $s$  (i.e.,  $s_i = \pm s$ ), and they are received in additive white Gaussian noise  $\{n_i, i = 1, \dots, S\}$  with zero mean and variance  $\sigma^2$ . The ratio  $\rho = s^2/\sigma^2$  is a signal-to-noise parameter. In terms of  $\rho$ , the channel symbol signal-to-noise ratio is  $E_s/N_0 = \rho/2$ , and the bit-energy-to-noise ratio is  $E_b/N_0 = N\rho/2$ . The received symbols  $\{r_i, i = 1, \dots, S\}$  are passed through a maximum likelihood convolutional decoder (Viterbi decoder) to obtain the decoded bits  $\{d_i, i = 1, \dots, B\}$ .

The general performance expressions in this article are derived for arbitrary combinations of the parameters  $K$ ,  $N$ ,  $L$ ,  $M$ ,  $B$ , and  $S$ . However, all explicit performance curves assume a 32-bit marker sequence and the NASA-standard constraint length 7, rate 1/2 code (i.e.,  $L = 32$ ,

$M = 52$ ,  $K = 7$ , and  $N = 2$ ). The curves in Section III are applicable to any frame length, while the curves in Section IV assume a frame length of 10000 bits or 20000 symbols (i.e.,  $B = 10000$ ,  $S = 20000$ ).

## B. Sync Methods Based on Decoded Bits

Frame sync observables based on decoded bits are a function of the  $L$ -vector of discrepancies  $\{\lambda_i \oplus d_i, i = 1, \dots, L\}$  between the sync marker bits  $\{\lambda_i, i = 1, \dots, L\}$  and a sliding  $L$ -bit window of actual decoded bits  $\{d_i, i = 1, \dots, L\}$ . Several reasonable statistics are described in this section. Further analysis of observables based on decoded bits is not carried out in this article, but performance curves are cited from [1, 2] for comparison with channel symbol methods.

### 1. Decoded Bit Discrepancy Count Statistic.

In this case, the statistic  $x$  simply counts the number of discrepancies between the sync marker bits and the current  $L$ -bit window:

$$x = \sum_{i=1}^L [\lambda_i \oplus d_i]$$

This is the statistic used in the current DSN frame sync algorithm.

### 2. Decoded Bit Discrepancy Span Statistic.

Viterbi decoders tend to make bursts of errors and the decoded bits are essentially random inside an error burst [3]. Thus, any agreements with the sync marker that occur inside a decoder error burst are completely accidental and should not be counted as true agreements. This observation leads to the definition of the discrepancy span statistic. The discrepancy span measures the distance in bits between the first and last disagreements with the sync marker. The statistic  $x$  is defined simply by

$$x = i_2 - i_1 + 1$$

where

$$i_1 = \min_{1 \leq i \leq M} \{i : (\lambda_i \oplus d_i) = 1\}$$

$$i_2 = \max_{1 \leq i \leq M} \{i : (\lambda_i \oplus d_i) = 1\}$$

This statistic was first suggested and analyzed in [2].

## C. Sync Methods Based on Channel Symbols

Frame sync observables based on channel symbols are obtained by comparing a sliding  $M$ -symbol window of received symbols  $\{r_i, i = 1, \dots, M\}$  with the sync marker symbols  $\{m_i, i = 1, \dots, M\}$ . Several reasonable channel symbol statistics are described in this section.

**1. Hard-Limited Channel Symbol Discrepancy Count.** The channel symbol statistic most similar to the decoded bit statistics is obtained by hard limiting each received symbol  $r_i$  to the nearest transmitted symbol value  $\pm s$  and counting disagreements with the sync marker sequence. A statistic  $x$  that simply counts the number of discrepant symbols is defined by

$$x = \frac{1}{2} \sum_{i=1}^M [1 - \text{sgn}(m_i) \text{sgn}(r_i)]$$

This statistic  $x$  counts one discrepancy for every disagreement in sign between a received symbol  $r_i$  and the corresponding sync marker symbol  $m_i$ .

Because the received symbols  $\{r_i\}$  contain white Gaussian noise and are not prone to error bursts, burst-inspired statistics such as the discrepancy span statistic are not useful in the channel symbol domain. However, statistics that make use of the soft-quantized information in the channel symbols can offer improvement.

**2. Weighted Symbol Discrepancies.** An interesting statistic using soft-quantized information is derived as a weighted sum of symbol discrepancies. The statistic  $x$  is defined by

$$x = \sum_{i=1}^M \max[0, -r_i \text{sgn}(m_i)]$$

Just like the hard-limited symbol discrepancy count statistic, this statistic  $x$  counts 0 every time the received symbol  $r_i$  and the corresponding sync marker symbol  $m_i$  agree in sign, no matter how inexact the agreement might be. However, this statistic does not count all sign discrepancies equally, but instead weights them by the magnitude of the corresponding received symbols.

The weighted symbol discrepancy statistic was first proposed by Massey [4] as a good approximation to the maximum likelihood detection statistic in the region of high signal-to-noise ratio. This statistic is not further analyzed in this article, but will be examined in the future by simulation or central limit theorem approximation.

**3. Channel Symbol Correlation Statistic.** One natural soft-quantized statistic measures the correlation between the received symbols and the sync marker symbols. The statistic  $x$  is defined by

$$x = \sum_{i=1}^M m_i(m_i - r_i)$$

The observed value of this statistic should be near 0 if  $\{r_i, i = 1, \dots, M\}$  contains the marker, and otherwise should be a large positive number around  $M s^2$ . With this definition, the correlation statistic exhibits the same qualitative behavior as the discrepancy-based statistics.

**4. Channel Symbol Squared Distance Statistic.** Another natural symbol-based statistic measures the squared distance between the received symbols and the sync marker symbols. The statistic  $x$  is defined by

$$x = \sum_{i=1}^M (m_i - r_i)^2$$

Again, a small value of  $x$  indicates agreement with the marker, while a large value indicates disagreement.

### III. Probability of Miss and Probability of False Alarm

#### A. Basic Definitions

All of the statistics  $x$  defined in the previous section are scalar; a value of  $x$  near 0 indicates a match with the sync marker, while a large positive value indicates disagreement. Thus, it is natural to use the observed values of  $x$  to make tentative yes-no decisions about the location of the marker, according to whether  $x$  falls below or exceeds

a threshold  $\theta$ . The tentative decision rule for all statistics defined above is of the general form

$$x \begin{array}{l} \leq \theta \\ > \theta \end{array} \begin{array}{l} \text{sync} \\ \text{no sync} \end{array}$$

If  $x$  exceeds  $\theta$  when  $x$  is measured at the true position of the marker, the tentative decision rule results in the true marker location being missed in the current frame. Conversely, if  $x$  falls below  $\theta$  when  $x$  is not measured at the true marker position, then the decision rule causes a false detection of sync or false alarm.

The intrinsic goodness of various statistics  $x$  can be evaluated by comparing the tradeoff between probability of miss ( $P_M$ ) and probability of false alarm ( $P_F$ ), which results as the threshold  $\theta$  is varied for a given statistic  $x$ . The miss probability and false alarm probability are defined as

$$P_M = \text{Prob} [x > \theta | \text{sync marker located at current } L\text{-bit or } M\text{-symbol window}]$$

$$P_F = \text{Prob} [x \leq \theta | \text{sync marker not located at current } L\text{-bit or } M\text{-symbol window}]$$

The miss probability  $P_M$  is the probability that the true marker will fail to pass the threshold test at the true location of the marker. Within any one data frame, there is only one opportunity for a miss to occur. The false alarm probability  $P_F$  is the probability that the threshold test will be passed at some particular location that contains only true data bits and symbols or possibly a combination of true data bits and symbols and marker bits and symbols. In each frame there are  $B - 1$  opportunities for false alarm for algorithms based on decoded bits, or  $S - 1$  opportunities for algorithms based on channel symbols.

#### B. Computation of $P_M$ versus $P_F$ for Various Observables

**1. Decoded Bit Discrepancy Count Statistic.** The tradeoff between miss probability and false alarm probability for the decoded bit discrepancy count statis-



tic has been reported in [1, 2]. A graph showing some of the results of [2] is presented in Fig. 2.

It must be noted that the miss probability and false alarm probability defined in [2] are not equal to those used in this article. In [2], the threshold test had to be passed at the same location in two consecutive frames for either a false alarm or detection of true sync to occur. The false alarm probability ( $O|M'$ ) and miss probability  $1-P(O|M)$  defined in [2] are related to the corresponding probabilities  $P_F$  and  $P_M$  used in this article as  $P(O|M') = P_F^2$ ,  $1-P(O|M) = 1-(1-P_M)^2$ .

## 2. Decoded Bit Discrepancy Span Statistic.

The miss probability versus false alarm probability curves for the decoded bit discrepancy span statistic are shown in Fig. 3. These curves are taken from [2], after adjusting for the different definitions of miss probability and false alarm probability described above.

**3. Hard-Limited Channel Symbol Discrepancy Count Statistic.** Measured at the marker, the hard-limited symbol discrepancy count statistic  $x$  is a binomial random variable,

$$x = \sum_{i=1}^M \frac{1}{2} [1 - \text{sgn}(m_i) \text{sgn}(m_i + n_i)] = \sum_{i=1}^M x_i$$

where  $\{x_i, i = 1, \dots, M\}$  are independent binary random variables,

$$\text{Prob}[x_i = 1] = q = 1 - \text{Prob}[x_i = 0]$$

with

$$q = \text{Prob}[n_i > m] = Q\left(\frac{m}{\sigma}\right) = Q(\sqrt{\rho})$$

and

$$Q(u) = \frac{1}{\sqrt{2\pi}} \int_u^{\infty} e^{-t^2/2} dt$$

The probability distribution for the statistic  $x$  is given by

$$\text{Prob}[x = j] = \binom{M}{j} q^j (1-q)^{M-j}, \quad j = 0, \dots, M$$

The miss probability is easily evaluated as

$$P_M = \sum_{j=\theta+1}^M \binom{M}{j} q^j (1-q)^{M-j}$$

At any given location not overlapping the marker, the statistic  $x$  is conditionally a sum of two binomial random variables whose statistics depend on the discrepancies  $\{\delta_i, i = 1, \dots, M\}$  between the marker symbols  $\{m_i, i = 1, \dots, M\}$  and the sliding window of  $M$  symbols  $\{s_i, i = 1, \dots, M\}$  at the current location

$$\delta_i = \frac{1}{2} [1 - \text{sgn}(m_i) \text{sgn}(s_i)], \quad i = 1, \dots, M$$

Then, if  $w$  is the Hamming weight of the discrepancy sequence  $\{\delta_i, i = 1, \dots, M\}$ ,

$$w = \sum_{i=1}^M \delta_i$$

the statistic  $x$  can be written as

$$\begin{aligned} x &= \sum_{i=1}^{M-w} \frac{1}{2} [1 - \text{sgn}(m_i) \text{sgn}(m_i + n_i)] \\ &\quad + \sum_{i=1}^w \frac{1}{2} [1 - \text{sgn}(m_i) \text{sgn}(-m_i + n_i)] \\ &= \sum_{i=1}^{M-w} x_i + \sum_{i=1}^w y_i \end{aligned}$$

where  $\{x_i, i = 1, \dots, M-w\}$  are independent binary random variables as defined above and  $\{y_i, i = 1, \dots, w\}$  are

also independent binary random variables with a different probability,

$$\text{Prob}[y_i = 1] = p = 1 - \text{Prob}[y_i = 0]$$

$$p = 1 - Q\left(\frac{m}{\sigma}\right) = 1 - q$$

The probability of false alarm is obtained by averaging the conditional probability that  $x$  will fall below threshold over the possible discrepancy weights  $w$ ,

$$P_F = \sum_{w=0}^M \text{Prob}[w] \times \sum_{\substack{j=0 \\ j+k \leq \theta}}^{M-w} \sum_{k=0}^w \binom{M-w}{j} q^j (1-q)^{M-w-j} \times \binom{w}{k} p^k (1-p)^{w-k}$$

The discrepancy weight probability distribution  $\text{Prob}[w]$  was obtained by exhaustive enumeration for a 32-bit sync marker sequence and the NASA-standard constraint length 7, rate 1/2 convolutional code. This distribution is approximately binomial except at the tails ( $w$  near 0 or  $M$ ). The exact distribution for the NASA-standard code is reported in Appendix A.

The tradeoff between miss probability and false alarm probability for the hard-limited channel symbol discrepancy count statistic is plotted in Fig. 4 for signal-to-noise ratios  $E_b/N_0$  ranging from -1.5 dB to +3.0 dB.

**4. Channel Symbol Correlation Statistic.** Measured at the marker, the channel symbol correlation statistic  $x$  is a sum of independent Gaussian random variables,

$$x = \sum_{i=1}^M (-m_i n_i) = \sum_{i=1}^M u_i$$

where  $u_i$  is  $N(0, s^2 \sigma^2)$ ,  $x/\sigma^2$  is  $N(0, M s^2/\sigma^2) = N(0, M\rho)$ . Thus, the miss probability is calculated simply as

$$P_M = Q\left(\frac{\theta}{\sqrt{M s \sigma}}\right) = Q\left(\frac{\theta/\sigma^2}{\sqrt{M\rho}}\right)$$

Away from the marker, the correlation statistic  $x$  is a sum of conditionally Gaussian random variables, some with zero mean and some with nonzero mean:

$$x = \sum_{i=1}^{M-w} (-m_i n_i) + \sum_{i=1}^w (2m_i^2 - m_i n_i) = \sum_{i=1}^{M-w} u_i + \sum_{i=1}^w v_i$$

where  $v_i$  is  $N(2s^2, s^2 \sigma^2)$ ,  $x/\sigma^2$  is  $N(2ws^2/\sigma^2, M s^2/\sigma^2) = N(2w\rho, M\rho)$ .

The false alarm probability is obtained by averaging the conditional Gaussian probability distribution for  $x$  over the discrepancy weight distribution  $\text{Prob}[w]$ ,

$$P_F = \sum_{w=0}^M \text{Prob}[w] Q\left(\frac{2w\rho - \theta/\sigma^2}{\sqrt{M\rho}}\right)$$

Figure 5 shows  $P_M$  versus  $P_F$  for the channel symbol correlation statistic.

### 5. Channel Symbol Squared Distance Statistic.

At the marker, the squared distance statistic  $x$  is a sum of squares of  $M$  zero-mean Gaussian random variables, i.e.,  $x$  is a chi-squared random variable with  $M$  degrees of freedom:

$$x = \sum_{i=1}^M n_i^2$$

where  $n_i$  is  $N(0, \sigma^2)$ ,  $x/\sigma^2$  is  $\chi^2(M)$ . Away from the marker, the squared-distance statistic is conditionally a noncentral chi-squared random variable,

$$\begin{aligned}
x &= \sum_{i=1}^{M-w} n_i^2 + \sum_{i=1}^w (2m_i - n_i)^2 \\
&= \sum_{i=1}^{M-w} n_i^2 + \sum_{i=1}^w z_i^2
\end{aligned}$$

where  $z_i$  is  $N(2m_i, \sigma^2)$ ,  $x/\sigma^2$  is noncentral  $\chi^2(M, 4w\rho)$ . The probability distribution functions for central and non-central chi-squared random variables with an even number of degrees of freedom can be related to the probability distributions for certain Poisson random variables, as shown in Appendix B. Assuming that the number of marker symbols  $M$  is even, the miss probability and false alarm probability can thus be expressed as

$$\begin{aligned}
P_M &= \text{Prob} \left[ \sum_{i=1}^M n_i^2 > \theta \right] \\
&= \text{Prob} \left[ k_1 \leq \frac{M}{2} - 1 \right] \\
P_F &= \sum_{w=0}^M \text{Prob}[w] \text{Prob} \left[ \sum_{i=1}^{M-w} n_i^2 + \sum_{i=1}^w z_i^2 \leq \theta \right] \\
&= \sum_{w=0}^M \text{Prob}[w] \text{Prob} \left[ k_1 - k_2(w) \geq \frac{M}{2} \right]
\end{aligned}$$

where  $k_1$  is Poisson  $(\theta/2\sigma^2)$ ,  $k_2(w)$  is Poisson  $(2w\rho)$ , and  $k_1, k_2(w)$  are conditionally independent given  $w$ . Evaluating the Poisson probabilities gives

$$\begin{aligned}
P_M &= \sum_{k_1=0}^{\frac{M}{2}-1} \frac{\left(\frac{\theta}{2\sigma^2}\right)^{k_1}}{k_1!} \exp\left(-\frac{\theta}{2\sigma^2}\right) \\
P_F &= \sum_{w=0}^M \text{Prob}[w] \sum_{k_2=0}^{\infty} \frac{(2w\rho)^{k_2}}{k_2!} \exp(-2w\rho) \\
&\quad \times \sum_{k_1=k_2+\frac{M}{2}}^{\infty} \frac{\left(\frac{\theta}{2\sigma^2}\right)^{k_1}}{k_1!} \exp\left(-\frac{\theta}{2\sigma^2}\right)
\end{aligned}$$

The resulting curves of  $P_M$  versus  $P_F$  are plotted in Fig. 6.

## C. Conclusions from the $P_M$ versus $P_F$ Curves

Several important conclusions about the relative performance achievable by the various sync statistics can be drawn by comparing the tradeoff curves in Figs. 2 through 6. A  $P_M$  versus  $P_F$  curve which is uniformly below and to the left of another indicates superiority in the corresponding sync statistic. For example, Figs. 2 and 3 appear to show that the decoded bit discrepancy span statistic is uniformly superior to the decoded bit discrepancy count statistic, although the advantage disappears at small  $P_F$ .

Among the channel symbol methods, the squared distance statistic yields superior performance for large  $E_b/N_0$ , while the correlation statistic overtakes it for very poor signal-to-noise ratio ( $E_b/N_0 < 0.5$  dB). The performance of the hard-limited symbol discrepancy count statistic resembles the correlation statistic's performance, but is decidedly inferior at very low signal-to-noise ratio ( $E_b/N_0 < 0$  dB). The theoretically near-optimum weighted symbol discrepancy statistic may be better than all three, but it was not analyzed here.

Comparing the performances of the decoded bit methods with those of the channel symbol methods is a more complicated issue, because the tradeoff curves are dissimilar in shape. The curves for the decoded bit statistics are much flatter than those for the channel symbol statistics. There is a distinct crossover point between the curve for one of the decoded bit statistics and the corresponding curve for one of the channel symbol statistics. To the left of the crossover point (small  $P_F$ ), the decoded bit statistic is superior, while the channel symbol statistic performs better on the other side (large  $P_F$ ). The judgment of which statistic is better depends on the desired operating point on the  $P_M$  versus  $P_F$  curves. As shown in Section IV, this desired operating point is a function of the frame length, the global acquisition strategy, and the preferred measure of acquisition timeliness.

The flatness of the  $P_M$  versus  $P_F$  curves for the decoded bit statistics is due to the bursty nature of the Viterbi decoder errors. A marker-length section of decoded bits is likely to be either entirely correct or mostly garbage. Thus, only a small improvement in miss probability can be obtained by increasing the threshold  $\theta$  above 0, while the

false alarm probability increases substantially. The consequence of the flat  $P_M$  versus  $P_F$  curves is that the performance of the decoded bit schemes is basically limited by the Viterbi decoder's error probability for decoding a 32-bit symbol. Arbitrarily small false alarm probability (down to a minimum of approximately  $2^{-L}$ , the probability that  $L$  data bits will accidentally reproduce the sync marker) can be purchased at little expense in  $P_M$ . However, reducing  $P_F$  too far has no benefit if the frame length is small. The acquisition failure probability is totally dominated by  $P_M$  in this case. Even with a large frame, there are other ways to reduce  $P_F$  dramatically, such as by requiring frame-to-frame verification before declaring sync. Such a technique has virtually no benefit if  $P_F$  can already be made small without it.

While the decoded bit curves are very flat, the channel symbol curves show a lot of elasticity. Appreciable changes in  $P_F$  result in appreciable changes in  $P_M$ , and vice-versa. These methods show stronger response to frame-to-frame verification strategies. Of course, the increased elasticity also means that all parameters must be more delicately adjusted for optimum behavior.

## IV. Probability of Acquisition of True Sync

### A. Assumptions About the Acquisition Strategy

The miss probability versus false alarm probability curves reflect the power of a given statistic  $x$  to discriminate in a pairwise manner between the true sync marker location and any one of the possible false locations. To compute the probability that sync will be acquired correctly, it is also necessary to consider the global strategy that combines all of the bit-by-bit or symbol-by-symbol tentative decisions based on  $x$  to arrive at a declared sync position.

The global strategy can be divided into two parts. First, how are the outcomes of all the tentative decisions within one frame period combined to select or rank the candidate marker locations based on a single frame of observations? Second, are observations from succeeding frames required to reevaluate the candidates from the first frame before declaring sync, and, if so, how are they applied? These two parts may be called the intraframe strategy and the frame-to-frame strategy, respectively.

In this article, it is assumed that no sync candidates emerge from any frame in which the threshold test is failed at all locations; when this happens, the sync search is restarted from scratch in the next frame. If the threshold test is passed at one or more locations within a frame, it is assumed that the *first* such location tested is the unique sync candidate based on that frame's observations. The current DSN sync algorithm can switch between this scheme and another slightly more accurate one that selects the *most probable* location (i.e., the one that passes the threshold test by the widest margin). The latter intraframe strategy is not considered in this article.

Two different frame-to-frame strategies are considered. The simplest frame-to-frame strategy requires no correlation with succeeding frames. Sync is declared or not declared solely on the basis of the intraframe results in a single frame. A second frame-to-frame strategy is one currently used by the DSN, which requires next-frame verification before declaration of sync. In this strategy, the unique sync candidate (if any) from any single frame is chosen as the sync location if and only if it is *verified once* in the next succeeding frame by repassing the threshold test at the corresponding location within that frame.

Each acquisition strategy and sync statistic must be evaluated relative to an appropriate measure of performance. The appropriate performance measure is not necessarily the same for different strategies. For example, minimum acquisition times under an elaborate frame-to-frame verification strategy are guaranteed to be longer than under a single-frame strategy. In this article, the probability of acquisition of true sync in one frame is used to compare the performance of different basic observables  $x$  under the single-frame acquisition strategy. On the other hand, the desired measure of performance for the strategy requiring next-frame verification is the probability of acquisition of true sync within *four* frames or less.

### B. Probability of Acquisition of True Sync in One Frame

For true sync to be declared after one frame length of observations, the true sync location must pass the threshold test, and no false sync location may pass the threshold test earlier. The probability that the true sync location passes the threshold test is  $1 - P_M$ . The probability that any particular false sync location passes the threshold test is  $P_F$ . As stated earlier, it is assumed that the same false

alarm probability  $P_F$  applies both to data-only locations and to locations partially overlapping the marker.

The probability that at least one false sync will be detected before the true sync location is tested can be upper bounded and approximated by multiplying the false alarm probability  $P_F$  by the total number of false sync locations tested prior to the true sync location. This number is a random variable that depends on the exact time the synchronization search was started relative to the true frame boundaries. If the synchronization search is started at a random time, the average number of false sync locations tested before the true sync location is  $(B - 1)/2$  for algorithms based on decoded bits or  $(S - 1)/2$  for algorithms based on channel symbols.

In summary, true sync can fail to be acquired in one frame if either the true sync location fails the threshold test, or if a false alarm occurs at any of the false sync locations tested earlier. These two events occur with probabilities  $P_M$  and  $[(B - 1)/2]P_F$  or  $[(S - 1)/2]P_F$ , respectively. Thus, the probability of acquisition of true sync within one frame is approximated for small  $P_M$  and  $P_F$  by

Prob[acquisition of true sync within one frame]

$$\approx 1 - P_M - \frac{B - 1}{2}P_F, \quad \text{for algorithms based on decoded bits}$$

$$\approx 1 - P_M - \frac{S - 1}{2}P_F, \quad \text{for algorithms based on channel symbols}$$

This probability is plotted in Fig. 7 for the five sync observables considered in Figs. 2 through 6, assuming a frame length of 10000 bits or 20000 symbols and an optimized threshold at each signal-to-noise ratio.

### C. Probability of Acquisition of True Sync Within Four Frames, With Required Next-Frame Verification

In this section, it is assumed that the frame-to-frame strategy requires that the threshold test be passed at the same location in two consecutive frames before sync is declared. For true sync to be declared within four frames,

the true sync location must pass the threshold test either in the first and second frames, or in the second and third frames after failing to pass in the first frame, or in the third and fourth frames after failing the test in the second frame. These three events happen with probabilities  $(1 - P_M)^2$ ,  $P_M(1 - P_M)^2$ , and  $P_M(1 - P_M)^2$ , respectively, assuming that observations are independent from frame to frame. In addition, it is necessary that no false sync location be detected and verified before true sync can be declared. Any particular false sync location is detected and verified on the second frame, third frame, or fourth frame with probabilities  $P_F^2$ ,  $P_F^2(1 - P_F)$ , or  $P_F^2(1 - P_F)$ , respectively. The corresponding probabilities that at least one false sync is detected and verified on the second frame, third frame, or fourth frame can be upper bounded and approximated by multiplying the former probabilities by the number of possible false sync locations.

To calculate the probability that a false sync declaration might preempt the possibility of declaring true sync, it is necessary to divide the possible false sync locations into two categories. Some false sync locations are first subjected to the threshold test prior to the true sync location, and some are first tested after the true sync location. The first category of false sync locations can preempt a true declaration of sync if both the false declaration and the true declaration occur in the *same* number of frames, while the second category causes trouble only if false sync is declared in *fewer* frames than true sync. If the frame sync process is started at a random point in the frame, the average number of false sync locations in each category is  $(B - 1)/2$  or  $(S - 1)/2$ .

The dominant contribution to the probability that a false sync will be declared before true sync comes from the probability that false locations in the first category are detected and verified at the first opportunity, i.e., within two frames. This probability is upper bounded and approximated by  $[(B - 1)/2]P_F^2$  or  $[(S - 1)/2]P_F^2$ . All other contributions to the probability of preemptive declaration of false sync involve terms of order  $P_M P_F^2$  and higher-order terms, and these terms are unimportant if both  $P_M$  and  $P_F$  are small.

The previous observations can be summarized as follows. True sync can fail to be acquired within four frames if the true sync location is not detected and verified within four frames or if a false sync location is detected and verified earlier. The first event happens with probability  $1 - (1 - P_M)^2(1 + 2P_M) \approx 3P_M^2$ , and the second event

happens with approximate probability  $[(B - 1)/2]P_F^2$  or  $[(S - 1)/2]P_F^2$ . Thus, the probability of true acquisition within four frames is approximated for small  $P_M$  and  $P_F$  by

Prob [acquisition of true sync within four frames

with one required verification]

$$\approx 1 - 3P_M^2 - \frac{B-1}{2}P_F^2, \text{ for algorithms based}$$

on decoded bits

$$\approx 1 - 3P_M^2 - \frac{S-1}{2}P_F^2, \text{ for algorithms based}$$

on channel symbols

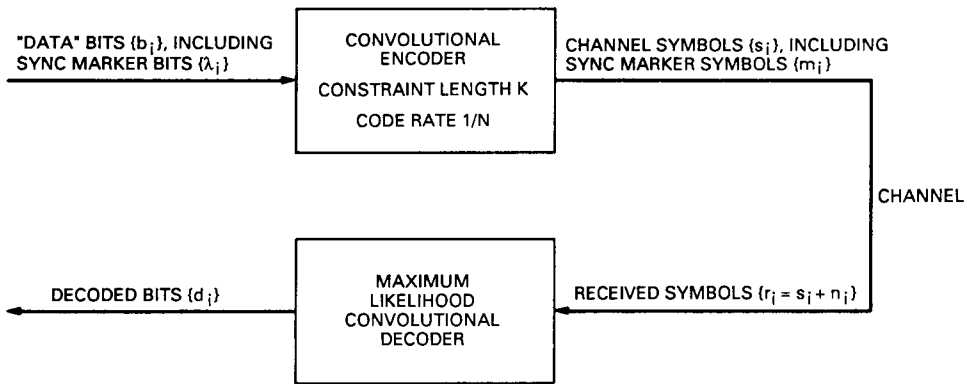
This probability is plotted in Fig. 8 for the same parameters considered in Fig. 7.

## V. Summary

This article has analyzed three different sync statistics based on measuring channel symbols. The perfor-

mance of these three statistics was analyzed without statistical approximation for the basic tradeoff curves of miss probability and false-alarm probability. These basic performance curves were then extended to yield expressions for the probability of timely acquisition under two different acquisition strategies.

It was shown that the statistics based on decoded bits are superior to the statistics based on channel symbols if the desired operating region utilizes a miss probability many orders of magnitude higher than the false-alarm probability. This operating point is applicable for very large frame lengths and minimal frame-to-frame verification strategy. On the other hand, the statistics based on channel symbols are superior if the desired operating point has a miss probability only a few orders of magnitude greater than the false-alarm probability. This happens for small frames or when frame-to-frame verifications are required. Among the three channel symbol methods examined, the squared-distance statistic offers the best performance in the range of normal signal-to-noise ratios, but it degrades more rapidly than the correlation statistic or the hard-limited symbol discrepancy count statistic when the signal-to-noise ratio is extremely low.

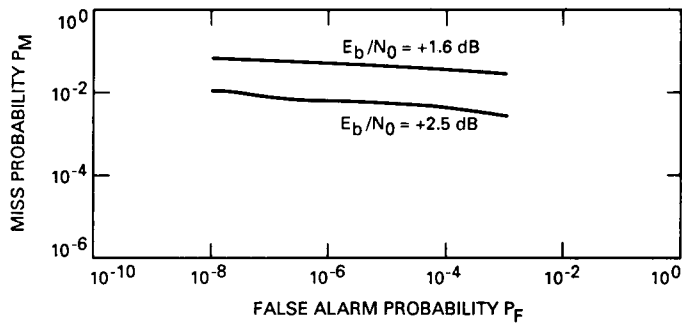


SYNC MARKER BITS  $\{\lambda_i, i = 1, \dots, L\}$  ARE TRANSMITTED ONCE EVERY DATA FRAME  $\{b_i, i = 1, \dots, B\}$

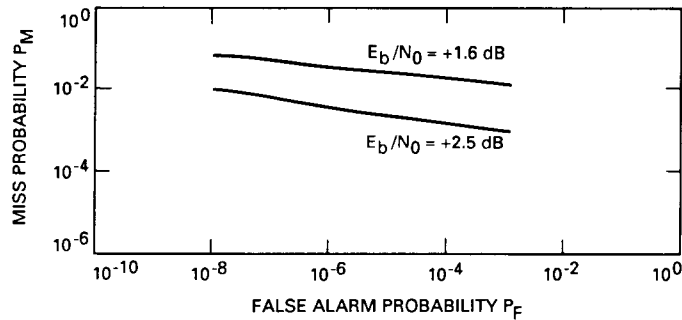
WITHIN EVERY SYMBOL FRAME  $\{s_i, i = 1, \dots, S\}$  THE SYNC MARKER SYMBOLS  $\{m_i, i = 1, \dots, M\}$  ARE THOSE SYMBOLS THAT ARE COMPLETELY DETERMINED BY THE SYNC MARKER BITS

$$S = NB \quad M = N(L - K + 1)$$

**Fig. 1. System diagram for two sync methods.**



**Fig. 2.  $P_M$  versus  $P_F$  for discrepancy count statistic based on decoded bits.**



**Fig. 3.  $P_M$  versus  $P_F$  for discrepancy span statistic based on decoded bits.**

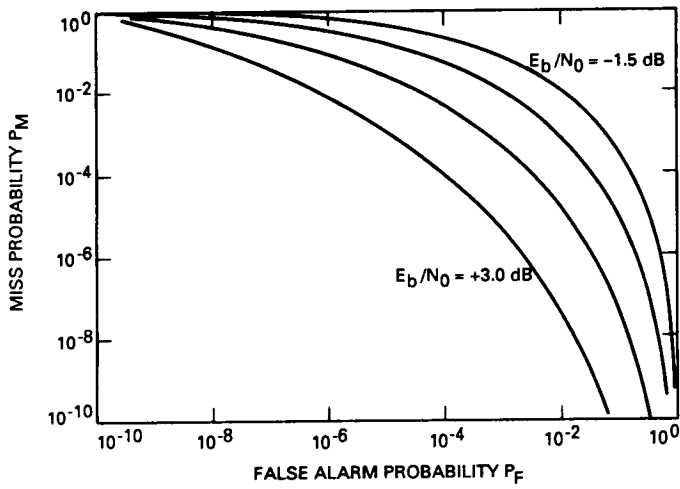


Fig. 4.  $P_M$  versus  $P_F$  for hard-limited symbol discrepancies.

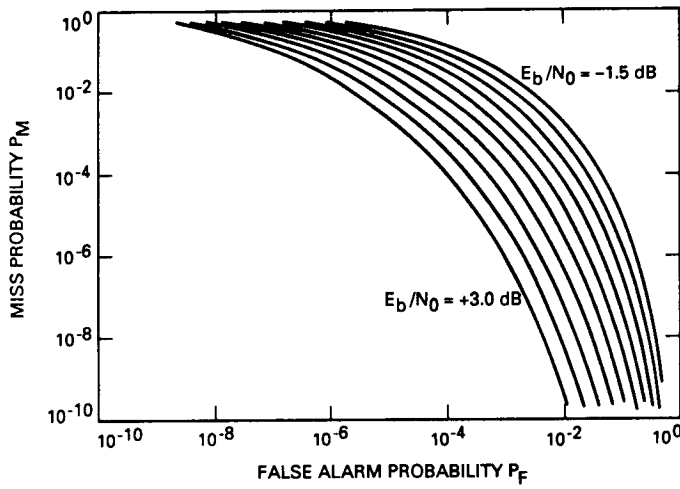


Fig. 5.  $P_M$  versus  $P_F$  for correlation statistic.

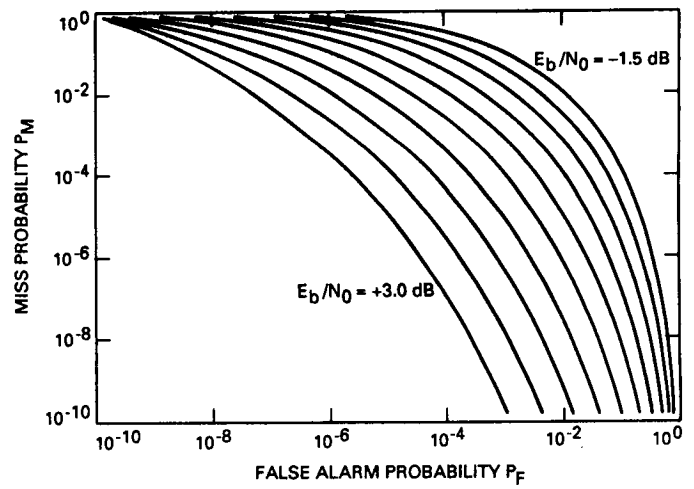


Fig. 6.  $P_M$  versus  $P_F$  for squared-distance statistic.



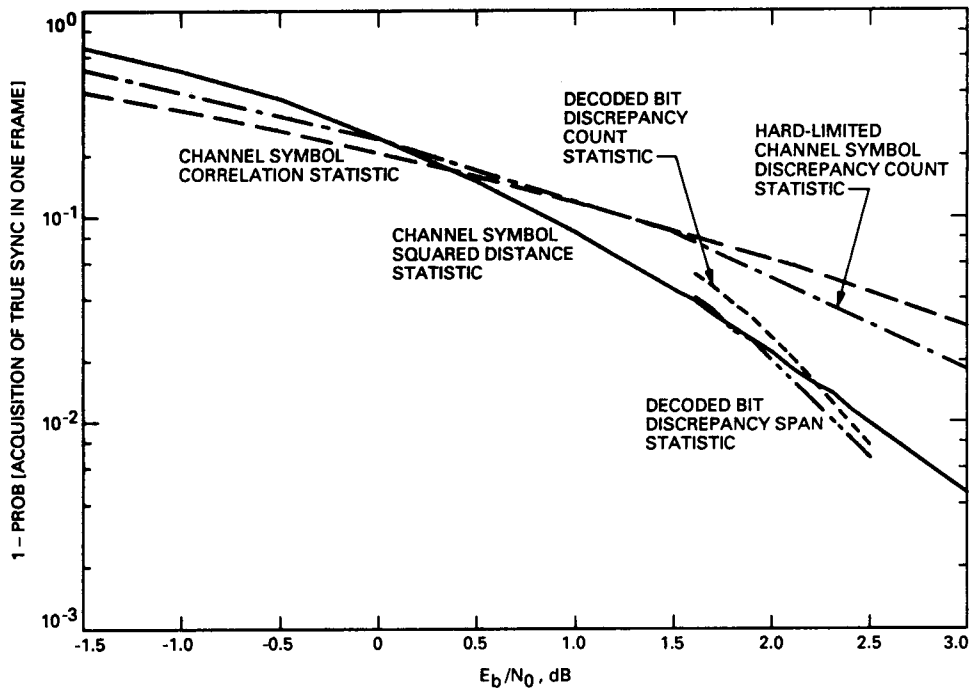


Fig. 7. Probability of failure to acquire true sync within one 10000-bit frame.

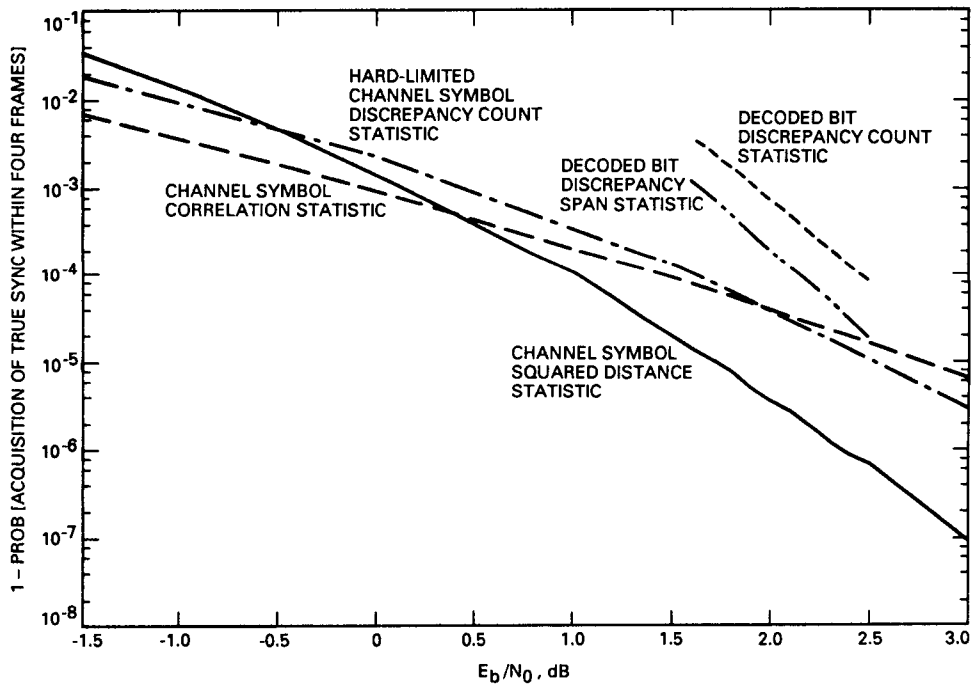


Fig. 8. Probability of failure to acquire true sync within four 10000-bit frames, with required next-frame verification.

## Appendix A

### 52-Symbol Weight Distribution for the NASA-Standard (7,1/2) Code

Evaluation of the false-alarm probabilities for the three channel symbol statistics analyzed in Section III requires knowledge of the probability distribution of the Hamming weights of the possible discrepancy sequences  $\{\delta_i, i = 1, \dots, M\}$  between the sync marker symbols  $\{m_i, i = 1, \dots, M\}$  and the sliding window of  $M$  symbols  $\{s_i, i = 1, \dots, M\}$  at the current location. The exact weight distribution for a 32-bit sync marker and the NASA-standard (7,1/2) code was generated by exhaustive enumeration. Since the convolutional code is linear, this weight profile is equivalent to the weight profile generated by encoding all possible 32-bit patterns. The  $2^{32}$  possible bit patterns were encoded one by one, and a complete histogram of their encoded weights was produced. This process required about 14 days of CPU time on a Sun 3 computer.

It is shown in [5] and [6] that the weight distribution of a binary primitive block code is approximately bino-

mial. Since the convolutional code is linear, and the 52-bit patterns represent truncated sequences from the output of the convolutional encoder, these 52-bit patterns can be regarded as the codewords of a (52,32) binary block code. Let  $A_w$  denote the number of patterns of weight  $w$ . Then according to [5] and [6]  $A_w$  can be approximated by the following binomial coefficient:

$$A_w \approx 2^{-20} \binom{52}{w} \equiv A'_w, \quad 0 \leq w \leq 52$$

A comparison between the exact weight distribution  $A_w$  and the approximate distribution  $A'_w$  is given in Table A-1. It is observed that  $A'_w$  is a good approximation to  $A_w$  except at the two extremes ( $w$  near 0 or 52).

**Table A-1. Encoded weight profile for all possible 32-bit input sequences, encoded by the NASA-standard (7,1/2) code**

Weight, $w$	Histogram totals		Ratio, $A_w/A'_w$
	Actual, $A_w$	Binomial, $A'_w$	
0	1	0	1048576.000
1	0	0	0.000
2	2	0	1581.563
3	7	0	332.128
4	18	0	69.718
5	40	2	16.138
6	122	19	6.284
7	333	128	2.610
8	963	718	1.342
9	3634	3509	1.036
10	14244	15087	0.944
11	54082	57605	0.939
12	191047	196819	0.971
13	600241	605596	0.991
14	1692962	1687018	1.004
15	4310848	4273778	1.009
16	9926772	9883112	1.004
17	20917283	20928944	0.999
18	40618430	40695169	0.998
19	72723310	72822934	0.999
20	120117817	120157841	1.000
21	183174400	183097662	1.000
22	258139236	258001251	1.001
23	336646298	336523371	1.000
24	406667062	406632406	1.000
25	455311348	455428295	1.000
26	472746296	472944768	1.000
27	455311348	455428295	1.000
28	406667062	406632406	1.000
29	336646298	336523371	1.000
30	258139236	258001251	1.001
31	183174400	183097662	1.000
32	120117817	120157841	1.000
33	72723310	72822934	0.999
34	40618430	40695169	0.998
35	20917283	20928944	0.999
36	9926772	9883112	1.004
37	4310848	4273778	1.009
38	1692962	1687018	1.004
39	600241	605596	0.991
40	191047	196819	0.971
41	54082	57605	0.939
42	14244	15087	0.944
43	3634	3509	1.036
44	963	718	1.342
45	333	128	2.610
46	122	19	6.284
47	40	2	16.138
48	18	0	69.718
49	7	0	332.128
50	2	0	1581.563
51	0	0	0.000
52	1	0	1048576.000
Total	$2^{32}$	$2^{32}$	

## Appendix B

### Relationship of Noncentral Chi-Squared Probability Distributions and Poisson Probability Distributions

Let  $\xi$  be a noncentral chi-squared random variable with  $M$  degrees of freedom and noncentrality parameter  $\mu$ . Then, from [7], the probability distribution function of  $\xi$  can be expanded in the form

$$\text{Prob}[\xi_k > \theta] = \sum_{j=0}^{(M/2)+k-1} e^{-\theta/2} \frac{(\theta/2)^j}{j!}$$

These two results combine to yield

$$\text{Prob}[\xi > \theta] = \sum_{k=0}^{\infty} e^{-\mu/2} \frac{(\mu/2)^k}{k!} \text{Prob}[\xi_k > \theta]$$

where  $\{\xi_k, k = 1, 2, \dots\}$  are central chi-squared random variables with increasing numbers of degrees of freedom,

$$\xi_k \text{ is } \chi^2(M + 2k), \quad k = 1, 2, \dots$$

Also from [7], the probability distributions for each of the central chi-squared random variables can be related to the Poisson distribution if  $M$  is even, according to

$$\begin{aligned} \text{Prob}[\xi > \theta] &= \sum_{k=0}^{\infty} e^{-\mu/2} \frac{(\mu/2)^k}{k!} \sum_{j=0}^{(M/2)+k-1} e^{-\theta/2} \frac{(\theta/2)^j}{j!} \\ &= \text{Prob} \left[ j - k \leq \frac{M}{2} - 1 \right] \end{aligned}$$

where  $j$  and  $k$  are independent Poisson random variables with means  $\theta/2$  and  $\mu/2$ , respectively.

510-32

239820

547663 p4

N90-12796

TDA Progress Report 42-98

August 15, 1989

98.

# Performance of the Image Statistics Decoder in Conjunction With the Goldstone-VLA Array

H. C. Wang and G. H. Pitt III

Communications Systems Research Section

*During Voyager's Neptune encounter, the National Radio Astronomy Observatory's Very Large Array (VLA) will be arrayed with Goldstone antennas to receive the transmitted telemetry data from the spacecraft. The telemetry signal from the VLA will drop out periodically, resulting in a periodic drop in the received signal-to-noise ratio (SNR). The Image Statistics Decoder (ISD), which assumes a correlation between pixels, can improve the bit error rate (BER) for images during these dropout periods. Simulation results have shown that the ISD, in conjunction with the Goldstone-VLA array can provide a 3-dB gain for uncompressed images at a BER of  $5.0 \times 10^{-3}$ .*

## I. Background

Throughout the Voyager mission, the DSN has been used for telemetry reception. As Voyager approaches Neptune, the signal must travel a long distance to reach earth, resulting in a low received SNR. In order to assure that the received Voyager data still achieves an acceptable BER, the National Radio Astronomy Observatory's Very Large Array (VLA) in the high New Mexico desert can assist the DSN station at Goldstone in receiving the Voyager telemetry data during Neptune encounter [1]. However, VLA telemetry has a dropout period of 1.6 millisecc every 52 millisecc. In [2], it was concluded that due to the loss of data in the dropout periods, the VLA alone would not be

able to support the (7,1/2) convolutional code for providing an acceptable BER for Voyager telemetry data.

Arraying the VLA with Goldstone antennas will result in a 3-dB gain in SNR, except during the VLA dropout period.

## II. BER Improvement Using the ISD

The Image Statistics Decoder (ISD) has been developed for decoding uncompressed images [3]. When decoding uncompressed images, the ISD has superior BER

## References

- [1] L. Swanson, *A Comparison of Frame Synchronization Methods*, JPL Publication 82-100, Jet Propulsion Laboratory, Pasadena, California, December 15, 1982.
- [2] M. Shahshahani and L. Swanson, "A New Method for Frame Synchronization," *TDA Progress Report 42-86*, Jet Propulsion Laboratory, Pasadena, California, pp. 111-123, August 15, 1986.
- [3] R. L. Miller, L. J. Deutsch, and S. A. Butman, *On the Error Statistics of Viterbi Decoding and the Performance of Concatenated Codes*, JPL Publication 81-9, Jet Propulsion Laboratory, Pasadena, California, September 1, 1981.
- [4] J. L. Massey, "Optimum Frame Synchronization," *IEEE Transactions of Communications*, vol. COM-20, no. 2, pp. 115-119, April, 1972.
- [5] K. Cheung, "The Weight Distribution and Randomness of Linear Codes," *TDA Progress Report 42-97*, Jet Propulsion Laboratory, Pasadena, California, pp. 208-215, May 15, 1989.
- [6] K. Cheung, "On the Decoder Error Probability of Linear Codes," *TDA Progress Report 42-98*, this issue.
- [7] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series, pp. 941-942, December 1972.

performance over the Maximum-likelihood Convolutional Decoder (MCD) which is currently used for decoding all data from Voyager. The ISD's algorithm is based on the fact that adjacent pixels in an image are very similar in their intensities; hence, a pixel is likely to be close to its neighbors in intensity. During the VLA's dropout periods, only Goldstone will receive the signal from Voyager. Were the SNR to drop to a point where the MCD could not decode the data at an acceptable BER, the ISD may be used for decoding the data. In addition, during other situations when the received SNR dropped too low for the MCD to decode a recognizable image, the ISD could also be used as an alternative to the MCD.

It should be pointed out that due to the large amount of computations and comparisons the ISD needs, as well as the fact that it currently only exists in software, the ISD is much slower than the MCD. The ISD takes three days on a SUN 3/260 to decode one image. Table 1 in [2] shows the speed of the ISD with different computers.

### III. Simulation

Simulations have been performed of image reception at Goldstone alone and of an equal aperture Goldstone-VLA array. In each case, the MCD and the ISD were used on identical data. An image of the Uranus moon Miranda (Fig. 1) was provided by the Voyager project. It was first packed into image frames, which were then packed into playback frames. After the playback frames

were convolutionally encoded (just as the spacecraft would do), Gaussian noise was added to the stream of data. For the Goldstone-alone case, the noise was stationary; for the Goldstone-VLA case, the SNR ratio was dropped by 3 dB for 1.6 millisecond every 52 milliseconds to simulate the VLA telemetry dropout. These simulated received telemetry streams were then decoded in software, either with an MCD or with the ISD inside images, and with MCD outside images. After decoding, the nonimage parts of the frame were stripped off, and the resulting image was compared to the original image to compute bit and byte error rates.

Figure 2 shows the BER for both decoders. The Goldstone stand-alone system has a constant SNR, while the Goldstone-VLA array is at its nominal level 97 percent of the time and 3 dB lower during the dropout period.

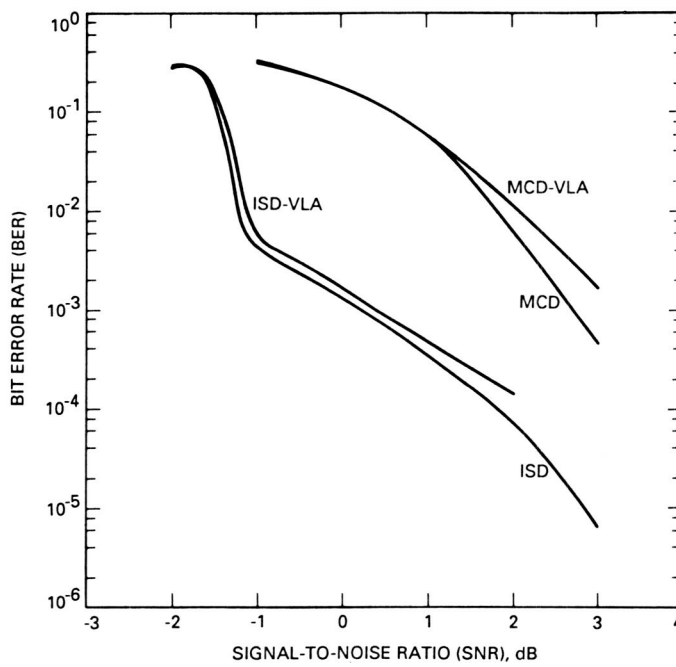
In the simulated SNR range, the ISD always has a better BER than the MCD. In particular, at a bit error rate of  $5.0 \times 10^{-3}$ , which is the maximum acceptable BER for image reception, the ISD has about a 3-dB gain. Note the steeper slope of the ISD curves at low SNRs. A conjecture for this phenomenon is a higher performance gain from a few good pixels among a large number of bad pixels at low SNRs. This is in contrast to the high SNR case where most pixels are good, hence yielding a lower performance gain with the addition of the same number of pixels. Figure 2 also shows that both the MCD and the ISD for the Goldstone-VLA array lose about 0.2 dB when compared to Goldstone alone at the same nominal SNR.

## References

- [1] J. W. Layland and D. W. Brown, "Planning for VLA/DSN Arrayed Support to the Voyager at Neptune," *TDA Progress Report 42-82*, vol. April-June 1985, pp. 125-135, Jet Propulsion Laboratory, Pasadena, California, August 15, 1985.
- [2] L. J. Deutsch, "An Update on the Use of the VLA for Telemetry Reception," *TDA Progress Report 42-72*, vol. October-December 1982, pp. 51-60, Jet Propulsion Laboratory, Pasadena, California, February 15, 1983.
- [3] G. H. Pitt III, L. Swanson, and J. H. Yuen, "Image Statistics Decoding for Convolutional Codes," *TDA Progress Report 42-90*, vol. April-June 1987, pp. 32-37, Jet Propulsion Laboratory, Pasadena, California, August 15, 1987.



**Fig. 1. Miranda.**



**Fig. 2. ISD and MSC performance compared at nominal SNR.**



# A Closed-Cycle Refrigerator for Cooling Maser Amplifiers Below 4 Kelvin

M. Britcliffe

Radio Frequency and Microwave Subsystems Section

*A helium refrigerator utilizing the Gifford-McMahon/Joule-Thomson (GM/JT) cycle was designed and tested to demonstrate the feasibility of using small closed-cycle refrigerators as an alternative to batch-filled cryostats for operating temperatures below 4 K. The systems could be used to cool low-noise microwave maser amplifiers located in large parabolic antennas. These antennas tilt vertically, making conventional liquid-filled dewars difficult to use. The system could also be used for a non-tilting beam waveguide antenna to reduce the helium consumption of a liquid helium cryostat. The prototype system is adjustable to provide 700 mW of cooling at 2.5 K to 3 W at 4.3 K. Performance of the unit is not significantly affected by physical orientation. The volume occupied by the refrigerator is less than 0.1 m<sup>3</sup>. Two JT expansion stages are used to maximize cooling capacity per unit mass flow. The heat exchangers were designed to produce minimum pressure drop in the return gas stream. Pressure drop for the entire JT return circuit is less than 5 kpa at a mass flow of 0.06 g/sec when operating at 2.5 K.*

## I. Introduction

The DSN currently uses low-noise maser amplifiers operating in 4.5-K closed-cycle refrigerators on 70- and 34-m Cassegrainian antennas. Planetary spacecraft programs planned for the 1990s would greatly benefit from improved downlink performance from these antennas. One strategy for improving downlink performance is to further lower the noise temperature of the antenna systems. Two methods

for reducing the system noise temperature are lowering the physical temperature of the maser well below the presently used 4.5 K, and cryogenically cooling the microwave components in the feed system between the feed horn and the maser.

New 34-m beam waveguide antennas are being implemented in the DSN that provide a large non-tipping location for the maser and feed components. For these

antennas, a batch-filled liquid helium cryostat provides a reliable method to accomplish both the above improvements. However, for the existing 70-m and 34-m antennas with feedcone-mounted masers, a new type of closed-cycle refrigerator that will operate at lower temperatures in a tilting environment is required. A similar closed-cycle refrigerator could also be used in conjunction with a liquid helium cryostat in the beam waveguide antennas to recondense the helium boil-off from the dewar and eliminate the need for manually refilling the system with liquid helium. This article describes the initial development of a refrigerator configuration that could satisfy these requirements.

One of the simplest approaches to providing closed-cycle cooling at temperatures below 4 K is to operate a conventional Joule-Thomson/Gifford-McMahon (JT/GM) cooler at a subatmospheric JT return pressure by pumping the gas returning from the JT circuit with a vacuum pump. The final temperature achieved using this configuration is a function of the pressure drop in the return path of the JT circuit heat exchangers and the capacity of the vacuum pump. The cooling capacity available at the final stage is limited by the JT mass flow, which is a function of the capacity of the GM expander and the efficiency of the heat exchangers. Unfortunately, the pressure drop in the heat exchangers also increases as the square of the mass flow, which limits the ultimate cooling capacity for a given operating temperature.

The refrigerator described in this article was designed specifically for operation at subatmospheric JT return pressures. A two-stage JT expansion was used to maximize the available cooling capacity per unit mass flow. The heat exchangers were designed to provide low pressure drop in the return path and to take maximum advantage of the GM engine capacity at each heat exchanger stage. Adjustable expansion valves are used to allow the cooling capacity and temperature to be varied from 700 mW at 2.5 K to 3 W at 4.4 K.

## II. Joule-Thomson Circuit Capacity

The refrigeration capacity of the final stage of a conventional Joule-Thomson refrigerator can be approximated by doing an enthalpy balance analysis of a system consisting of the final heat exchanger, the JT expansion valve, and the final heat station (see Fig. 1). If the capacity of the refrigerator is considered to be the constant temperature capacity (i.e., the the maximum heat load that can be applied without warming the refrigerator above the saturation temperature), the fluid at the outlet of the heat

station can be considered to be saturated vapor. In this case the refrigeration capacity is equal to the heat of vaporization of the liquid fraction of the flow through the JT valve. The constant temperature capacity is given by the following:

$$q = h_3 - h_2 \text{ or } q = h_{\text{sat}} - h_2 \quad (1)$$

where  $h_{\text{sat}}$  is the enthalpy of saturated vapor at the final heat exchanger return side inlet and  $h_2$  is the enthalpy of the fluid exiting the JT valve. If the JT expansion is isenthalpic then  $h_2$  is equal to the value of the enthalpy leaving the cold side of the heat exchanger. Then,  $h_2$  can be expressed as follows:

$$h_2 = h_1 - q_{\text{he}} = h_1 - (h_4 - h_{\text{sat}}) \quad (2)$$

Where  $h_1$  is the enthalpy of the fluid at the inlet of the heat exchanger,  $q_{\text{he}}$  is the heat transferred from the warm stream to the cold stream in the exchanger, and  $h_4$  is the enthalpy at the outlet of the cold return of the heat exchanger. Combining Eqs. (1) and (2) gives the following:

$$q = h_4 - h_1 \quad (3)$$

Heat exchanger efficiency is defined by the following:

$$e = \frac{q_{\text{he}}}{q_{\text{max}}} = h_{\text{out}} - \frac{h_{\text{in}}}{h_{\text{out}'}} - h_{\text{in}} \quad (4)$$

where  $q_{\text{he}}$  is the actual heat exchange between the warm and cold streams in the heat exchanger and  $q_{\text{max}}$  is the maximum possible heat transfer with no thermal resistance between the warm and cold stream. If the heat capacities of the supply stream and the return stream are different, the heat exchanger is "pinched." This means that the heat transfer between the two streams is limited to the smaller of the two enthalpy changes. The primed enthalpy value indicates the fluid properties that would be achieved if the thermal resistance in the heat exchanger is zero.

In the case of operation near 4 K (100-kpa return pressure) a JT supply pressure can be chosen (1.6 Mpa) so that the heat exchanger pinch that occurs in any of the

stages is small. Equation (4) can be written in terms of the enthalpy change as follows:

$$h_4 = e_{h_4'} \quad (5)$$

Substituting this expression into Eq. (3) yields the following:

$$q = e_{h_4'} - h_1 \quad (6)$$

This shows that the final capacity of the Joule-Thomson stage is a function of the enthalpy of the fluid entering the final heat exchanger and the heat exchanger efficiency.

At temperatures below 4 K there is no single operating point for the JT supply pressure that will prevent an imbalance of the heat capacity between the supply and return gas in either the final heat exchanger or the second-stage heat exchanger. If the enthalpy change ( $h_4 - h_{\text{sat}}$ ) of the final stage return stream is smaller than the enthalpy change ( $h_2' - h_1$ ) in the supply stream, the temperature of the fluid entering the expansion valve will be higher. This will result in a smaller fraction of the JT flow being liquified and therefore reduce the capacity. A pinch in the second-stage heat exchanger upstream of the GM second stage will increase the loading on the GM stage and increase the temperature of the gas entering the final stage.

Collins [1] suggested using multiple stages of JT expansion to optimize the heat transfer in the JT heat exchangers as a method to avoid this problem. The heat capacity in the warm and cold streams in the heat exchangers can be balanced more closely by expanding the fluid in several steps along the final heat exchanger. This process results in the enthalpy of the fluid being lowered to the minimum level possible before the final expansion valve. This causes the largest possible fraction of the flow through the final expansion to be liquefied and yields the maximum final stage cooling capacity. Figure 2 shows the schematic of a refrigerator using two stages of JT expansion. Figure 3 represents the cycle on a temperature-entropy diagram. The dotted line indicates the path for a single JT expansion. The area under the constant temperature line at 2.5 K represents the cooling capacity of the refrigerator. The shaded area between the final expansion end points represents the cooling capacity gained by using two expansions.

### III. Heat-Exchanger Return-Path Pressure Drop

Return-path pressure drop is a critical parameter in the design of JT refrigerators because it determines the pressure and therefore the temperature of the helium bath in the final stage. Designing heat exchangers for these refrigerators requires a compromise between heat exchange and pressure drop. Factors that improve heat exchange (small cross-sectional area and long path lengths) tend to increase pressure drop.

The pressure drop for heat exchangers with flow across tubes is given by Hogan [2] to be the following:

$$\Delta P = f \frac{L}{D_e} \frac{M^2}{A^2 \rho} \quad (7)$$

where  $f$  is the friction factor, which is a function of the Reynolds number that involves the tube size and the fluid properties in the heat exchanger;  $L$  is the heat exchanger length, and  $D_e$  is the hydraulic diameter of the flow passage;  $M$  is the JT mass flow rate,  $A$  is the passage cross-sectional area, and  $\rho$  is the fluid density. The mass flow squared dependence of pressure drop shows the need to keep the JT flow as low as possible. Table 1 shows the calculated pressure drop for the three heat exchangers in the JT circuit for a nominal mass flow rate of 0.08 g/sec. The pressure drop in the first-stage exchanger is the largest contributor to pressure drop by a factor of 10.

### IV. Hardware Description

A prototype closed-cycle refrigerator was built to study the feasibility of using the multiple JT valve concept to obtain temperatures of 2 K and below. The device features a two-stage JT expansion with adjustable expansion valves to optimize cooling capacity as a function of mass flow. The heat exchangers were designed to provide minimum pressure drop and still provide adequate heat transfer. Many of the components used in the fabrication of the prototype such as the final cold stage and the thermal switch used to aid in the cooldown of the system were adopted from existing JPL designs and are described by Higa and Weibe [3].

The entire refrigerator is less than 70 cm long and 20 cm wide and occupies a volume of less than 0.1 m<sup>3</sup>. The JT circuit is contained in a cylindrical vacuum housing

that is 20 cm in diameter and 40 cm long. The complete assembly, shown in Fig. 4, weighs 20 kg.

### A. Gifford-McMahon Expander

The refrigerator uses a Varian two-stage GM expander designed for cryo-pump applications. The device was designed to provide cooling at 20 K and 80 K with nominal cooling capacities of 5 W and 35 W respectively. The measured cooling capacity of the device as a function of temperature using a JPL 5-hp compressor is plotted in Figs. 5 and 6. The JT circuit heat sinks were bolted to the displacer cylinder flanges rather than soldered directly to the cylinder.

### B. Heat Exchangers

The heat exchangers were adapted from those developed for the JPL prototype 2-W, 4-K refrigerator [4, 5]. They consist of a helix of corrugated bronze tubing wound on a Micarta mandrel. This assembly is housed in a tightly fitting stainless steel outer tube. The high pressure fluid travels through the inside diameter of the bronze tubing and the low pressure stream returns over the outside of the tube in the annulus between the inner mandrel and the outer tube in a cross-flow arrangement. The sizes and measured efficiencies of the heat exchangers are tabulated in Table 2.

### C. Compressor and Vacuum Pump

A JPL 5-hp two-stage compressor was combined with a 21-l/sec direct-drive vacuum pump (Leybold-Heraeus S 65 BC) to provide the proper operating pressures for the refrigerator. A schematic diagram of the system is shown in Fig. 7. The helium compressor will supply the GM expander with 2.0-Mpa helium and maintain a 700-kpa return pressure. The JT circuit is supplied with 1.0- to 2.0-Mpa gas. Gas returning from the JT circuit is evacuated by the vacuum pump and then returned to the compressor at 110 kpa. An oil separator returns any vacuum

pump oil vapor to the pump. The pressure at the vacuum pump inlet can be maintained below 50 torr for any flow rate below 0.25 g/sec, which is the highest practical flow rate for the refrigerator.

## V. Performance

The adjustable expansion valves allowed comparisons of the refrigerator's performance using single and double JT expansions. Final-stage heat capacity is plotted as a function of temperature for both cases in Fig. 8. Using a two-stage expansion results in a 30-percent increase in cooling capacity at 2.5 K. These measurements were taken with a fixed mass flow of 0.06 g/sec. Supply pressure for the single-stage measurement was 1.6 Mpa. For the two-stage measurement, the high pressure was 1.6 Mpa for the first stage and 600 kpa for the second stage.

Tests performed with an additional heat load applied to the first GM engine stage show that the efficiency of the first stage heat exchanger could be reduced by 30 percent and still allow normal JT stage operation. Utilizing a less efficient but lower pressure drop perforated-plate-type heat exchanger in the first stage should allow the refrigerator to operate below 2 K.

## VI. Conclusions

Tests of the prototype refrigerator confirm the feasibility of using a small GM/JT refrigerator for temperatures of 2.5 K and below. Two of the major factors involved with designing refrigerators for this application are minimizing the JT circuit return-path pressure drop and optimizing the refrigeration capacity per unit mass flow. Ninety percent of the pressure drop in the JT return path occurs in the first-stage (300-40 K) heat exchanger. This pressure drop can be lowered by using a less efficient plate-fin type device and allowing the first stage of the GM expander to absorb the excess heat. The heat capacity per unit mass flow can be improved by using a two-stage JT expansion.

## Acknowledgments

Thanks are due to Frank McCrea of Varian and Associates for providing technical support on the GM refrigerator and to Ted Hanson of Bendix Field Engineering for assembling the prototype.

## References

- [1] S. C. Collins, R. W. Stuart, and M. H. Streeter, *Review of Scientific Instruments*, no. 38, pp. 1654–1655, 1967.
- [2] W. H. Hogan, “Refrigerators and Cryostats for Superconducting Devices,” *Applied Superconductivity*, vol. 1, pp. 309–385, 1975.
- [3] W. Higa and E. Wiebe, “A Simplified Approach to Heat Exchanger Construction,” *Cryogenic Technology*, vol. 3, pp. 47–51, March/April 1967.
- [4] M. Britcliffe, “Two-Watt 4-Kelvin Refrigerator Performance,” *TDA Progress Report 42-91*, vol. July-September 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 312–317, November 15, 1987.
- [5] M. Britcliffe, “A Two-Watt 4-Kelvin Refrigerator System for Microwave Low-Noise Amplifiers,” *Cryogenic Processes and Equipment*, pp. 131–134, January 1989.

**Table 1. Calculated heat exchanger pressure drop**

Heat exchanger	Temperature, K	Pressure drop, Kpa
1 stage	175	3.9
2 stage	20	0.24
3 stage	7	0.06

**Table 2. Prototype heat exchanger efficiencies and sizes**

Heat exchanger	Efficiency, percent	Length, cm	Diameter, cm
1 stage	90	25	3.75
2 stage	97	20	3.10
3 stage	98	35	3.10

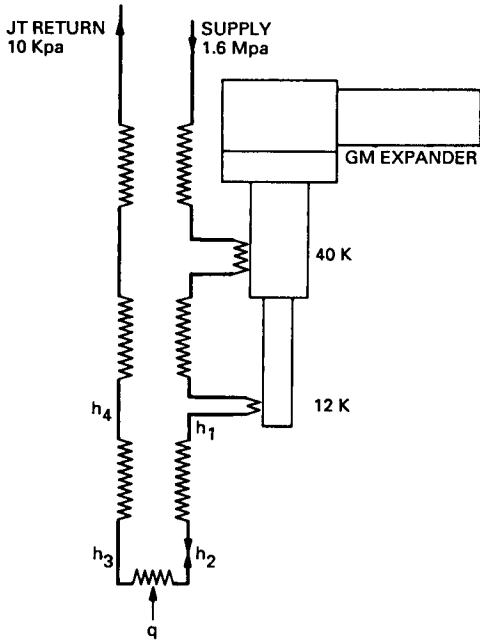


Fig. 1. Single-stage JT schematic diagram.

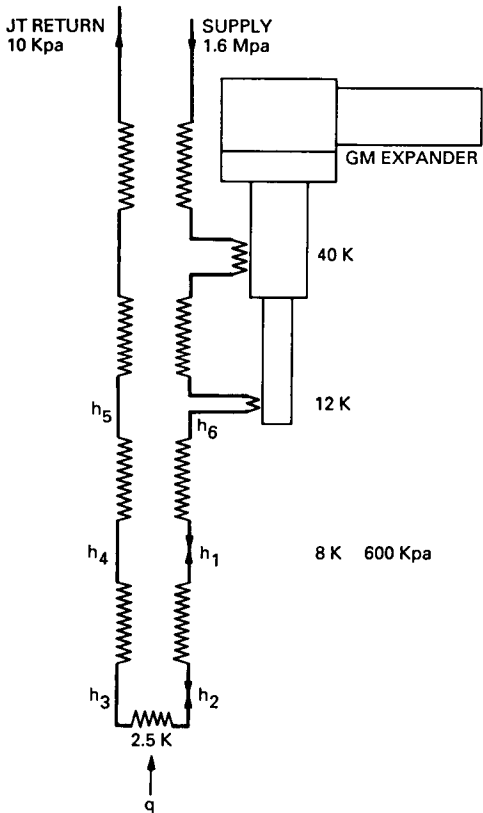


Fig. 2. Two-stage JT circuit schematic diagram.

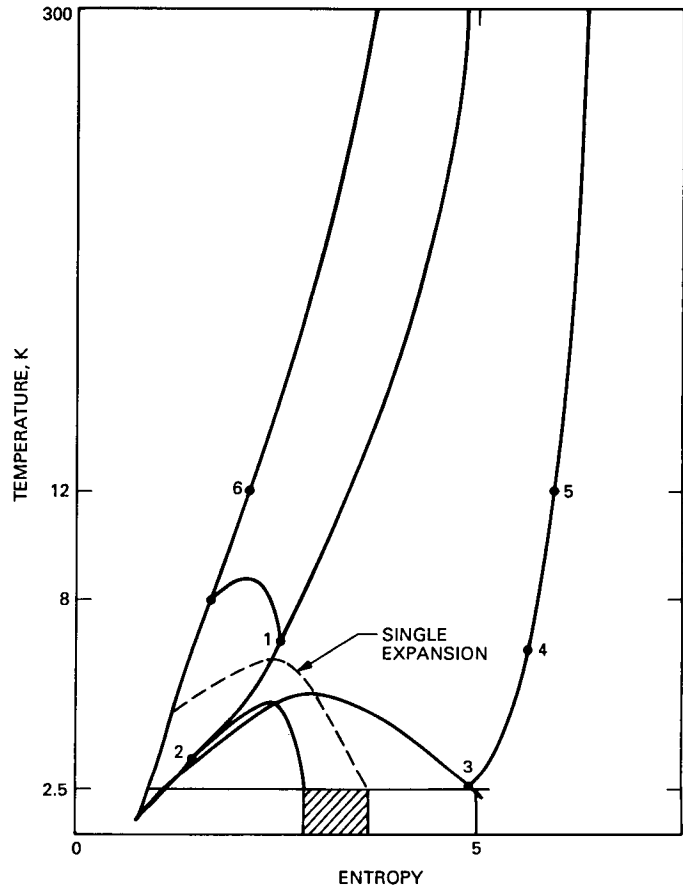
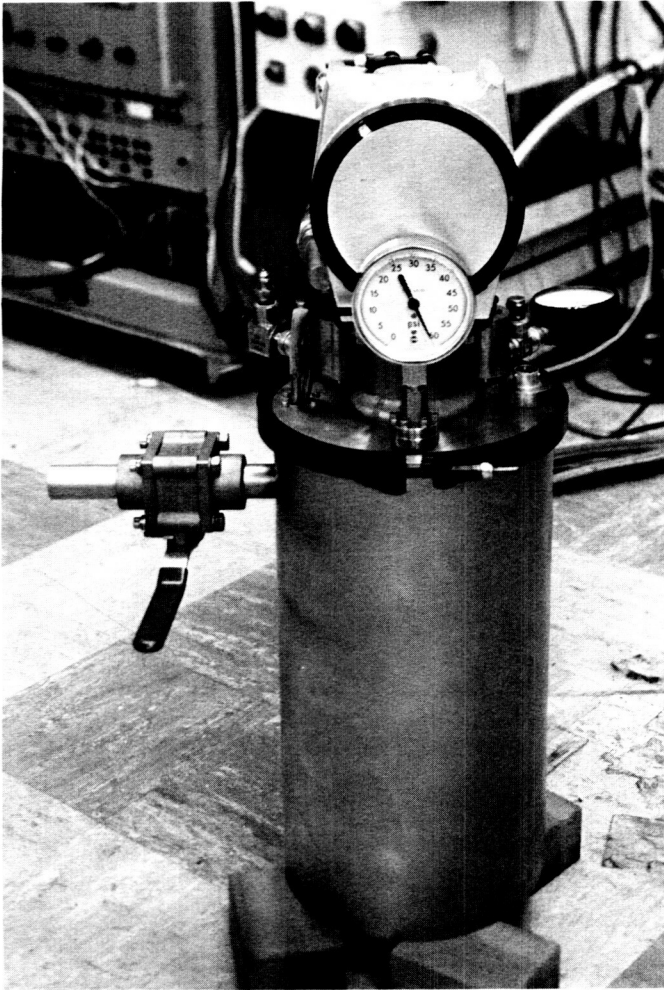
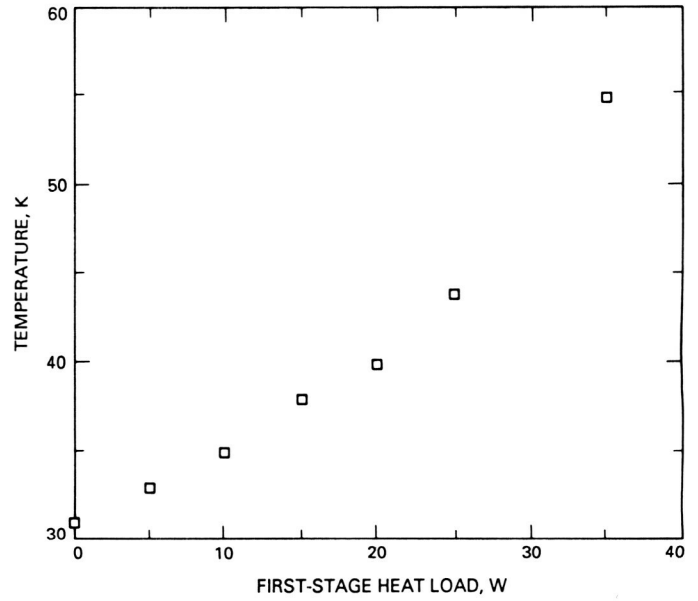


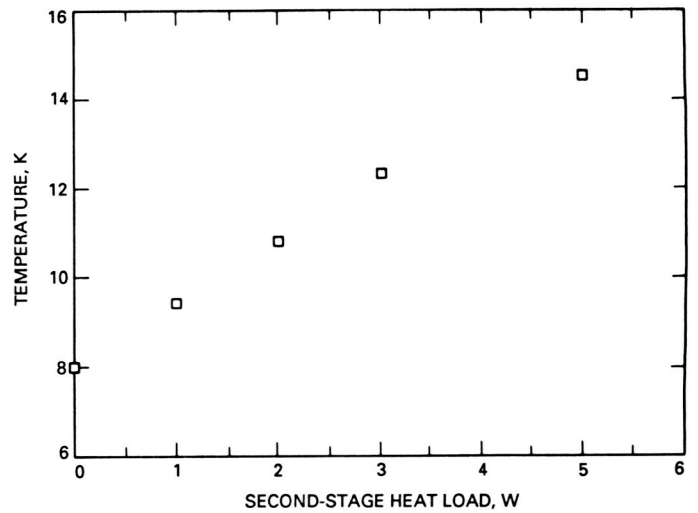
Fig. 3. Temperature-entropy diagram.



**Fig. 4. Prototype refrigerator.**



**Fig. 5. Varian GM refrigerator first-stage capacity.**



**Fig. 6. Varian GM refrigerator second-stage capacity.**



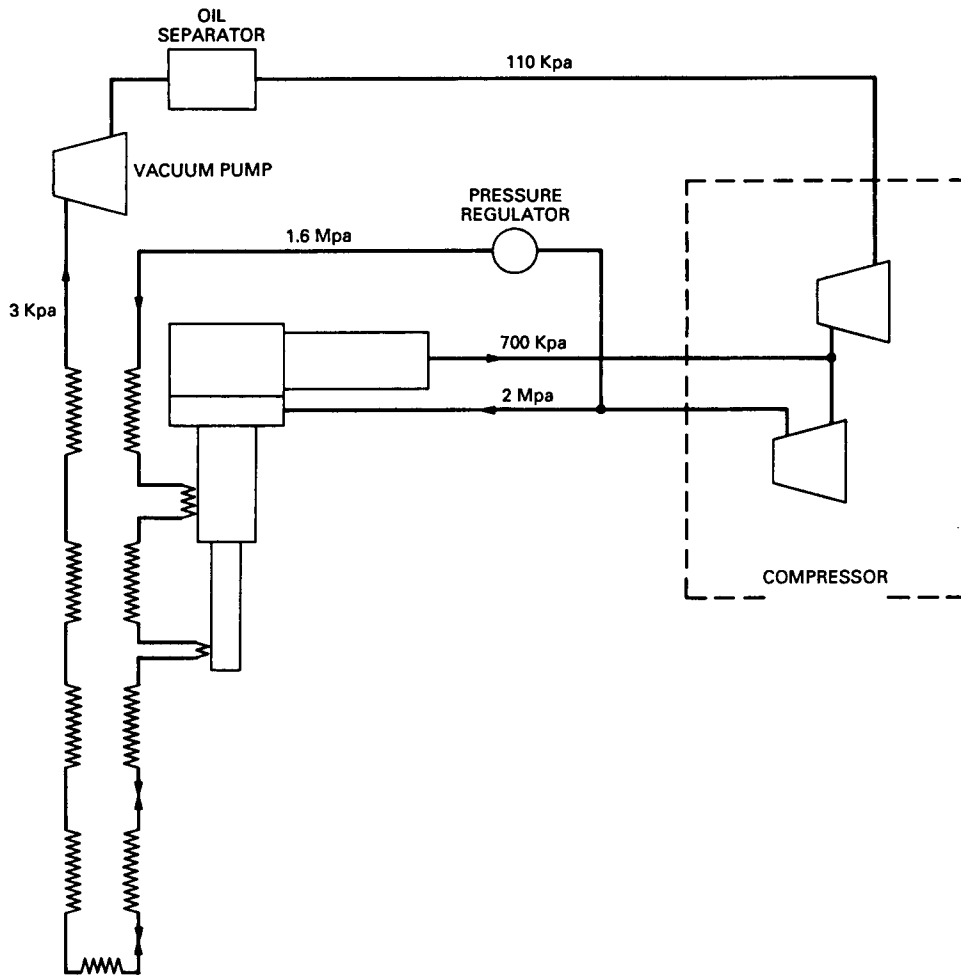


Fig. 7. Compressor and vacuum pump schematic diagram.

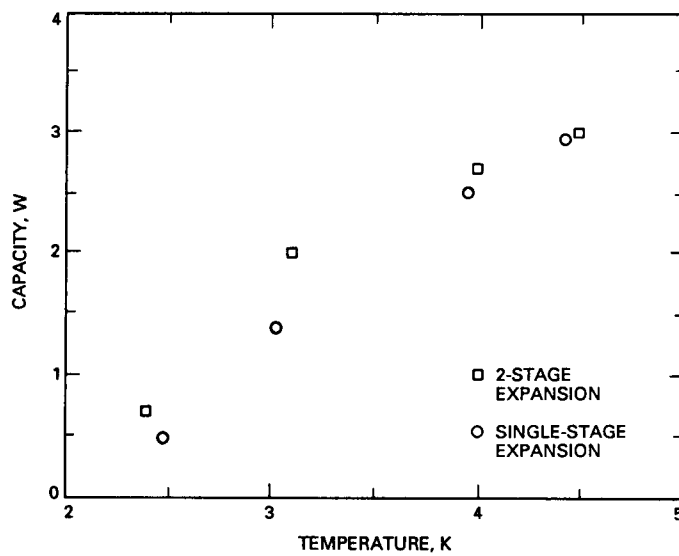


Fig. 8. Final-stage cooling capacity as a function of temperature.

72-32  
239822  
10.

547669 P7  
N90-12798

# Wideband Phase-Locked Angular Modulator

L. Nguyen

Radio Frequency and Microwave Subsystems Section

*A phase-locked loop (PLL) angular modulator scheme has been proposed which has the characteristics of wideband modulation frequency response. The modulator design is independent of the PLL closed-loop transfer function  $H(s)$ , thereby allowing independent optimization of the loop's parameters as well as the modulator's parameters. A phase modulator implementing the proposed scheme was built to phase modulate a low-noise phase-locked signal source at the output frequency of 2290 MHz. The measurement results validated the analysis by demonstrating that the resulting baseband modulation bandwidth exceeded that of the phase-locked loop by over an order of magnitude. However, it is expected to be able to achieve much wider response still.*

## I. Introduction

Phase-locked loops (PLL) are sometimes designed to mechanize the generation of phase modulation (PM) or frequency modulation (FM) signals. In a synthesized signal generator, the same frequency synthesis loops are often also used for providing PM and/or FM capability. As such an angular modulator, the PLL has the following desirable characteristics: the modulated carrier frequency can be stabilized by a stable reference signal; and a large modulation index can be obtained by suitable phase compression in the loop [1]. In a typical phase-locked angular modulator, shown in Fig. 1, the modulating signal is added to the appropriate baseband points in the loop to produce the desired PM or FM [2]. An analysis of this scheme shows that the modulation transfer function relating the output phase to the input phase modulating voltage is propor-

tional to the PLL closed-loop transfer function  $H(s)$  (a low-pass function), whereas the function relating the output instantaneous frequency to the input frequency modulating voltage is proportional to the complementary high-pass of  $H(s)$ . This dependence of the modulation transfer function on the PLL closed-loop characteristics has the following consequences:

- (1) The configuration of Fig. 1 does not lend itself to easy implementation of both PM and FM using the same loop. For PM, a sufficiently wide loop bandwidth is required to accommodate the high frequency modulating signal, whereas for FM a narrowband loop is needed to prevent the loop from tracking out the low frequency modulating signal.

- (2) Wideband modulation frequency response is often not possible since the design is usually constrained by considerations such as spurious signal rejection, loop stability, settling time, and noise performance that place upper and lower limits on the loop bandwidth.
- (3) The loop introduces phase distortion to the output modulating envelope because of its band-limited nature.

A phase-locked modulator scheme is proposed that eliminates the above limitations. It will be shown that the modulation transfer function resulting from this scheme, which has the characteristics of wideband frequency response, is independent of the PLL transfer function. An experimental phase modulator was built. The measured data substantiate the analysis by demonstrating that the resulting baseband modulation bandwidth exceeded that of the PLL by over an order of magnitude.

## II. Analysis

Figure 2 shows the generalized block diagram of the proposed modulator scheme. A noise-free and linearized model of a PLL consisting of a phase detector, loop filter, and a voltage-controlled oscillator (VCO) is assumed. The following notation will be used:

- $\Theta_R(s)$  = reference phase, rad  
 $\Theta_0(s)$  = output phase, rad  
 $\Theta_M(s)$  = output phase modulation, rad  
 $\Omega_M(s)$  = output frequency modulation, rad/sec  
 $K_D$  = phase detector gain, volt/rad  
 $K_0$  = VCOs gain constant, rad/sec/volt  
 $F(s)$  = loop filter  
 $P(s), Q(s)$  = modulation filters  
 $V_M(s)$  = input modulating signal, volt

The output phase for the loop in Fig. 2 is given by

$$\Theta_0(s) = \frac{K_0}{s} \left\{ P(s)V_M(s) + F(s) \left[ Q(s)V_M(s) + K_D (\Theta_R(s) - \Theta_0(s)) \right] \right\} \quad (1)$$

Simplifying yields

$$\begin{aligned} \Theta_0(s) &= \frac{K_0 K_D F(s)}{s + K_0 K_D F(s)} \Theta_R(s) \\ &+ \frac{K_0 P(s)}{s + K_0 K_D F(s)} V_M(s) \\ &+ \frac{K_0 F(s) Q(s)}{s + K_0 K_D F(s)} V_M(s) \end{aligned} \quad (2)$$

The quantity  $K_0 K_D F(s) / [s + K_0 K_D F(s)]$  is the closed-loop transfer function, denoted  $H(s)$ . Thus,

$$\begin{aligned} \Theta_0(s) &= H(s)\Theta_R(s) + \frac{K_0 P(s)}{s + K_0 K_D F(s)} V_M(s) \\ &+ \frac{K_0 F(s) Q(s)}{s + K_0 K_D F(s)} V_M(s) \end{aligned} \quad (3)$$

$\Theta_R(s)$  defines the reference phase for the loop to maintain lock and is constant as far as the modulation is concerned. So, the output phase modulation due to the input modulating signal is

$$\begin{aligned} \Theta_M(s) &= \frac{K_0 P(s)}{s + K_0 K_D F(s)} V_M(s) \\ &+ \frac{K_0 F(s) Q(s)}{s + K_0 K_D F(s)} V_M(s) \end{aligned} \quad (4)$$

which can be rearranged to yield

$$\begin{aligned} \Theta_M(s) &= \frac{s}{s + K_0 K_D F(s)} \frac{K_0 K_D}{s} P(s) \frac{V_M(s)}{K_D} \\ &+ \frac{K_0 K_D F(s)}{s + K_0 K_D F(s)} Q(s) \frac{V_M(s)}{K_D} \end{aligned} \quad (5)$$

or

$$\Theta_M(s) = \left\{ \left[ 1 - H(s) \right] \frac{K_0 K_D}{s} P(s) + H(s) Q(s) \right\} \frac{V_M(s)}{K_D} \quad (6)$$

The output instantaneous frequency is the derivative of the phase. Hence the output frequency modulation is

$$\begin{aligned}\Omega_M(s) &= s\Theta_M(s) \\ &= \left\{ [1 - H(s)]P(s) + H(s)\frac{s}{K_0K_D}Q(s) \right\} K_0V_M(s)\end{aligned}\quad (7)$$

It is evident from Eqs. (6) and (7) that the design strategy is to choose the correct modulation filters  $P(s)$  and  $Q(s)$  that would enable the appropriate compensations to eliminate the terms involving  $H(s)$  and thus allow the modulation response to be independent of the PLL transfer function parameters. That is, it is desirable to choose  $P(s)$  and  $Q(s)$  such that the sums inside the brackets of Eqs. (6) and (7) simplify to unity. The two cases of PM and FM are considered.

### A. Phase Modulation

Choose  $P(s) = s/K_0K_D$ . This is the transfer function of a differentiator where the time constant equals  $1/K_0K_D$ , and choose  $Q(s) = 1$ . Then Eq. (6) becomes

$$\Theta_M(s) = \left\{ [1 - H(s)]\frac{K_0K_D}{s}\frac{s}{K_0K_D} + H(s) \right\} \frac{V_M(s)}{K_D} \quad (8)$$

or

$$\Theta_M(s) = \frac{1}{K_D}V_M(s) \quad (9)$$

The output phase is thus proportional to the input modulating signal, scaled by the constant  $1/K_D$ . This is the desired result.

### B. Frequency Modulation

Choose  $P(s) = 1$ , and choose  $Q(s) = K_0K_D/s$ . This is the transfer function of an integrator with the time constant equal to  $1/K_0K_D$ . Then, Eq. (7) becomes

$$\Omega_M(s) = \left\{ [1 - H(s)] + H(s)\frac{s}{K_0K_D}\frac{K_0K_D}{s} \right\} K_0V_M(s) \quad (10)$$

or

$$\Omega_M(s) = K_0V_M(s) \quad (11)$$

The output instantaneous frequency is thus proportional to the input modulating signal, scaled by the constant  $K_0$ . This is the desired result.

## III. Discussion

Equations (9) and (11) demonstrate that the output PM and FM frequency responses are independent of the PLL transfer function  $H(s)$ . Specifically, the detail of the loop filter  $F(s)$  is completely irrelevant as far as the overall modulation process is concerned. To the extent that the linearized model of the PLL is valid, and to the degree that  $P(s)$  and  $Q(s)$  can be realized, the design of the modulator can be completely independent from the PLL closed-loop characteristics; the important parameters, as shown by the analysis, are the phase detector and the VCO gain constants. This independence of the modulation frequency response from the transfer function  $H(s)$  is extremely desirable. Since Eqs. (9) and (11) are implied, the proposed modulator scheme is a wideband configuration and therefore does not suffer from the inherent limitations associated with the configuration of Fig. 1 as mentioned above.

It is possible to obtain wideband response in the configuration of Fig. 1 by introducing the appropriate compensation networks to the input modulating signal prior to the PLL modulation terminals and external to the PLL. These networks would be in the form of  $1/H(s)$  or  $1/[1 - H(s)]$  for PM or FM, respectively. The drawback is that  $H(s)$  must be accurately known. Also, if a high-order loop is required, the resulting high-order compensation network could be difficult to design. Furthermore, the loop would invariably have spurious poles contributed by the loop filter network. This would make extremely difficult the task of fully characterizing  $H(s)$  to the degree required for wideband compensation.

Based on the results of Eqs. (9) and (11) it is possible to propose an entirely new modulator configuration which is shown in Fig. 2. This scheme takes advantage of the feedback loop and applies the modulation to both baseband positions before and after the loop filter such that the modulator design is free from the effect and constraint of  $F(s)$ . The compensation networks are simply the differentiator and the integrator whose time constants are solely determined by the product  $K_0K_D$ . They remain unchanged regardless of the specific detail of  $F(s)$ , and hence the loop order. In particular, as far as the modulator design is concerned, any spurious poles of  $F(s)$  or of any networks in the signal path from the phase detector output to the VCO input are lumped together with  $F(s)$  as part of an equivalent loop filter. This also applies to any time delay functions in that path. However, as will be shown in the experimental results discussed in Section IV, the design is not free of the effect of spurious poles in the loop that are not made part of the loop filter.

This scheme is ideal since  $H(s)$  is self-negating to unity as shown by Eqs. (8) and (10) and obviates the necessity for an accurate external loop compensation to  $H(s)$  or the need to compromise between the PLL and the modulator performance.

Figures 3 and 4 show block diagrams of the proposed phase modulator and frequency modulator. It is evident from the diagrams that the frequency modulator can be derived from the phase modulator by inserting one integrator in each of the phase-modulating signal-injection arms (one of which nullifies the differentiator).

The linearity requirement constrains the peak modulation index seen at the phase detector, hereafter denoted  $\beta_p$ . This means that the sum of the peak phase error in the absence of modulation and  $\beta_p$  must be within the linear range of the phase detector. For a sinusoidal phase detector transfer function, this sum should not exceed 30 deg for good linearity. Also, since the PLL is a carrier tracking loop, the design must prevent the range of modulation index from causing a carrier-null at the phase detector input. The constraint on  $\beta_p$  can be mitigated by judicious choice of the phase compression factor  $N$  in the feedback loop that would permit a broader range of modulation index at the loop output.

Linearity of the integrator and the differentiator also constrains  $\beta_p$ . In the PM case, it can be shown that  $\beta_p$  must satisfy  $\beta_p < K_0 V_{SAT} / (2\pi f_{max})$ , where  $V_{SAT}$  is the saturation voltage of the differentiator and  $f_{max}$  is the highest input modulating frequency. Similarly in the FM case,  $\beta_p$  must satisfy  $\beta_p < V_{SAT} / K_D$ , where  $V_{SAT}$  is the saturation voltage of the integrator (here  $\beta_p$  is the ratio of the maximum frequency deviation and the lowest input modulating frequency).

It would appear from Eq. (11) for the FM case that it would be possible to apply modulation at very low frequency all the way down to DC. Then, one could generate a constant frequency shift at the VCO output by applying a constant DC voltage at the modulating input. However, a DC response is not possible in reality because the integrator (in Fig. 4) would eventually saturate. Secondly, the phase error would eventually exceed the dynamic range of the phase detector and the loop would then no longer remain linear or in lock. However, by careful design to prevent component saturation and to remain within the linear range of the PLL, it is possible to modulate the frequency as close to DC as desired.

## IV. Experiments

A phase modulator implementing the proposed scheme was built to verify the analysis. It was designed to phase modulate a phase-locked signal source whose design emphasis is for phase-noise and spurious signal power spectral density performance. A block diagram of the circuit is shown in Fig. 5. Basically, the loop is a second-order PLL and is of the mix-down-and-divide design for low-noise performance. The VCO is a cavity oscillator operating at a frequency of 2290 MHz. The  $\times 24$  multiplies the 95-MHz phase-locked crystal oscillator output frequency to 2280 MHz. The 2280-MHz signal downconverts the cavity oscillator output at 2290 MHz to an IF of 10 MHz. The  $\div 2$  frequency divides the 10-MHz IF to 5 MHz and effects a phase compression by a factor of 2. A temperature-stabilized low-noise crystal oscillator provides the 5-MHz reference signal. From the phase-noise data of the 95-MHz phase-locked crystal oscillator and the free-running cavity oscillator, the loop parameters were designed for optimum phase noise performance at the output, with the constraint of adequate rejection of the 5-MHz reference signal feed-through. The loop natural frequency was selected to be about 35 kHz and the damping ratio to be 0.707. Measured gain constants of the sinusoid phase detector and the VCO are 0.30 volt/rad and 300 kHz/volt, respectively, which yields the time constant  $1/K_0 K_D$  of 1.77  $\mu$ sec. An op-amp differentiator with a 3-dB bandwidth of about 14 MHz was used. A gain adjustment was included as part of the differentiator to effectively provide the adjustment of the time constant.

To test the circuit, a baseband sinusoidal signal from a signal generator was applied to phase modulate the 2290-MHz carrier output. To assure linearity at the phase detector and not to exceed the linear range of the differentiator at high modulation frequency, the modulation index was kept small, at about 3 deg. The output was then observed on a spectrum analyzer while sweeping the input modulating signal frequency. The spectrum analyzer thus displays the magnitude of the modulation transfer function being measured at the carrier frequency.

Figure 6 shows the result when the modulating signal is applied to the loop filter input only. As would be expected, the output modulation spectrum is symmetrical about the carrier frequency. Note that in this case the display spectrum envelope is proportional to  $|H(jw)|$ . It can be seen that the actual natural frequency is about 40 kHz, and the damping ratio is about 0.50. The 3-dB bandwidth of the loop is about 75 kHz. Figure 7 shows the result when the modulating signal is applied to the differentiator input only; here the spectrum envelope is

proportional to  $|1 - H(j\omega)|$ . These two cases correspond to the configuration in Fig. 1, and the results in Figs. 6 and 7 demonstrate the band-limited nature of that circuit.

Figure 8 shows the spectrum analyzer display when the modulating signal is applied to both inputs of the loop filter and the differentiator. Note that the previous high- and low-frequency roll-offs in Figs. 6 and 7 as well as the amplitude peakings occurring at the loop's natural frequency have disappeared. The flat spectrum indicates that wideband frequency response has been achieved as predicted by the analysis. Figure 9 shows the same results but with the spectrum analyzer set at a 10-MHz span. As can be seen by comparing Figs. 6 and 8 (or 9), the phase modulation transfer function bandwidth has increased by more than an order of magnitude over that of the PLL. Specifically, the baseband modulation 3-dB bandwidth as shown in Fig. 9 is about 1.50 MHz in comparison with the 3-dB bandwidth of only 75 kHz for the loop as shown in Fig. 6. The rolling-off of the output modulation spectrum in Fig. 9 (instead of its remaining flat as it should) was puzzling at first. It was later realized that this is because the driving circuit of the VCO varactor diode as seen from the output of the differentiator network approximates an RC lowpass function with a 3-dB bandwidth of

about 1.50 MHz. This forms a spurious pole as discussed in Section III. Suffice it to say that this part of the circuit must be designed properly or be lumped as part of the loop filter network in order to improve high frequency modulation performance.

## V. Conclusion

A phase-locked angular modulator scheme has been proposed which has the characteristics of wideband modulation frequency response. The modulator design is independent of the PLL closed-loop transfer function  $H(s)$ , thereby allowing independent optimization of the PLL parameters as well as the modulator's parameters. A phase modulator implementing the proposed scheme was built to phase modulate a phase-locked signal source that was designed for low-noise performance at 2290 MHz. The measurement results validated the analysis by demonstrating that the resulting baseband modulation 3-dB bandwidth exceeded that of the PLL by over an order of magnitude. With an additional implementation effort and measurements to improve the high frequency performance, it is expected that flat wideband modulation response from DC well into the MHz regions can be achieved.

## Acknowledgment

The author expresses gratitude to Carl Johns and North Ham for the valuable comments and discussions.

## References

- [1] G. R. Vaughan and E. Osborne, "Phase Locked Phase Modulator," *IEEE International Communications Conference Record, Digest of Papers*, New York, pp. 206-207, 1966.
- [2] F. M. Gardner, *Phaselock Techniques*, New York: John Wiley & Sons, 2nd edition, 1979.

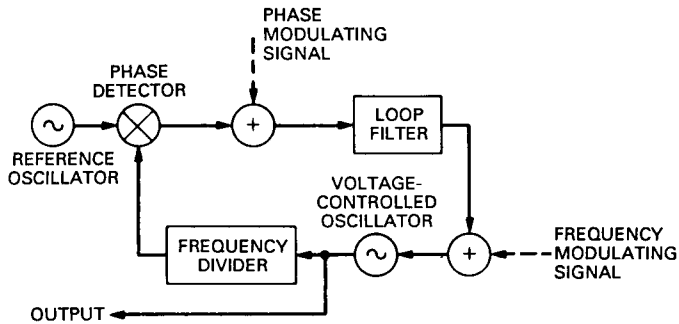


Fig. 1. Typical phase-locked modulator [2].

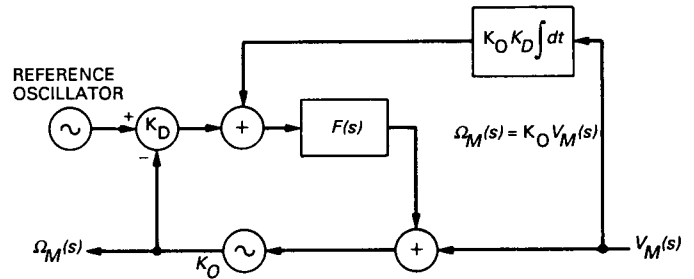


Fig. 4. Proposed wideband frequency modulator.

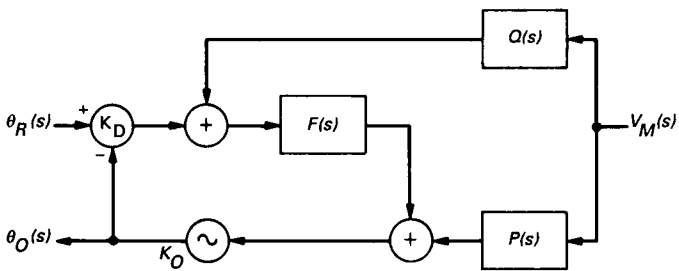


Fig. 2. Proposed phase-locked modulator.

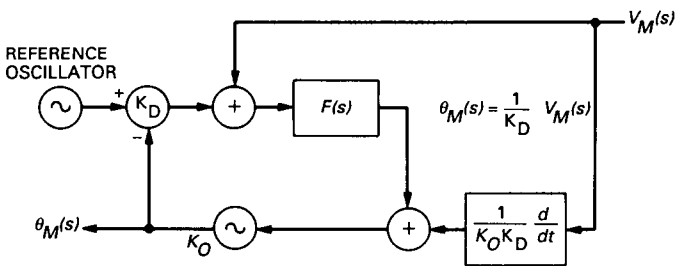


Fig. 3. Proposed wideband phase modulator.

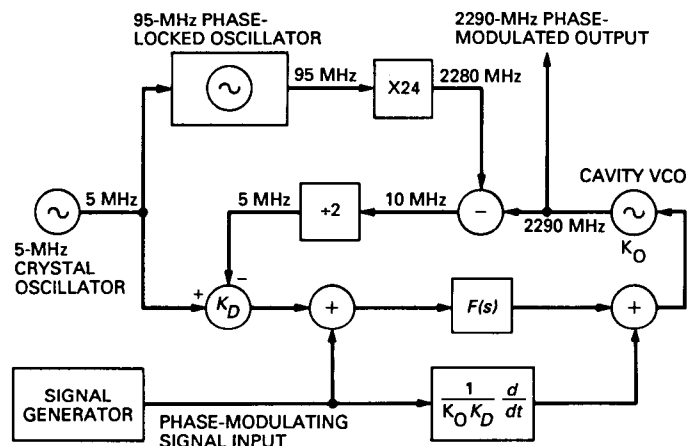
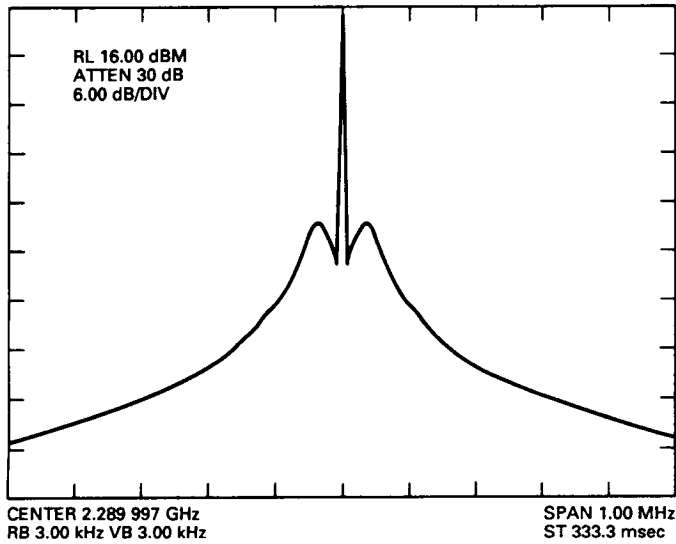
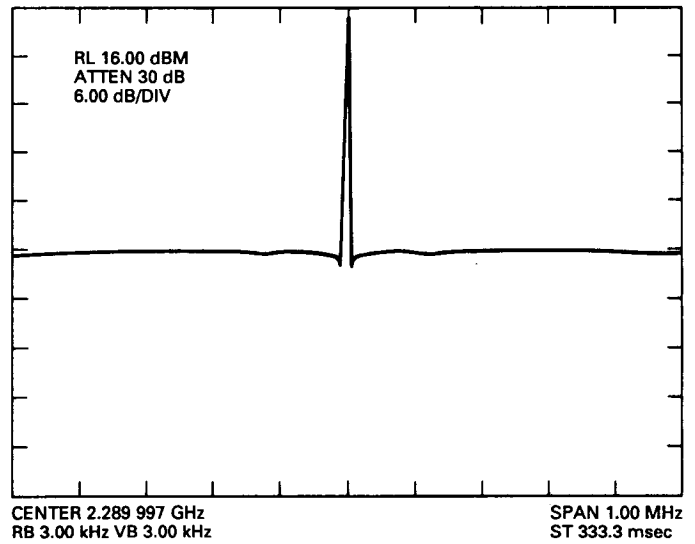


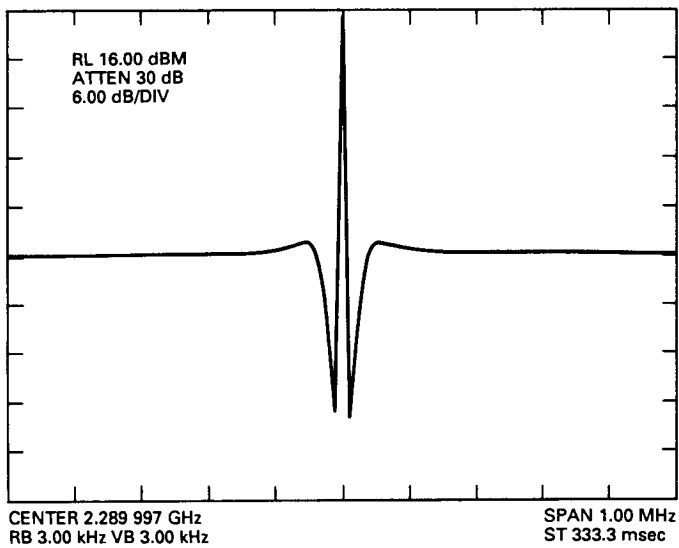
Fig. 5. Experimental wideband phase modulator.



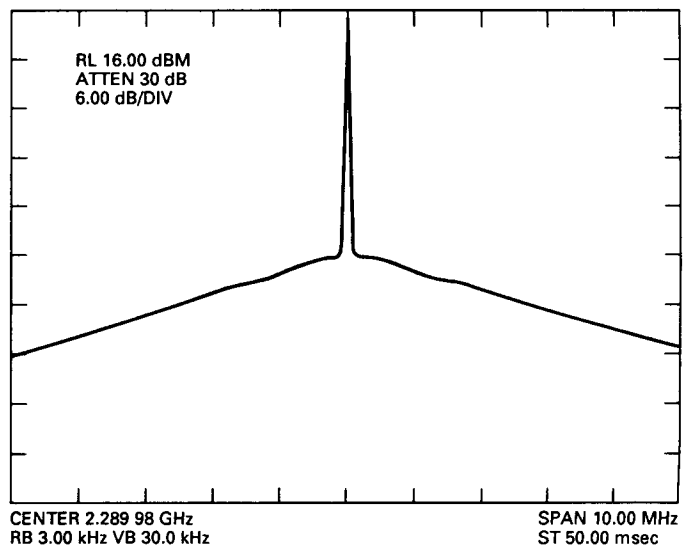
**Fig. 6. Output phase modulation spectrum without compensating differentiator.**



**Fig. 8. Output phase modulation spectrum of the proposed wideband scheme.**



**Fig. 7. Output phase modulation spectrum with compensating differentiator only.**



**Fig. 9. Output phase modulation spectrum of the proposed wideband scheme at 10-MHz span.**



547678  
P18

N90-12799 5/3-32

237823

170

# Performance of the Split-Symbol Moments SNR Estimator in the Presence of Inter-Symbol Interference

B. Shah

Radio Frequency and Microwave Subsystems Section

S. Hinedi

Communications Systems Research Section

*The Split-Symbol Moments Estimator (SSME) is an algorithm that is designed to estimate symbol signal-to-noise ratio (SNR) in the presence of additive white Gaussian noise (AWGN). This article examines the performance of the SSME algorithm in band-limited channels and quantifies the effects of the resulting inter-symbol interference (ISI). All results obtained herein are in closed form and can be easily evaluated numerically for performance prediction purposes. Furthermore, they are validated through digital simulations.*

## I. Introduction

The Deep Space Network Baseband Assembly (BBA) and Advanced Receiver (ARX) use a Split-Symbol Moments Estimator (SSME) algorithm to estimate the symbol signal-to-noise ratio (SNR) of the received signal. The algorithm was designed to operate in the presence of additive white Gaussian noise (AWGN) and at data rates low enough so that bandwidth limiting is insignificant. Its principle of operation relies on using the outputs of two accumulators each operating on a separate half of the same symbol. The product of the outputs provides, after av-

eraging, an estimate of the signal power. An estimate of total power is obtained by summing both accumulations, thus integrating over the full symbol period, squaring the sum, and then averaging to obtain an estimate of signal mean-square plus noise variance. By processing both these averages, an estimate of the symbol SNR can be easily derived.

This article examines the performance of the SSME when the data stream is filtered and inter-symbol interference (ISI) results. This is a likely scenario in the ARX II (current version of ARX) which has a 20-MHz processing

rate and a data rate goal of 6.6 Msymbols/sec. The ratio of the bandwidth  $W$  to the highest data rate ( $R_h$ ) expected is  $W/R_h = 1.33$  assuming an 8.8-MHz 3-dB telemetry bandwidth. Analysis and simulations are used simultaneously to quantify the performance of the SSME in band-limited channels. Throughout this article, perfect carrier, subcarrier, and symbol synchronization are assumed, since the purpose of the analysis is to quantify the effects of filtering only.

## II. The Split-Symbol Moments Estimator (SSME)

The input to the SSME consists of a sequence of samples obtained by processing the received signal. The processing can be broken into two stages: analog and digital. In the ARX II, for instance, the received signal is first downconverted to an intermediate frequency and filtered using analog circuits. This is followed by an analog-to-digital converter after which the carrier, subcarrier, and symbol synchronization are performed digitally. It is assumed that the synchronization function is performed in an ideal fashion and is thus ignored in this article. Figure 1 depicts a basic model for the telemetry signal processing up to the SSME input. The analog filtering function is lumped into a single equivalent filter  $H(s)$  and the digital processing is represented by the filter  $H(z)$ . Given symbol timing, the SSME estimates the symbol signal-to-noise ratio, which is then used to assess the performance of the overall receiver. Quantifying the performance of the SSME, especially at high data rates, enables the user to determine the source of erroneous SNR measurements. These erroneous measurements may be attributed to the fundamental limitations of the SNR estimator or to the various tracking loops not being in-lock. This is particularly important for monitoring the symbol synchronization loop, as the only indication of its proper or improper operation is provided by the estimates of symbol SNR available from the SSME. In order to simplify the problem, it is assumed that the digital filter  $H(z)$  is dominating and the effects of the analog filter are ignored. This is not the case in the ARX II, as both filters contribute equally. However, dealing with one filter will provide some insight into the behavior of the SSME in the presence of ISI.

### A. Signal Model

Assuming that the carrier and subcarrier synchronization have been performed ideally, the baseband signal  $r(t)$  can be modeled as

$$r(t) = \sqrt{(S)}d(t) + n(t) \quad (1)$$

where  $S$  is the data power and  $d(t)$  denotes the actual data given by

$$d(t) = \sum_{n=-\infty}^{\infty} d_n p(t - nT) \quad (2)$$

where  $p(t)$  is the baseband pulse limited to  $T$  sec (data rate  $R = 1/T$ ) and  $d_n$  denotes the binary symbols  $\pm 1$ . The additive noise  $n(t)$  is zero-mean white Gaussian noise with two-sided power spectral density (PSD)  $N_0/2$  watts/Hz. The continuous signal  $r(t)$  is then sampled once every  $T_s$  seconds (sampling rate  $f_s = 1/T_s$ ) to produce the sequence  $r_{ij}$ , as shown in Fig. 2. Consequently, the samples  $r_{ij}$  are given by

$$r_{ij} = \sqrt{(S)}d_{ij} + n_{ij} \quad (3)$$

where  $d_{ij}$  takes on the value  $\pm 1$  with equal likelihood and  $n_{ij}$  is a sequence of independent identically distributed zero mean Gaussian random variables with variance  $N_0/2T_s$ . The double subscript  $ij$  denotes the  $i$ th sample of the  $j$ th symbol. In this case the digital samples  $r_{ij}$  are then filtered by a general filter  $H(z)$  to produce the string of samples  $y_{ij}$ , which form the input to the SSME. The digital filter  $H(z)$  can be either a FIR or an IIR filter. In either case, the analysis to follow is applicable. Assume an integer number  $N_s$  of samples per symbol. This is a valid assumption when the sampling clock is controlled by the symbol synchronization loop to produce an even number of samples per symbol as in ARX I, (the previous version of the advanced receiver) [4]. However, in ARX II this is not the scenario, as the sampling clock is free running and is not synchronized with the symbol rate. As a result, the number of samples per symbol can change and hence, is not fixed. It is not the goal of this article to investigate the effects of noninteger samples per symbol on the performance of the SSME. Thus, only the integer number of samples per symbol case is considered. In writing Eq. (3), it is further assumed that the pulse  $p(t)$  is an NRZ or a Manchester pulse. In either case,  $d_{ij}$  remains constant (1 or  $-1$ ) for  $N_s$  or  $N_s/2$  samples. By including an additional weighting factor, more general pulse shapes can also be accommodated. Note however that when the baseband pulse  $p(t)$  contains nonzero rise or fall times due to analog filtering, the sampling offset (defined as the distance in time between the start of the pulse and the time at which the

first sample is taken) becomes an additional variable and should be considered in a performance prediction analysis.

## B. The SSME Structure

In this section, the SSME is introduced and discussed briefly. Key performance parameters are also defined to evaluate the SSME performance in the presence of ISI. To the authors' knowledge, the SSME was first suggested by L. Howard [1] and later analyzed in white Gaussian noise by Simon and Mileant [2].

The SSME structure is depicted in Fig. 3. The upper "arm" sums the input samples  $y_{ij}$  over the first half of a symbol to produce  $Y_{\alpha j}$ , and the lower "arm" sums the input samples over the second half of a symbol to produce  $Y_{\beta j}$ . Subsequently,  $Y_{\alpha j}$  and  $Y_{\beta j}$  are first summed and squared to produce  $X_{s,sj}$ , an estimate of total power, and secondly multiplied to yield  $X_{pj}$ , an estimate proportional to signal power. Next,  $X_{s,sj}$  and  $X_{pj}$  are averaged over  $n$  symbols to obtain better estimates, namely,  $m_{s,s}$  and  $m_p$ . Finally,  $m_p$  is appropriately weighted and subtracted from  $m_{s,s}$  to obtain an estimate of noise power. This is then used with the signal power estimate to obtain an estimate of the symbol signal-to-noise ratio.

In order to evaluate the performance of the SSME, the SNR should be defined at various points in the system (see Fig. 2), in particular, the symbol SNR before the filter and the symbol SNR at the output of the filter which forms the input to the SSME. Obviously, it is expected that the SNR estimates of the SSME will coincide with the symbol SNR of the filtered signal, as the latter constitutes the input to the SSME. As the data rate is increased and the number of samples per symbol is reduced (assuming an approximately fixed sampling clock), it is expected that the SSME estimates will diverge from the symbol SNR; i.e., there is a data rate threshold above which the SSME breaks down.

The symbol SNR most commonly used in telecommunications is defined as the ratio of symbol energy  $E_s$  to the one-sided PSD level  $N_0$ , i.e.,  $SNR = E_s/N_0$ . For rectangular symbols,  $E_s = ST$ . This can also be expressed in terms of the symbol noise variance as  $SNR = S/2\sigma_n^2$  where  $\sigma_n^2 = N_0/2T$ . In terms of Fig. 2, this is the symbol SNR at the input to the filter. The symbol SNR at the output of the sum- and-dump filter, denoted by  $SNR'$  (' denotes filtered), can be approximated by

$$SNR' \approx \frac{S'}{2\sigma_n'^2} \quad (4)$$

where  $S'$  and  $\sigma_n'^2$  denote the average signal power and the noise variance at the output of the sum-and-dump filter, respectively. Equation (4) is an approximation because it does not include the "self noise" due to ISI (denoted  $\sigma_{ISI}^2$ ) from adjacent symbols. An exact definition of  $SNR'$  would include  $\sigma_{ISI}^2$  in the denominator of Eq. (4). However, in the case of low SNR,  $\sigma_n'^2$  is much larger than  $\sigma_{ISI}^2$ , and the latter can be ignored. On the other hand, at high SNR,  $\sigma_{ISI}^2$  dominates and Eq. (4) is invalid. In DSN applications, the signal is typically dominated by noise and not by ISI, hence, Eq. (4) is valid. Also note that the sum-and-dump filter is not a "true" matched filter. As a result,  $SNR'$  is not optimized. It is shown in Section I of the Appendix that  $SNR'$  as approximated in Eq. (4) can be expressed as

$$SNR' = \frac{S \sum_{k=-\infty}^{\infty} \left( \sum_{i=0}^{N_s-1} h_p(i + kN_s) \right)^2}{2 \sum_{i=0}^{N_s-1} \sum_{m=0}^{N_s-1} R_{n'}(i-m)} \quad (5)$$

where  $h_p(i)$ ,  $i = 0, 1, \dots$ , is the pulse response of the digital filter and  $R_{n'}(k)$  is the autocorrelation function of the filtered noise  $n'(m)$ , defined by  $R_{n'}(k) \triangleq E[n'(m)n'(m+k)]$ . The indices of  $h_p(\ )$  and  $R_{n'}(\ )$  correspond to sampling intervals. The double summation in the denominator accounts for the correlation between the  $i$ th noise sample and the  $m$ th noise sample of the same symbol. The summation over  $k$  in the numerator incorporates the contribution of ISI from all symbols. Strictly speaking, this sum is infinite for an IIR filter and finite for an FIR filter. In either case, the sum can be truncated to include only those symbols whose ISI contribution is significant. In the absence of filtering,  $h(0) = 1$  and  $h(i) = 0$  for  $i \neq 0$ . Consequently, the autocorrelation  $R_{n'}(k) = (N_0/2T_s)[\delta(k) * h(k) * h(k)]$  simplifies to  $(N_0/2T_s)\delta(k)$  and Eq. (5) reduces to  $ST/N_0$ , as expected.

## C. The Performance of the SSME Algorithm

In this section, the performance of the SSME is evaluated in the presence of ISI. All results derived herein reduce to their counterpart for a wideband channel as in [2]. The input samples to the SSME are given by

$$y_{ij} = \sqrt{S} \sum_{m=-\infty}^{\infty} d_{j-m} h_p(i + mN_s) + n'_{ij} \quad (6)$$

where  $n'_{ij}$  are the filtered noise samples. We chose to represent  $y_{ij}$  in terms of the pulse response  $h_p(n)$  to save the additional summation present when written in terms of the impulse response  $h(n)$ . In any case,  $h_p(n)$  is related to the impulse response  $h(n)$  via

$$h_p(n) = \sum_{m=0}^{N_s-1} h(n-m) \quad (7)$$

From Fig. 2, the input samples  $y_{ij}$  are summed over the first half of a symbol to form the random variable  $Y_{\alpha j}$ ; samples in the second half form  $Y_{\beta j}$ , i.e.,

$$Y_{\alpha j} = \sum_{i=0}^{(N_s/2)-1} y_{ij} \quad (8a)$$

$$Y_{\beta j} = \sum_{i=N_s/2}^{N_s-1} y_{ij} \quad (8b)$$

As discussed earlier, we form from the half-symbol accumulators the product and the sum squared of the outputs to get

$$X_{ssj} = (Y_{\alpha j} + Y_{\beta j})^2 \quad (9a)$$

$$X_{pj} = Y_{\beta j} \cdot Y_{\alpha j} \quad (9b)$$

These variables are the estimates of total power and signal power on a per symbol basis. To improve these estimates, we average over  $n$  symbols and obtain

$$m_{ss} = \frac{1}{n} \sum_{j=1}^n X_{ssj} \quad (10a)$$

$$m_p = \frac{1}{n} \sum_{j=1}^n X_{pj} \quad (10b)$$

Note that  $m_{ss}$  and  $m_p$  change at one  $n$ th of the data rate and  $n$  is a design parameter. Finally, an SNR estimate,  $SNR^*$ , is obtained by scaling and combining  $m_p$  and  $m_{ss}$  as follows:

$$SNR^* = \frac{m_p}{2(\frac{1}{4}m_{ss} - m_p)} \quad (11)$$

Equation (11) describes the SSME SNR estimator which has been designed to operate in a wideband channel. Depending on its performance in the presence of ISI, modifications might be required to adopt the estimator for use in band-limited channels.

The mean and variance of  $SNR^*$  have been derived in [2] and are repeated below for convenience,

$$E[SNR^*] = \widehat{SNR} + \frac{1}{n} (1 + 2\widehat{SNR}) \quad (12)$$

$$\text{Var}[SNR^*] = \frac{1}{n} (1 + 4\widehat{SNR} + 2\widehat{SNR}^2) \quad (13)$$

where

$$\widehat{SNR} \triangleq \frac{E[m_p]}{2[\frac{1}{4}E[m_{ss}] - E[m_p]]} \quad (14)$$

Equations (12) through (14) have been computed in the presence of AWGN and assume no ISI. Furthermore, they are approximate expressions which are valid for large  $n$ . Improved approximations could be obtained by retaining higher-order terms of the expansion as discussed in [3]. In the presence of filtering, these equations are not valid in the strict sense but can still serve as a guide, the quality of which is assessed via computer simulations at a later point. They are expected to be "good" approximations because we are operating in a noise-dominant rather than ISI-dominant scenario. It is shown in the Appendix that

$$E[m_p] = S \sum_{m=-\infty}^{\infty} \left[ \sum_{i=0}^{(N_s/2)-1} h_p(i+mN_s) \cdot \sum_{k=N_s/2}^{N_s-1} h_p(k+mN_s) \right] + \sum_{i=0}^{(N_s/2)-1} \sum_{k=N_s/2}^{N_s-1} R_{n'}(i-k) \quad (15a)$$

and

$$E[m_{s,s}] = S \sum_{m=-\infty}^{\infty} \left[ \sum_{i=0}^{N_s-1} h_p(i + mN_s) \right]^2 + \sum_{i=0}^{N_s-1} \sum_{k=0}^{N_s-1} R_{n'}(i-k) \quad (15b)$$

which gives

$$\widehat{SNR} = \frac{S \sum_{m=-\infty}^{\infty} \left[ \sum_{i=0}^{(N_s/2)-1} h_p(i + mN_s) \sum_{k=N_s/2}^{N_s-1} h_p(k + mN_s) \right] + \sum_{i=0}^{(N_s/2)-1} \sum_{k=N_s/2}^{N_s-1} R_{n'}(i,k)}{2 \left[ \frac{1}{4} \left( S \sum_{m=-\infty}^{\infty} \left[ \sum_{i=0}^{N_s-1} h_p(i + mN_s) \right]^2 + \sum_{i=0}^{N_s-1} \sum_{k=0}^{N_s-1} R_{n'}(i,k) \right) - S \sum_{m=-\infty}^{\infty} \left( \sum_{i=0}^{(N_s/2)-1} h_p(i + mN_s) \sum_{k=N_s/2}^{N_s-1} h_p(k + mN_s) \right) - \sum_{i=0}^{(N_s/2)-1} \sum_{k=N_s/2}^{N_s-1} R_{n'}(i,k) \right]} \quad (16)$$

All these equations reduce to their respective counterparts in a wideband channel as in [2]. Namely,  $E[m_p] = N_s^2 S/4$ ,  $E[m_{s,s}] = N_s^2 S + N_s N_0/2T_s$  and  $\widehat{SNR} = ST/N_0$ . Since the estimate  $SNR^*$  itself is a random variable, a measure of its quality is its signal-to-noise ratio given by

$$SNR(SNR^*) = \frac{E^2[SNR^*]}{\text{Var}[SNR^*]} \quad (17a)$$

Using the approximations for  $E[SNR^*]$  (Eq. 12), and  $\text{Var}[SNR^*]$  (Eq. 13), one obtains

$$SNR(SNR^*) = \frac{\left[ \widehat{SNR} + \frac{1}{n} (2\widehat{SNR} + 1) \right]^2}{\frac{1}{n} (1 + 4\widehat{SNR} + 2\widehat{SNR}^2)} \quad (17b)$$

Note that in either limit, as SNR approaches 0 or  $\infty$ , one has

$$\widehat{SNR}_L = \frac{A_L}{2(\frac{1}{4}B_L - A_L)} \quad L = 0, \infty \quad (18)$$

where

$$A_0 = \sum_{i=0}^{(N_s/2)-1} \sum_{k=N_s/2}^{N_s-1} R_{n'}(i-k) \quad (19a)$$

$$B_0 = \sum_{i=0}^{N_s-1} \sum_{k=0}^{N_s-1} R_{n'}(i-k) \quad (19b)$$

$$A_\infty = S \sum_{m=-\infty}^{\infty} \left[ \sum_{i=0}^{(N_s/2)-1} h_p(i + mN_s) \times \sum_{k=N_s/2}^{N_s-1} h_p(k + mN_s) \right] \quad (19c)$$

$$B_\infty = S \sum_{m=-\infty}^{\infty} \left[ \sum_{i=0}^{N_s-1} h_p(i + mN_s) \right]^2 \quad (19d)$$

and

$$\lim_{SNR \rightarrow L} SNR(SNR^*)$$

is given by Eq. (17b) with  $\widehat{SNR}$  replaced by its respective limit. In the absence of filtering,  $A_0 = 0$ ,  $B_0 = (N_s, N_0/2T_s)$ ,  $A_\infty = (N_s/2)^2$  and  $B_\infty = N_s^2$ , which simplifies to

$$\lim_{SNR \rightarrow 0} SNR(SNR^*) = \frac{1}{n} \quad (20a)$$

$$\lim_{SNR \rightarrow \infty} SNR(SNR^*) \simeq \frac{n}{2} + 2 \quad (20b)$$

in agreement with [2]. Equations (18) and (19) depend on the filter  $H(z)$  and cannot, in general, be simplified any further. Intuitively, however,  $SNR(SNR^*)$  is expected to improve with  $n$  for high SNR even in the presence of ISI.

### III. Numerical Results and Discussion

As mentioned earlier, the performance of the SSME needs to be assessed in the presence of ISI for various data rates. The digital filter used in the simulations is the half-band filter described in [5] with transfer function  $H(z)$  given by

$$\begin{aligned} H(z) = & -0.017 + 0.038z^{-2} - 0.089z^{-4} \\ & + 0.312z^{-6} + 0.5z^{-7} + 0.312z^{-8} \\ & - 0.089z^{-10} + 0.038z^{-12} - 0.017z^{-14} \end{aligned} \quad (21)$$

and plotted in Fig. 4 versus normalized frequency. Note that the ratio of 3-dB bandwidth to sampling frequency is about 0.23. Assuming an effective sampling rate of 20 MHz (actual sampling rate is 40 MHz but telemetry processing at the output of the half-band filter is performed at 20 MHz due to decimation by two), the number of samples per symbol is given by  $N_s = f_s T$ . Thus, defining  $N_s$  is equivalent to specifying the data rate (for a fixed sampling rate). To get a feeling of the filtering involved, Fig. 5 depicts the pulse response and Fig. 6 shows a typical filtered sequence for  $N_s = 2, 4, 10$  in the absence of noise. Note from Fig. 5 that a filtered pulse experiences interference from both past and future symbols due to the inherent filter delay.

The performance of the SSME in the presence of filtering is depicted in Fig. 8 versus the number of samples per symbol  $N_s$ . The figure shows analysis and simulation results for input SNRs of  $-5, 0$ , and  $5$  dB. As  $N_s$  decreases, the data rate increases and more ISI results. As  $N_s$  decreases below 6,  $SNR^*$  diverges from  $SNR'$  and the SSME breaks down. Surprisingly, the estimated  $SNR^*$  are several decibels higher than  $SNR'$  in the breakdown region. However, it is unclear whether the divergence of the estimates is due to the ISI or colored noise. In order to gain more insight into the causes, the effects of ISI and colored noise are considered separately. Figures 10 and 12 depict the SSME response when the data or the noise are filtered, but not both. When the data is filtered (Fig. 10), the estimated SNR, denoted by  $SNR'_d$ , is still valid even for  $N_s < 6$ . In this special scenario, Eqs. (5) and (16) reduce to give

$$SNR'_d = \frac{S \sum_{k=-\infty}^{\infty} \left( \sum_{i=0}^{N_s-1} h_p(i + kN_s) \right)^2}{2N_s^2 \sigma_n^2} \quad (22a)$$

and

$$\widehat{SNR}_d = \frac{S \sum_{m=-\infty}^{\infty} \left[ \sum_{i=0}^{(N_s/2)-1} h_p(i + mN_s) \sum_{k=N_s/2}^{N_s-1} h_p(k + mN_s) \right]}{2 \left[ \frac{1}{4} \left( S \sum_{m=-\infty}^{\infty} \left[ \sum_{i=0}^{N_s-1} h_p(i + mN_s) \right]^2 \right) + N_s^2 \sigma_n^2 - \sum_{m=-\infty}^{\infty} \left( \sum_{i=0}^{(N_s/2)-1} h_p(i + mN_s) \sum_{k=N_s/2}^{N_s-1} (k + mN_s) \right) \right]} \quad (22b)$$

The subscript  $d$  is to indicate that only the data is filtered. Both equations are shown in Fig. 10 along with simulation points. As  $N_s$  decreases,  $SNR'_d$  and  $SNR^*_d$  decrease because more signal energy is lost due to filtering. Also note that at  $N_s = 2$ ,  $SNR^*$  and  $SNR'$  coincide better at low SNR and they diverge for high SNR. The curvature of the

graph in Fig. 10 is opposite to that of Fig 8. That tends to indicate that ISI itself is not the reason why  $SNR^*$  of Fig. 8 increases for low  $N_s$ . This phenomenon is better explained by Fig. 12, which depicts the SSME response in the presence of correlated noise, but unfiltered data. In that case, Eqs. (5) and (16) simplify to

$$SNR'_n = \frac{SN_s^2}{2 \sum_{i=0}^{N_s-1} \sum_{m=0}^{N_s-1} R_{n'}(i-m)} \quad (23a)$$

$$\widehat{SNR}_n = \frac{\frac{SN_s^2}{4} + \sum_{i=0}^{(N_s/2)-1} \sum_{k=N_s/2}^{N_s-1} R_{n'}(i-k)}{2 \left[ \frac{1}{4} \left( SN_s^2 + \sum_{i=0}^{N_s-1} \sum_{k=0}^{N_s-1} R_{n'}(i-k) \right) - \frac{SN_s^2}{4} - \sum_{i=0}^{N_s/2} \sum_{k=N_s/2}^{N_s-1} R_{n'}(i-k) \right]} \quad (23b)$$

where the subscript  $n$  indicates that only noise is filtered. For  $N_s < 6$ , the SSME seems to break down due to noise correlation. Recall that the SSME estimates noise power by subtracting signal power estimate from total power estimate. The signal power estimate  $X_{pj}$  (see Fig. 3) is formed by the product of the accumulator outputs. In the absence of filtering, that product contains independent noise samples whose effect is reduced by averaging over  $n$  symbols. In the presence of filtering and specifically for low  $N_s$ , the noise samples in that product are highly correlated and contribute to the signal power estimate even after averaging. That in turn reduces the noise power estimate after subtracting from the estimate of total power, hence significantly increasing  $SNR^*_n$ . Obviously, the noise correlation also increases the total power estimate but that increase is still less than its counterpart in estimating signal power, taking into account the appropriate weighting. In Figs. 13 and 14,  $SNR^*$  is plotted versus  $SNR$  (Fig. 13) and  $SNR'$  (Fig. 14) for a fixed  $N_s$ , when both the data and noise are filtered. As expected, a linear relationship is obtained for large values of  $N_s$  and a nonlinear relationship as the data rate increases.

Finally, as an indication of the quality of the estimates, Figs 15 and 16 depict  $SNR(SNR^*)$  versus  $SNR$  for

a fixed  $N_s$  (Fig.15) and fixed  $n$  (Fig. 16). As more symbols are used to provide an SNR estimate, an improved performance is expected and this is clearly shown in Fig. 15. From Fig. 16, it is also clear that  $SNR(SNR^*)$  improves with decreasing  $N_s$ . This is a rather surprising result at first glance. But remember that  $SNR(SNR^*)$  is an indication of how the estimate jitters about its mean, no matter where the mean is. Also recall that for  $N_s = 4$  the definition of SNR in Eq. (4) is invalid as the variance due to ISI might be significant. The simulations were conducted using 25,000 symbols for  $N_s \geq 10$  and 50,000 symbols for  $N_s < 10$ , with varying  $n$  depending on the SNR.

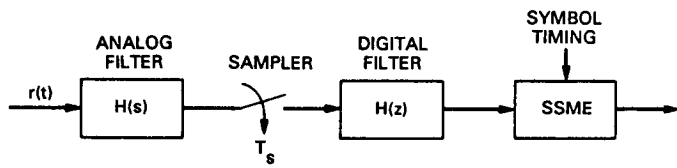
## IV. Conclusion

This article deals with the performance of the SSME in the presence of ISI, evaluated both analytically and by simulation. It is shown that the estimator performs well with more than 6 samples per symbol, but poorly with fewer than 6 samples per symbol due to the high correlation among noise samples of the same symbol. Further work would be required to reduce the bias in the estimates that occur with few samples per symbol.

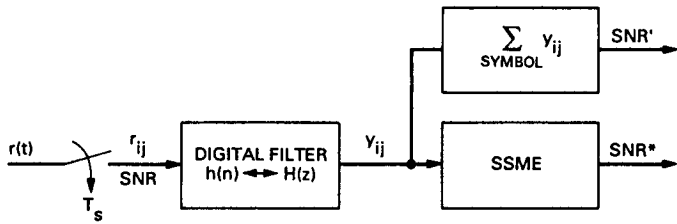
## References

- [1] L. D. Howard, "Signal and Noise Measurements in the NASA Deep Space Network," *20th URSI Conference*, Washington, D.C., August 10-19, 1981.
- [2] M. K. Simon and A. Mileant, "SNR Estimation for the Baseband Assembly," *TDA Progress Report 42-85*, vol. January-March 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 118-126, May 15, 1986.
- [3] M. G. Kindel and A. Stuart, *The Advanced Theory of Statistics*, vol. I, New York: Hafner Publishing Co., 1973.
- [4] D. H. Brown and W. J. Hurd, "DSN Advanced Receiver: Breadboard Description and Test Results," *TDA Progress Report 42-89*, vol. January-March 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 48-66, May 15, 1987.
- [5] R. Sadr and W. J. Hurd, "Digital Carrier Demodulation for the DSN Advanced Receiver," *TDA Progress Report 42-93*, vol. January-March 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 45-63, May 15, 1988.

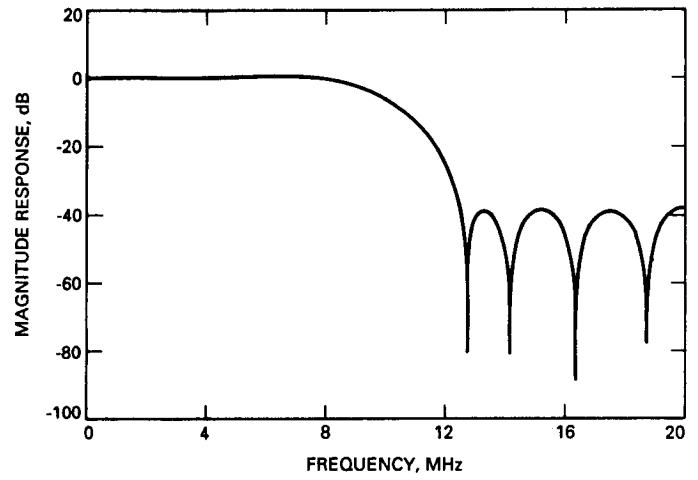




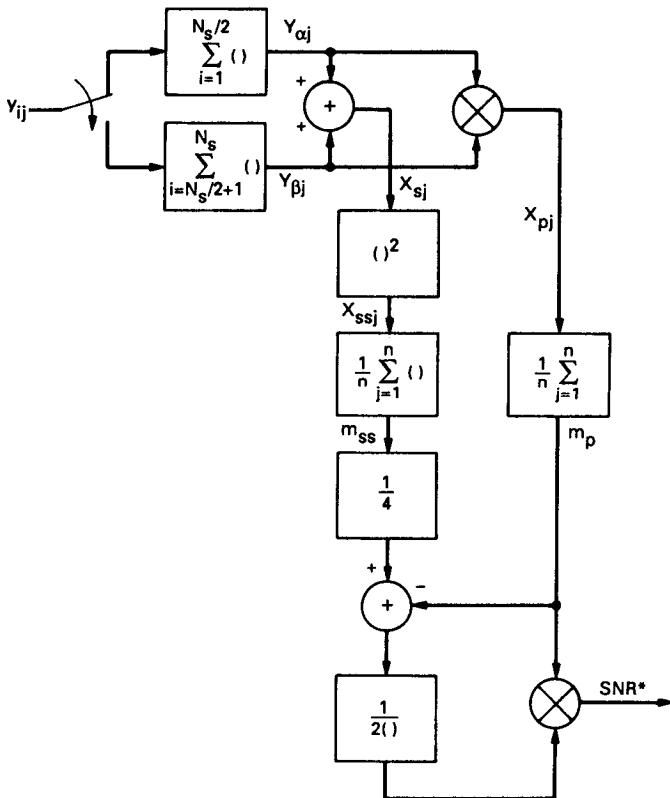
**Fig. 1. General signal processing model.**



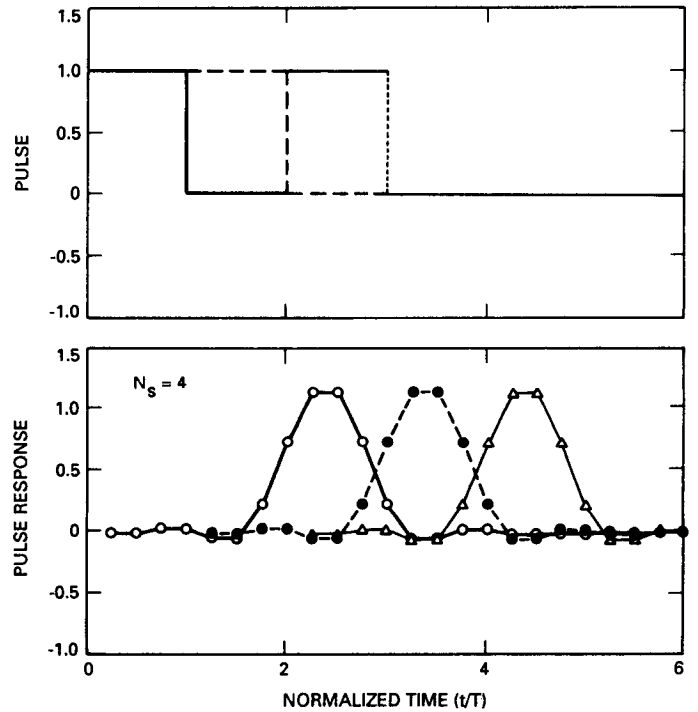
**Fig. 2. Signal model.**



**Fig. 4. Digital filter frequency response (20-MHz processing rate).**



**Fig. 3. The SSME structure.**



**Fig. 5. Pulse response of the half-band filter.**

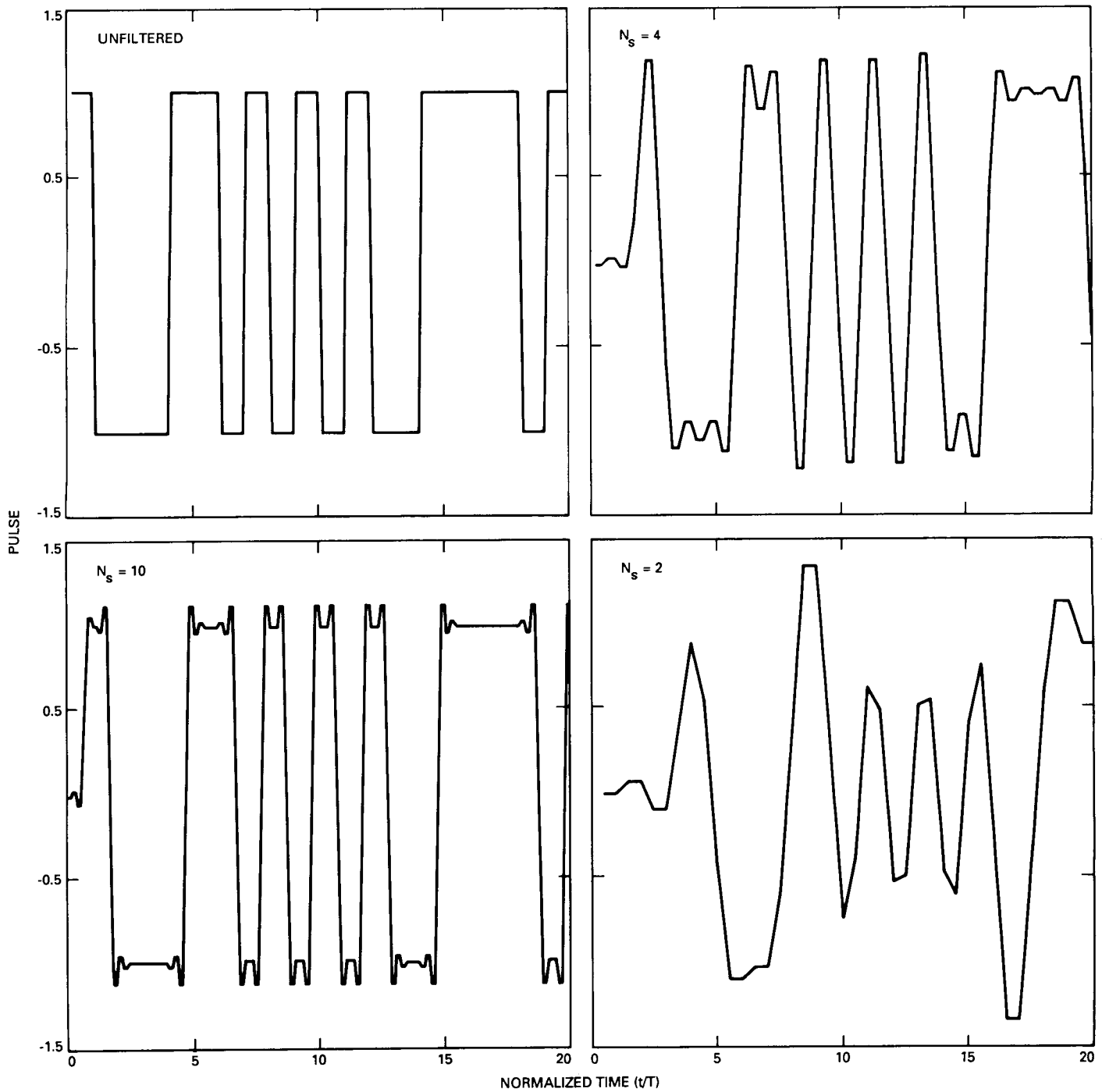


Fig. 6. Filtered data sequences for  $N_s = 2, 4, 10$ .

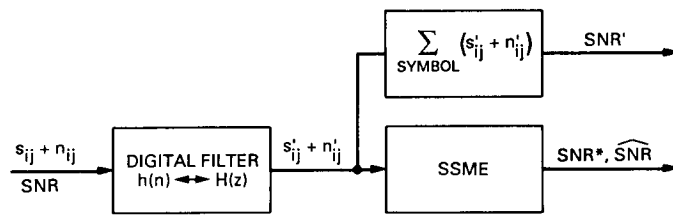


Fig. 7. System model for Fig. 8.

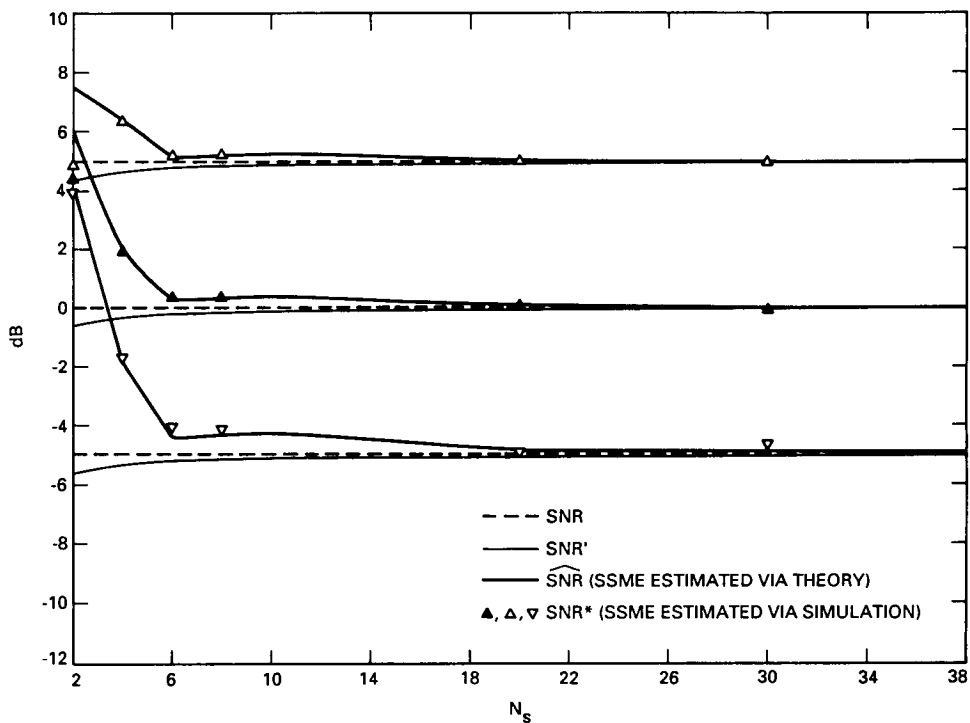


Fig. 8. SNR, SNR', SNR\*, and  $\widehat{SNR}$  versus  $N_s$  (data and noise filtered).

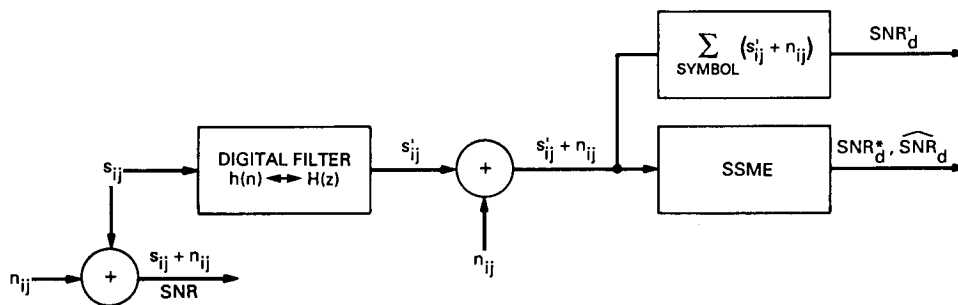


Fig. 9. System model for Fig. 10.

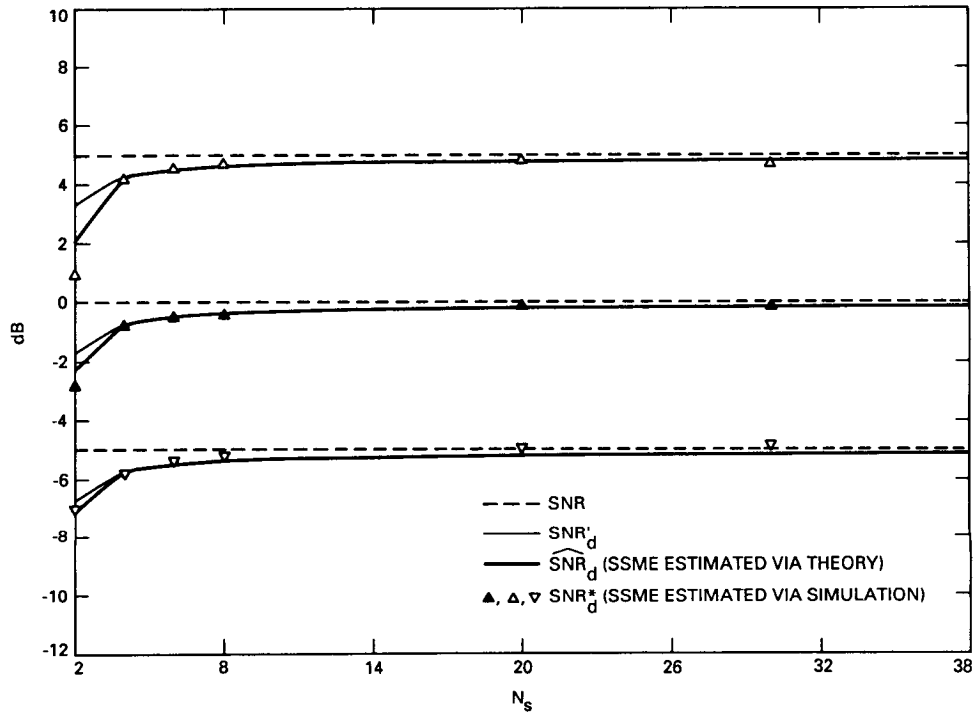


Fig. 10. SNR,  $SNR'_d$  and  $\widehat{SNR}_d$  versus  $N_s$  (data filtered).

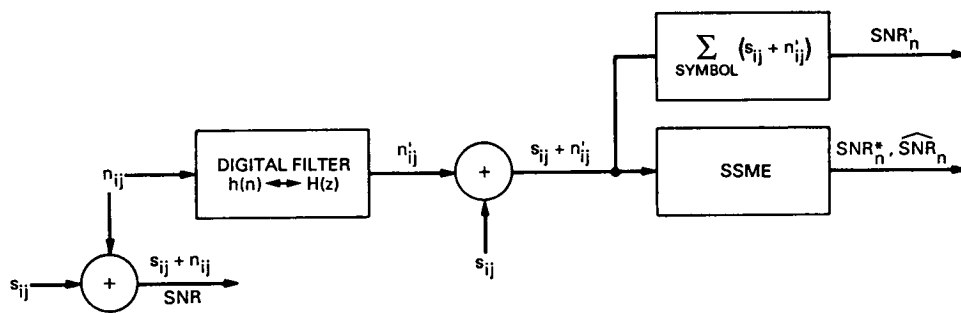


Fig. 11. System model for Fig. 12.

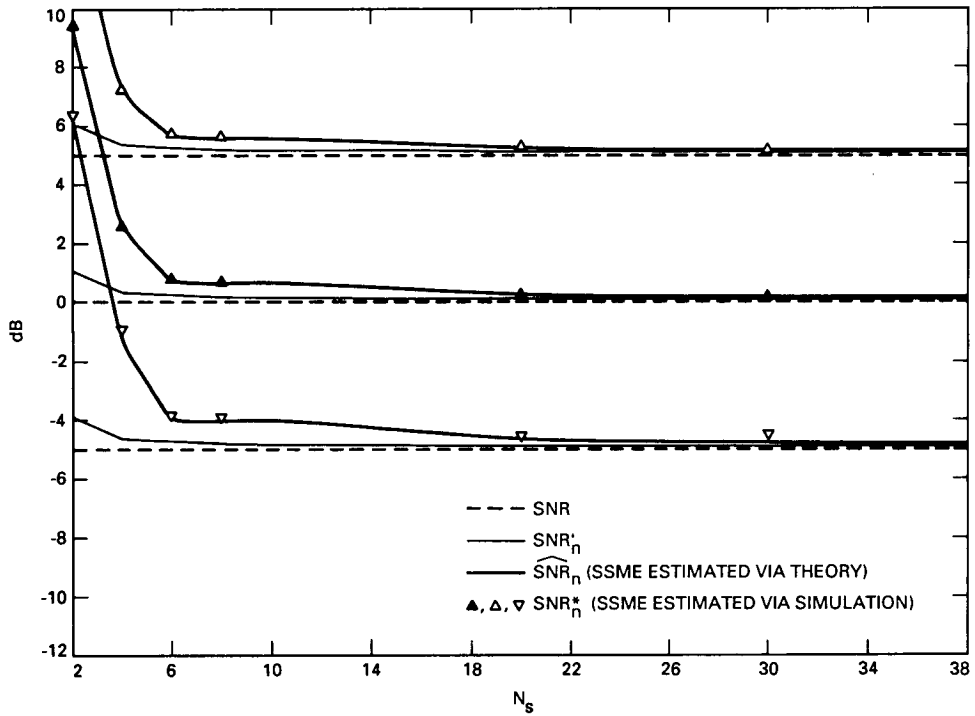


Fig. 12. SNR,  $SNR'_n$ , and  $\widehat{SNR}_n$  versus  $N_s$  (noise filtered).

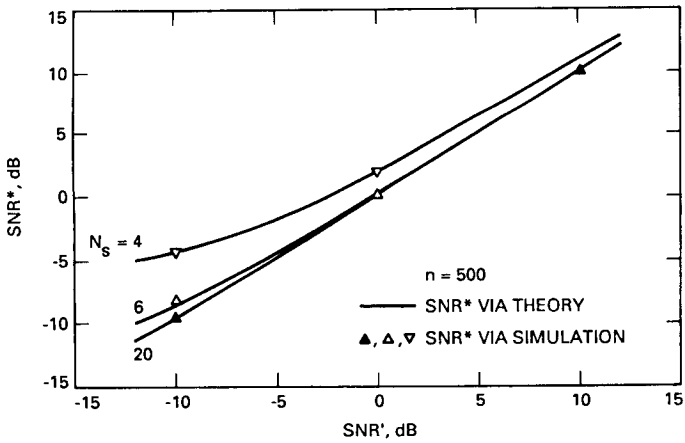


Fig. 13.  $SNR^*$  versus SNR (data and noise filtered).

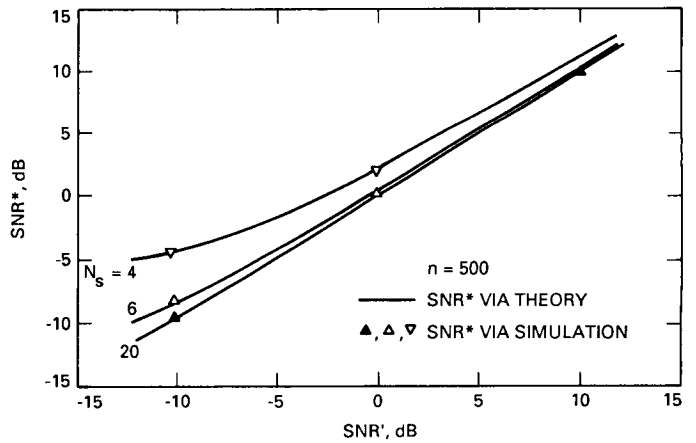


Fig. 14.  $SNR^*$  versus  $SNR'$  (data and noise filtered).

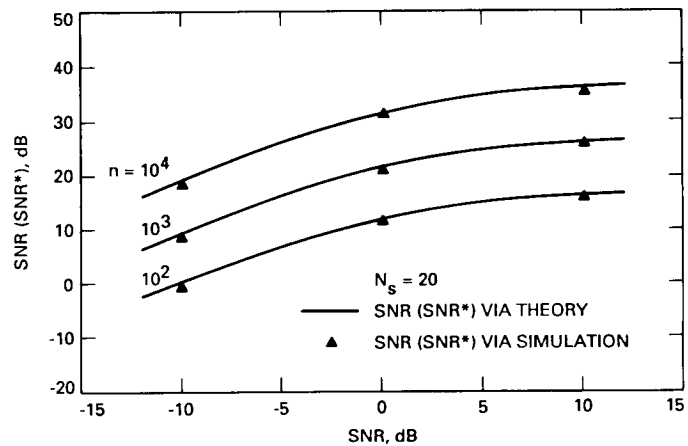


Fig. 15. SNR(SNR\*) versus SNR (fixed  $N_s$ , varying  $n$ ).

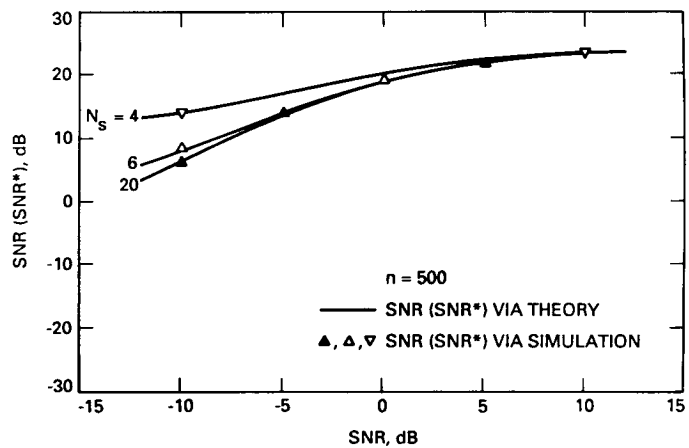


Fig. 16. SNR(SNR\*) versus SNR (varying  $N_s$ , fixed  $n$ ).

# Appendix

## Derivation of Equations

### I. Derivation of Equation (5)

The approximate symbol signal-to-noise ratio at the output of the digital filter,  $SNR'$ , is given by

$$SNR' = \frac{S'}{2\sigma_n^2} \quad (\text{A-1})$$

where  $S'$  and  $\sigma_n^2$ , denote the average signal power and the noise variance respectively at the output of the sum-and-dump filter. We begin our derivation of the signal power in the filtered symbol with an expression for the  $i$ th sample of the  $j$ th filtered symbol,  $y_{ij}$ . Namely,

$$\begin{aligned} y_{ij} &= \sqrt{S} [d_j h_p(i) + d_{j-1} h_p(i + N_s) \\ &\quad + d_{j-2} h_p(i + 2N_s)] + \dots + n'_{ij} \\ &\quad i = 0, 1, 2, \dots, N_s - 1 \\ &= \sqrt{S} \sum_{m=-\infty}^{\infty} d_{j-m} h_p(i + mN_s) + n'_{ij} \\ &\quad i = 0, 1, 2, \dots, N_s - 1 \quad (\text{A-2}) \end{aligned}$$

where  $d_j$  is the  $j$ th received symbol,  $h_p(i)$  is the pulse response at time  $i$ , and  $n'_{ij}$  is the noise sample at the output

of the digital filter. Let  $s'_{ij}$  be the signal portion of  $y_{ij}$ . Clearly then,

$$s'_{ij} = \sqrt{S} \sum_{m=-\infty}^{\infty} d_{j-m} h_p(i + mN_s) \quad i = 0, 1, 2, \dots, N_s - 1 \quad (\text{A-3})$$

and the signal power,  $S'$ , is

$$S' = E \left[ \left( \sum_{i=0}^{N_s-1} s'_{ij} \right)^2 \right] \quad (\text{A-4})$$

where  $E[ \ ]$  is the expected value operator to average over the random data. Substituting Eq. (A-3) into Eq. (A-4) yields

$$S' = E \left[ \left( \sqrt{S} \sum_{i=0}^{N_s-1} \sum_{m=-\infty}^{\infty} d_{j-m} h_p(i + mN_s) \right)^2 \right]$$

Separating the above equation into four sums and moving the expected value operator inside the summations yields:

$$S' = S \sum_{i=0}^{N_s-1} \sum_{\ell=0}^{N_s-1} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} E [d_{j-m} d_{j-n}] h_p(i + mN_s) h_p(\ell + nN_s) = S \sum_{n=-\infty}^{\infty} \left[ \left( \sum_{i=0}^{N_s-1} h_p(i + nN_s) \right)^2 \right] \quad (\text{A-5})$$

where the last step follows since  $E[d_{j-n}d_{j-m}] = \delta(n-m)$  for all  $j$ . Using Eq. (7),  $S'$  can be expressed in terms of the impulse response as

$$S' = S \sum_{n=-\infty}^{\infty} \left[ \left( \sum_{i=0}^{N_s-1} \sum_{m=0}^{N_s-1} h(i+nN_s-m) \right)^2 \right] \quad (\text{A-6})$$

The filtered noise variance  $\sigma_{n'}^2$ , on a per symbol basis, is defined as the expected value of the squared sum of  $N_s$  noise samples, i.e.,

$$\sigma_{n'}^2 = E \left[ \left( \sum_{i=0}^{N_s-1} n'_{ij} \right)^2 \right] = \sum_{i=0}^{N_s-1} \sum_{\ell=0}^{N_s-1} R_{n'}(i-\ell) \quad (\text{A-7})$$

where  $R_{n'}(m)$  is the autocorrelation function of the filtered noise samples. Using Eqs. (A-5) and (A-7) in Eq. (A-1),

$$SNR' = \frac{S \sum_{n=-\infty}^{\infty} \left[ \left( \sum_{i=0}^{N_s-1} h_p(i+nN_s) \right)^2 \right]}{2 \sum_{i=0}^{N_s-1} \sum_{\ell=0}^{N_s-1} R_{n'}(i-\ell)} \quad (\text{A-8})$$

## II. The Expected Value of $X_{ss}$ , $X_p$ , $m_{ss}$ , and $m_p$

As indicated in Fig. 3, the random variable  $X_{ssj}$  is the squared sum of  $N_s$  input samples. Hence, the expected value of  $X_{ssj}$  is

$$E[X_{ssj}] = E \left[ \left( \sum_{i=0}^{N_s-1} y_{ij} \right)^2 \right] \quad (\text{A-9})$$

Substituting for  $y_{ij}$  from Eq. (A-2) in Eq. (A-9) gives

$$E[X_{ssj}] = E \left[ \sqrt{S} \sum_{i=0}^{N_s-1} \left( \sum_{m=-\infty}^{\infty} d_{j-m} h_p(i+mN_s) + n'_{ij} \right) \right]^2 \quad (\text{A-10})$$

Carrying out the square in the above equation, moving the expectation operator through the summations, and noting the following:  $E[a_{j-m}a_{j-\ell}] = \delta(m-\ell)$  for all  $j$ ,  $E[a_{j-m}n'_{ij}] = 0$  for any  $i, j$ , and  $m$ , and  $E[n'_{ij}n'_{kj}] = R_{n'}(i-k)$  for all  $j$ , the desired result is obtained, namely,

$$E[X_{ssj}] = S \sum_{m=-\infty}^{\infty} \left[ \sum_{i=0}^{N_s-1} h_p(i+mN_s) \right]^2 + \sum_{i=0}^{N_s-1} \sum_{k=0}^{N_s-1} R_{n'}(i-k) \quad (\text{A-11})$$

Note that  $E[X_{ssj}]$  is independent of  $j$  and henceforth the subscript  $j$  can be dropped from  $E[X_{ssj}]$  and the expected value of  $X_{ssj}$  can be denoted by  $\bar{X}_{ss}$ . As shown in Fig. 3, the random variable  $X_{pj}$  is the product of the sum of samples from the first half of a symbol with the sum of samples from the second half of a symbol. Consequently, the expected value of  $X_{pj}$  is given as

$$E[X_{pj}] = E \left[ \sum_{i=0}^{(N_s/2)-1} y_{ij} \sum_{k=N_s/2}^{N_s-1} y_{kj} \right] \quad (\text{A-12})$$

Substituting for  $y_{ij}$  from Eq. (A-2),

$$E[X_{pj}] = E \left[ S \sum_{i=0}^{(N_s/2)-1} \left( \sum_{m=-\infty}^{\infty} d_{j-m} h_p(i+mN_s) + n'_{ij} \right) \sum_{k=N_s/2}^{N_s-1} \left( \sum_{\ell=-\infty}^{\infty} d_{j-\ell} h_p(k+\ell N_s) + n'_{kj} \right) \right] \quad (\text{A-13})$$



Expanding the product in the equation above and taking the expectation of the received data symbols, filtered noise,

as well as the product of data and noise yields the desired result:

$$E[X_{pj}] = S \sum_{m=-\infty}^{\infty} \left[ \sum_{i=0}^{(N_s/2)-1} h_p(i + mN_s) \sum_{k=N_s/2}^{N_s-1} h(k + mN_s) \right] + \sum_{i=0}^{(N_s/2)-1} \sum_{k=N_s/2}^{N_s-1} R_{n'}(i - k) \quad (\text{A-14})$$

Note that  $E[X_{pj}]$  is independent of  $j$  and henceforth the subscript  $j$  can be dropped from  $E[X_{pj}]$  and the expected value of  $X_{pj}$  can be denoted by  $\overline{X_p}$ . The random variables  $m_{ss}$  and  $m_p$  are formed by scaling and summing the random variables  $X_{ssj}$  and  $X_{pj}$  over  $n$  symbols. Consequently, the expected values of  $m_{ss}$  and  $m_p$  are the expected values of these scaled sums. But, since  $E[X_{ssj}]$  and

$E[X_{pj}]$  are constant from symbol to symbol (i.e., independent of  $j$ ),

$$E[m_{ss}] = \overline{X_{ss}} \quad (\text{A-15a})$$

and

$$E[m_p] = \overline{X_p} \quad (\text{A-15b})$$

514-74

239824

547679 p16

N90-12800

TDA Progress Report 42-98

August 15, 1989

168

# DSN 70-Meter Antenna Microwave Optics Design and Performance Improvements Part II: Comparison With Measurements

D. A. Bathker

Ground Antenna and Facilities Engineering Section

S. D. Slobin

Telecommunications Systems Section

*This article compares the measured DSN 70-meter antenna performance at S- and X-bands with the design expectations presented in Part I of this report. A discussion of natural radio-source calibration standards is given. New estimates of DSN 64-meter antenna performance are given, based on improved values of calibration source flux and size correction. A comparison of the 64- and 70-meter performances shows that average S-band peak gain improvement is 1.94 dB, compared with a design expectation of 1.77 dB. At X-band, the average peak gain improvement is 2.12 dB, compared with the (coincidentally similar) design expectation of 1.77 dB. The average measured 70-meter S-band peak gain exceeds the nominal design-expected gain by 0.02 dB; the average measured 70-meter X-band peak gain is 0.14 dB below the nominal design-expected gain.*

## I. Introduction

This article is the second part of a two-part report [1] detailing the design performance expectations and comparison with actual performance of the DSN 70-meter antennas in their upgrade from the original 64-meter configuration. The actual 70-meter antenna gain and noise temperature performance is documented in [2, 3].

Two of the important elements of the 64-70 meter antenna performance upgrade project were (1) to determine

if the project specifications were met and (2) to quantify, with an estimated absolute accuracy, the actual efficiency/gain levels achieved. The project specifications of 1.9-dB X-band and 1.4-dB S-band gain increases are "improvement" specifications and do not effectively quantify absolute performance levels; there remains the issue of quantitatively defining the baseline performance level of the 64-m antennas. In this situation, the DSN had three differently performing 64-m antennas [4-6]. For example, it was clear (but not specifically measured) that the Goldstone 64-m antenna had reflecting panels which

were manufactured and/or adjusted better than those on the Madrid and Canberra 64-m antennas. Also, the Goldstone and Canberra 64-m antennas had been previously upgraded with a gravity-distortion-reducing major structural brace system which generally performed successfully, but was not accompanied with the exact required sub-reflector active focus command programs. The Madrid 64-m antenna was further unique in that the structural brace upgrade was installed but certain superfluous older bracing was intentionally not removed due to time constraints. This combination was known to be counter-productive. Because of noncritical tracking requirements, this was an acceptable configuration for a short interim period prior to planned 70-m upgrade but, in fact, it produced an antenna system with degraded shape and nonoptimum sub-reflector focus performance. In short, the 64-70 meter upgrade project began with three quite differently performing "baseline" 64-m antennas, and these needed to be carefully calibrated to assist settling project baseline definition questions.

## II. Natural Radio Source Calibration Standards

DSN antenna calibrations are wholly dependent upon viewing natural radio noise sources (radio galaxies, quasars) with estimated microwave flux densities. Obtaining such calibrations by the alternative of overall deep space telecommunication link performance has proven a more difficult matter because of the great number of variables involved. The absolute accuracy of natural radio-source flux densities has been estimated to be 0.50 dB, ( $3\text{-}\sigma$ ) at X-band [4]. It was believed even with the difficult absolute accuracy issue at the half-dB level, at least very accurate comparison measurements would ensue, and the 70-m calibrations would be no worse on an absolute basis than the 64-m calibrations.

The 70-m antennas provided more detailed radio-source observation by virtue of much narrower antenna X-band beamwidth (0.0305 deg versus 0.036 deg). Narrower beamwidth results from both larger diameter and substantially higher area efficiency. Various sources which formerly appeared consistent in strength on the 64-m and smaller DSN antennas immediately appeared with inconsistent apparent strengths on the new 70-m antennas—an indication of source size/beam width (source resolution)

increase. Although never an issue at S-band, the DSN radio source list "standards" document<sup>1</sup> in fact was shown inadequate at X-band for the very narrowbeam 70-m. The approximately 15-percent narrower X-band 70-m beamwidth is about 0.0305 deg or 1.83 arcmin. This proved to better resolve the primary calibrator 3C274 and reveal the previous inaccuracies, especially in the source size/beamwidth correction factor, but to a lesser degree in the absolute flux scale as well.

Table 1 provides a summary of data corrections needed when using natural radio sources with these extremely large, high-efficiency X-band antennas. The resolution provided by significantly sharper (high-efficiency) 70-m beamwidth, together with careful analytical work at JPL (M. Klein and P. Richter, private communication), revises the primary and secondary DSN calibrator sources, as detailed in Table 1. The effects on gain are shown in parentheses.

Updated Table 1 values, and others, will be published in the Revision C version of JPL Document D-3801. Accordingly, within this article, previously published [4-6] 64-m X-band results are rescaled according to Table 1 values. The 70-m X-band results herein are also based on Table 1. At S-band, due to wider beamwidth, no noticeable source size/beamwidth or flux effects were expected or determined. At S-band, previous DSN calibrations with National Bureau of Standards (now NIST) certified standards have proven that natural radio-source S-band flux values are known to 0.22 dB,  $3\text{-}\sigma$  [7].

A rigorous analysis of absolute accuracy presently attained in DSN antenna gain measurements using natural sources is not available. The three dominant error contributors are natural source flux, instrumentation accuracy, and data repeatability. Tests for and corrections of these errors are applied in data reduction for pointing errors as needed and receiver linearity. It is believed the S-band flux accuracy is 0.22 dB and that the X-band accuracy is improved, due to the 70-m narrowbeam experience, to not worse than 0.40 dB, both numbers quoted at high ( $3\text{-}\sigma$ ) confidence. It is believed that the instrumentation accuracy is 0.15 dB, and that data repeatability is generally 0.09 dB, both  $3\text{-}\sigma$ . Taking these dominant error sources

---

<sup>1</sup> *DSN Radio Source List for Antenna Calibration*, JPL D-3801, Rev. B (internal document), Jet Propulsion Laboratory, Pasadena, California, September 25, 1987.

as independent, it is reasonable to accept overall absolute accuracies of 0.28 dB (S-band) and 0.44 dB (X-band),  $3\text{-}\sigma$ , for normally appearing data sets. At both frequency bands, it is significant that the natural-source flux uncertainty is the dominant component of the total measurement uncertainty.

### III. Adjustment of 64-m Antenna Calibrations

All three 64-m antennas were calibrated at S- and X-bands within a month or two prior to removal from service for 70-m conversion. DSS-63 was calibrated in July 1986; DSS-43 in January 1987; and DSS-14 in September 1987. The JPL DSN radio-source flux standard D-3801, Rev. A, was in effect throughout the DSS-43/63 period, and Rev. B during the DSS-14 period. Consistency in the use of equipment, technique, and standards was sought and is believed to have been achieved. The results are initially reported in [4-6].

In view of the natural radio source size and flux corrections, the 64-m calibrations are now adjusted in order to allow accurate comparison with the 70-m results. DSS-14 64-m X-band results are increased +0.313 dB, changing the X-band peak efficiency and gain values in [6] from 0.498 (+72.01 dBi) to 0.535 (+72.33 dBi). DSS-43 is adjusted +0.329 dB, changing the X-band peak efficiency and gain values in [5] from 0.454 (+71.61 dBi) to 0.490 (+71.94 dBi). DSS-63 is also adjusted +0.329 dB, changing the X-band peak values in [4] from 0.451 (+71.58 dBi) to 0.487 (+71.91 dBi). All X-band results apply at a frequency of 8420 MHz, and the peak-to-peak spread, 0.42 dB, was rather large.

At S-band, the 64-m antennas yielded an average of 58.2 (+1.2/ - 0.8)-percent area efficiencies (61.36 dBi at the test frequency of 2285 MHz, or 61.40 dBi adjusted to 2295 MHz for comparison with 70-m results). DSS-14 peaked at 59.4 percent, DSS-43 at 57.7 percent, and DSS-63 at 57.4 percent, a peak-to-peak spread of 0.15 dB.

### IV. 64-m Calibrations Compared With Design Expectations

Figure 1 shows the 64-m antenna X-band gain with elevation angle response, adjusted as described above. Both

DSS-14 and DSS-43 exhibited generally flat response, not unlike the design-expected shape, albeit everywhere below the expectation. DSS-63 clearly exhibited the interim dual-brace arrangement, known to degrade antenna gain response away from the setting angle. The reader is reminded that the 64-m design-expected peak value is based on 1.14-mm rms (normal) small-scale surface roughness. This was the original design value and was likely never achieved. Limited holographic imaging on the DSS-63 64-m antenna indicates the small-scale effective roughness was near 1.7-mm rms, consistent with the actual performance described in [4], although the DSS-14 64-m antenna was clearly better, perhaps 1.35-mm normal rms.

Two X-band design expectations were discussed in Part I [1] of this article. The first is considered a "nominal" expectation, in terms of likelihood, and the second is considered an "estimated minimum-likely" value. Both are highly dependent on the small-scale roughness term, which was poorly known for the 64-m antennas. The procedure adopted to form the estimated minimum-likely value, as discussed, takes one-third of the estimated peak limits. This is equivalent to the nominal estimate, minus approximately  $1\text{-}\sigma$ . In the 64-m case, this is believed to lead to overestimating the real performance to be expected. If high-resolution holographic imaging of all 64-m antennas was available, the estimates could be improved; however, such imaging is not available on all those antennas. Accordingly, in the absence of further quantitative information, the original design rms, the nominal design-expected, and the estimated minimum-likely 64-m efficiency values must be accepted. Comparisons with measurements are shown in Table 2.

The S-band average measured difference from estimated minimum-likely is 0.07 dB. The X-band difference is 0.05 dB at DSS-14 and about 0.38 dB at DSS-43 and DSS-63. Clearly, DSS-43 and DSS-63 had surfaces whose roughness was larger than both the original specification and the one-third peak value adopted in the minimum-likely estimating procedure. It requires 1.24 dB of small-scale roughness loss to reconcile DSS-43/63 X-band 64-m actual performance with expectations. This indicates an effective normal surface roughness of 1.6 mm (0.063 inch), not an unreasonable value. At S-band, it should be noted that the DSS-14 64-m antenna in particular met and appeared to slightly exceed the estimated minimum-likely efficiency (0.594 versus 0.591).

## V. 70-m Antenna Calibrations

Immediately upon conclusion of final upgrade structural work, the 70-m antenna calibrations commenced. Inclement weather at DSS-63, the interim fiberglass subreflector at DSS-43, and other factors delayed completion until August 1988. Early informal performance reportings, especially for X-band, were gauged against the then-existing D-3801, Rev. B flux standard document, even though the very narrowbeam 70-m antennas immediately revealed inconsistencies among the X-band calibration sources. In the following discussion, the revised flux and source size/beamwidth corrections discussed earlier are properly used. The 70-m X-band peak efficiencies and gains as reported in [2] are: DSS-14, 68.4 percent (74.17 dBi); DSS-43, 67.3 percent (74.10 dBi); and DSS-63, 70.1 percent (74.28 dBi). These values are more closely grouped than the three 64-m antenna X-band performances were. The DSS-14 peak occurs at 46-deg elevation, DSS-43 at 46-deg elevation, and DSS-63 at 47-deg elevation.<sup>2</sup> The standard deviation of DSS-14 efficiency data (318 points) is 0.7 percent, or 0.04 dB. The DSS-43 data standard deviation (540 points) is likewise 0.7 percent. The DSS-63 standard deviation (176 points) is 1.0 percent, or 0.06 dB.

The 70-m X-band gains as functions of elevation angle are shown in Fig. 2. The design-expectation curve has a peak value as determined in Table 4 of [1] and a shape determined from GTD calculations, which consider main reflector structural deformation due to gravity loading. The DSS-14 elevation function is seen to follow the design-expected elevation function, albeit 0.15 dB low at the adjustment angle and less than 0.2 dB low from 0 to 80-deg elevation. Both DSS-43 and DSS-63 exhibit elevation functions which are considerably steeper than design expectations.

In view of the very small standard deviations in the measured data, the measured X-band elevation functions are judged statistically significant and indicate a real difference in structural or focus performance among the 70-m antennas. It should be noted in Fig. 2 that small atmospheric loss terms have been removed from the observed data, according to on-site meteorological measurements taken during the tests. In short, the data in Fig. 2 are

<sup>2</sup> The main reflector surfaces were adjusted based on holographic imaging as follows: DSS-14 at 47.0 deg, April 1988; DSS-43 at 47.3 deg, June 1988; and DSS-63 at 41.8 deg, July 1987.

as self-consistent as can be expected, and the DSS-43 and DSS-63 antennas indeed behave alike and somewhat differently with elevation angle than does the DSS-14 antenna.

At S-band, limited 70-m data were collected, in favor of the more revealing X-band data. At S-band, the peak efficiency is 0.761 (+63.34 dBi at 2295 MHz), and the elevation response is essentially flat, as expected. Figure 3 shows the average result from two antennas, DSS-14 and DSS-43. The DSS-63 S-band 70-m data were self-inconsistent for reasons not understood, therefore judged invalid. Operationally, it is not expected that DSS-63 will perform differently from the other stations at S-band, and indeed no significant difference has been observed.

## VI. 70-m Calibrations Compared With Design Expectations

Table 3 summarizes the nominal design-expected and estimated minimum-likely 70-m efficiency values and measurements. The "best" X-band antenna, DSS-63 with 0.701 efficiency, is 0.09 dB above the measurement average while the "worst" antenna at 0.673 efficiency is 0.08 dB below the measurement average. In reality, these are exceedingly small differences, and the qualitative descriptions above should not be interpreted too literally. At X-band, all antennas are below the nominal design expectation by an average of 0.14 dB. The S-band average is 0.02 dB above the nominal design expectation.

The improved agreement in 70-m actual performance against design expectations is directly traceable to improved knowledge gained through use of high-lateral-resolution microwave holographic imaging. The holography illumination amplitude image gives a clear view of quadripod shadowing while the phase image gives an excellent estimate of small-scale surface roughness. All 70-m antennas yielded high-quality holographic data and the nearly 1,300 individual reflector panels on each antenna were adjusted according to those data.

In actuality, all 70-m antenna surface (and focus) adjustments were seriously time-compromised. The high-resolution holographic data type inherently provides panel-setting data based on individual-panel 3-degree-of-freedom best fits, within a global best-fit reference system. Rather precise information (approximately 50 microns or

0.002-inch repeatability) for each panel corner is provided. In the time-constrained 70-m project a compromise was necessary. At each intersection of four panel corners, the necessary individual settings, which are in general different, were simply averaged. Thus, all panels are, in general, misset and in fact very slightly twisted. Further, panel corners in *average* error by 0.015 inch (0.37 mm) or less were not adjusted at all. This is clear from the resultant 0.7-mm rms small-scale roughness achieved compared with the initial Project goal of 0.5-mm rms, and further compared to the holography-determined theoretical minimum (given perfect adherence to setting directions) of 0.3-mm rms. An rms of 0.3 mm is effectively the manufacturing precision of the panels and subreflector, as independently determined during manufacture, plus a small incoherent holography data-system noise component. If time had permitted refinement of the 0.7-mm quick setting to a practical 0.4 mm, the X-band gain would increase 0.13 dB, changing the average measured 68.6 percent efficiency to 70.7 percent. At a higher frequency such as 32 GHz, the gain difference developed by improving a 0.7-mm surface to 0.4 mm is nearly 1.9 dB.

## VII. Gain Performance Increases

### A. S-Band

Tables 1 and 2 of [1] give the design-expected S-band gain increase (70-m over 64-m) as +1.77 dB. From measurements, the efficiencies determined are 0.582 (+61.40 dBi at 2295 MHz) for the 64-m antennas and 0.761 (+63.34 dBi) for the 70-m antennas. The measured increase is thus +1.94 dB. Both 70-m measurement sets show very small differences between the antennas. The Project specification (+1.4 dB) was, of course, based on predesign estimates. Detailed design extracted 0.5 dB more gain than initial estimates for the 70-m antennas.

### B. X-Band

Tables 3 and 4 of [1] give the design-expected X-band gain increase (70-m over 64-m) as +1.77 dB. However, this is based on the originally specified 64-m surface tolerance of 1.14 mm which is unrealistically low, as discussed before. Likewise, the estimated minimum-likely efficiencies in Table 6 of [1] for 64-m X-band (0.533 for DSS-43/63 and 0.541

for DSS-14) are considered too high, due to the adopted one-third peak estimating method used. The measurements (simply averaged) show +74.18 dBi for 70-m and +72.06 dBi for 64-m antennas. On this simple averaged basis, the 70-m upgrade project provided +2.12 dB X-band gain improvement at or near the adjustment (peak gain) elevation angle. The best 64-m antenna, DSS-14, provided +72.33 dBi. After upgrade to 70-m, DSS-14 now provides +74.17 dBi. On this single antenna basis, the 70-m upgrade project provided +1.84 dB at DSS-14, the minimum improvement realized at X-band. For DSS-43 and DSS-63, the average X-band gain increase provided by the 70-m project is +2.27 dB. These improvements clearly exceed expectations simply because the 64-m surface tolerance was somewhat worse than our limited previous knowledge and estimates.

Table 4 presents a summary of the S- and X-band 64- and 70-meter antenna gains and efficiencies, plus the increases gained by the upgrade activities. Each reader can (and perhaps will) select his or her baseline of choice in this matter of three quite differently performing 64-m X-band antennas at the start of the project, and three very slightly differently performing 70-m X-band antennas at project conclusion. Nevertheless, perhaps the best value to use is the simple average of the before-to-after measures, that is, +2.1 dB at X-band. The specification (+1.9 dB) was, of course, based on predesign estimates. Detailed design extracted a bit less than the predesign X-band estimate from DSS-14. The necessary stovepipe feed and surface setting compromises, both of which are recoverable, would boost DSS-14 improvement from +1.84 to +2.08 dB with, of course, similar increases to the other 70-m antennas.

Figure 4 presents 64-m and 70-m measured X-band data for comparison. Included in Fig. 4 are the 64-m and 70-m design expectations (nominal, not estimated minimum-likely values). The peak design expectations are obtained directly from Tables 3 and 4 of [1]. The design-expected elevation functions are obtained from finite-element structural gravity distortion data, analytically smoothed for input to the GTD scattering program [8-10]. After smoothed GTD scattering a small-scale (nominally random) roughness term is added for the overall design estimate. Thus the large lateral-scale distortions are handled by direct integration, while the small-scale distortion is handled on the familiar statistical basis, according to the commonly used Ruze formulation [11]. This is the best technique known and possible at this time. Apply-

ing the simplified statistical method of Ruze is considered quite proper in this instance of 1,300 individual reflecting panels per 70-m antenna.

In Fig. 4 it is clear that the 64-m and 70-m design expectations compared with DSS-14 data are in general agreement. The DSS-63 64-m antenna was not expected to follow the design elevation function due to the temporary dual structural bracing configuration as discussed.

Both DSS-43 and DSS-63 70-m antennas do not presently follow the design-expected elevation function. Whether this is due to a real structural deficiency, a subreflector focus control deficiency, poor pointing during data collection, or other reasons is not known. The indications are that pointing was acceptable. Additional work is necessary to understand the DSS-43/63 elevation functions, and focus tests at high and low elevation angles are the likely priority. Separately, some very valuable holography data was obtained at low elevation angles at DSS-14 and to a lesser extent at DSS-63. These data, with a lateral resolution comparable with structural rib and hoop dimensions, may assist in detailed understanding of the gravity behavior of the DSN 70-m antenna structures.

## VIII. System Noise Temperatures

No formal predictions of 64- to 70-m antenna upgrade noise improvement (or degradation) were made. The rear spillover noise, a component of the total, was expected to increase roughly 1 K at S-band, a consequence of the X-band optimum design (the former 64-m systems were S-band optimized). At X-band, the rear spillover was expected to decrease roughly 0.5 K. The deeper main reflector resulting from 64- to 70-m diameter increase, with approximately the same effective focal length, could be expected to slightly reduce noise with the 70-m antenna pointing below about 30-deg elevation and at zenith. The reduced quadripod shadowing might provide slightly reduced noise at all elevation angles. An upper bound on this might be 1 K.<sup>3</sup>

<sup>3</sup> The 64-m systems utilized a special beamshaping flange around the periphery of the hyperboloidal portion of the subreflector, which was optimized for S-band. In any event, such appendages tend to be narrowband on a scale of an octave, thus of limited utility for wideband antennas.

Despite the above 70-m design tendencies, the JPL Technical Division and Project position was that the system noise temperature change, if any, was too close to quantify without significant additional work. This section will therefore briefly record measured noise results. The results presented here are slight modifications of the initial results presented in [3]. During the period of mechanical upgrade, some of the maser preamplifiers were serviced, and in some cases, outright exchanges were made. With at least twelve masers disposed among three stations, such servicings and exchanges require subcalibrations and become a difficult measurement, bookkeeping, and reporting task. Also during the period of mechanical upgrade, the X/S-band dichroic plates were serviced and upgraded. The paint in some cases appeared seriously checked and badly flaking from over 10 years exposure to weather. All three dichroic plates were stripped of paint and the aluminum alloy base metal was irridited. The irridite process is a thin film and considered sufficient weather protection for several years. It is unknown how much steady-state (dry) noise degradation was being experienced with the paint seriously flaking. The expectation is that at least following rain, the recovery time to normal system noise temperature will be considerably improved due to more rapid drying of the irridited noncapillary surfaces.

In summary, one might expect the S-band zenith noise to remain roughly the same and expect a small X-band noise zenith decrease. X-band noise recovery following rainy conditions should improve. Table 5 summarizes available measured zenith noise temperatures for 64- and 70-m antennas at S- and X-bands. From the measured values, composite S- and X-band noise temperature models were created, as described in [3]. The table is divided into two parts: measured totals (including atmosphere) and those values inferred for no atmosphere. The no-atmosphere data is based on local meteorological conditions existing at the time of the measurements and should be the more reliable (weather-independent) data for comparison purposes.

From Table 5, the DSS 63/64-m S-band noise appears abnormally high, and remains so after 70-m conversion. The  $20.7 \pm 0.6$ -K DSS-14/43 S-band 64-m average (with atmosphere) is reasonable. After 70-m conversion, the DSS-14/43 average (with atmosphere) is again reasonable, at  $18.9 \pm 0.6$  K. The average improvement of 1.8 K is a distinct S-band performance benefit. However, the atmosphere-free data is the more reliable. Here, the av-

erage difference is 1.5 K, still indicating a distinct S-band 70-m performance benefit.

At X-band, the DSS-14/64-m noise was abnormally high. However, after 70-m conversion all three antennas provided quite similar zenith noise performance of  $20.9 \pm 0.3$  K. Ignoring the abnormal DSS-14 64- to 70-m difference, contaminated by the high 64-m data, and taking the atmosphere-free data as the more reliable, the average difference of 0.7 K (70-m improvement) again provides another distinct 70-m performance benefit.

It is noted that the DSS-43 70-m antenna has the best S-band zenith noise performance, at 18.3 K, while the others are considerably higher. In decibels, DSS-14 is 0.2 dB worse, while DSS-63 is fully 1 dB worse at S-band, clearly a candidate for maintenance attention to either operational or calibration equipment.

It is emphasized that considerable noise data are required for a fully reliable performance description. Selected 70-m data are presented here. Figure 5 shows 70-m S-band total system noise temperatures (including atmosphere) taken during the calibration measurements at three different stations. The DSS-63 operational or instrumentation anomaly is clear, and these data are ignored for system noise temperature modeling. The DSS-43 18.3-K zenith temperature and the elevation dependence should be achievable in principle, and probably in practice as well, by all stations.

Figure 6 shows three selected 70-m X-band total system noise temperature data sets, again including atmosphere. At the time the DSS-14 data were obtained, the dichroic plate had not yet been reinstalled. Thus the 19.7-K measured total DSS-14 zenith noise shown in Fig. 6 projects to 20.9 K in the normal operating configuration with the dichroic plate installed. Remarkably, the three very low noise antennas perform quite uniformly with a zenith noise temperature of  $20.9 \pm 0.3$  K at X-band. Special efforts should be made to maintain this highly desirable and valuable performance condition.

From the limited data in Figs. 5 and 6, and the total system temperatures at zenith discussed above, the best present estimates of the DSN 70-m S- and X-band system noise temperatures are plotted in Fig. 7. The upper pair of plots are total system temperatures (with atmosphere) as functions of elevation angle. These are the operational noise temperatures (with slight adjustments) experienced in the nominal clear weather which existed at

the time of the calibration measurements. In the same figure, the lower pair of plots represents antenna performance alone, in what may be referred to as the "vacuum" condition, that is, with the nominal clear atmosphere removed. From atmosphere-removed plots, factors such as spillover and quadripod or other scatter may be estimated for detailed design evaluation purposes. For example, at 30-deg elevation, the atmosphere-removed 17.5-K (S-band) measurement is only 0.9 K above the the zenith (vacuum) of 16.6 K. Similarly, again at 30-deg elevation, the atmosphere-removed 20.6-K measurement (X-band) is only 2.2 K above the zenith (vacuum) of 18.4 K. These small noise terms are indeed difficult to know with great confidence, especially with limited data, but appear remarkably small and indicate the 70-m design has provided excellent low noise antennas.

In any careful future measurements and comparisons of antenna noise performance, it is recommended the feed-cone hoist cables and heavy large lifting sheaves and hooks be removed from the topmost two quadripod legs. These operationally needless (service) tools can only serve to degrade antenna noise performance, due to an unknown amount of ground noise pickup from additional scattering.

It can be concluded that the reduced cross-section 70-m quadripod, in the slightly deeper main reflector environment, is the key agent leading to 70-m system noise improvements. At zenith, 1.5-K S-band improvement and 0.7-K X-band improvement is observed. Observing noise as a function of elevation angle, one sees very small spillover and scatter components. It is quite clear, based on gain and noise temperature measurements, that the efforts expended to improve the 70-m quadripod structural design have yielded multiple benefits to microwave performance, at both frequency bands, to the three major antenna systems.

## IX. Conclusion

The detailed design-expected area efficiencies for the high-efficiency (dual-shaped reflector) DSN 70-m antennas range from 74.9 to about 75.8 percent at S-band, and from 68.8 to about 70.8 percent at X-band, based on the best estimating practices available. Measurements carried out during 1988 show 76.1 percent has been achieved at S-band, and an average of 68.6 percent achieved at X-band. The highest measured peak efficiency (DSS-63) is 70.1 percent, while the lowest measured efficiency is 67.3 percent



(at the X-band) where small-scale surface roughness effects are dominant.

The 70-m upgrade project achieved all objectives and more. The noise performance of all three antennas was improved, and it is clear from the limited gain and noise temperature measurements that the reduced cross-section quadripod contributed significantly to the overall performance improvement.

Additional X-band microwave performance is easily obtainable: the feedhorn design can be improved by 0.1 dB; and main reflector panels can be reset to obtain an additional 0.1 dB. Both the feedhorn and surface adjustment compromises were based on time and financial resource limitations. Nevertheless, those remain important, but recoverable, performance compromises.

One question remains unanswered. The structural/gravity (or focus) performance with elevation angle change is not as good as expected for two of the antennas (DSS-43 and DSS-63). Further work in this area will be accomplished in a post-project phase, beginning with focus tests at high and low elevation angles, and the study of valuable low-elevation holography images. These are expected to reveal actual structural gravity behavior for comparison with theoretical models.

The DSN 70-m project was a significant undertaking, and by any reasonable standard, was severely time constrained to fit between the Voyager Uranus (1986) and Neptune (1989) planetary encounters. Good design, management, contracting, scheduling, fieldwork, and technology formed a powerful combination. The use of microwave holography proved to be vital to the Project success.

## Acknowledgments

It would be impossible to individually acknowledge all of the people on three continents who contributed to the success of the DSN 70-m antenna upgrade project. Each of those many individuals can be assured that his or her efforts have contributed to greatly improved performance of the NASA/JPL Deep Space Network.

## References

- [1] D. A. Bathker and S. D. Slobin, "DSN 70-Meter Antenna Microwave Optics Design and Performance Improvements Part I: Design Optimization," *TDA Progress Report 42-97*, Jet Propulsion Laboratory, Pasadena, California, pp. 306-313, May 15, 1989.
- [2] P. H. Richter and S. D. Slobin, "DSN 70-Meter Antenna X- and S-Band Calibration Part I: Gain Measurements," *TDA Progress Report 42-97*, Jet Propulsion Laboratory, Pasadena, California, pp. 314-351, May 15, 1989.
- [3] S. D. Slobin and P. H. Richter, "DSN 70-Meter Antenna X- and S-Band Calibration Part II: System Noise Temperature and Telecommunications Link Evaluation," *TDA Progress Report 42-97*, Jet Propulsion Laboratory, Pasadena, California, pp. 352-366, May 15, 1989.
- [4] S. D. Slobin, "DSS 63 64-m Antenna S- and X-band Efficiency and System Noise Temperature Calibration, July 1986," *TDA Progress Report 42-90*, Jet Propulsion Laboratory, Pasadena, California, pp. 103-115, August 15, 1987.

- [5] S. D. Slobin, "DSS 43 64-Meter Antenna S- and X-Band Efficiency and System Noise Temperature Calibrations, January 1987," *TDA Progress Report 42-90*, Jet Propulsion Laboratory, Pasadena, California, pp. 116-126, August 15, 1987.
- [6] S. D. Slobin, "DSS 14 64-Meter Antenna S- and X-Band Efficiency and System Noise Temperature Calibrations, September 1987," *TDA Progress Report 42-92*, pp. 138-146, February 15, 1988.
- [7] P. D. Batelaan, A. J. Freiley, and D. A. Bathker, "Absolute Flux Density Calibrations of Radio Sources: 2.3 GHz," *JPL Technical Memorandum 39-806*, Jet Propulsion Laboratory, Pasadena, California, 1977.
- [8] J. M. Schredder, "Seventy-Meter Antenna Performance Predictions: GTD Analysis Compared With Traditional Ray-Tracing Method," *TDA Progress Report 42-92*, Jet Propulsion Laboratory, Pasadena, California, pp. 166-174, February 15, 1988.
- [9] S. D. Slobin and W. A. Imbriale, "DSS 43 Antenna Gain Analysis for Voyager Uranus Encounter: 8.45-GHz Radio Science Data Correction," *TDA Progress Report 42-90*, Jet Propulsion Laboratory, Pasadena, California, pp. 127-135, August 15, 1987.
- [10] S. D. Slobin and D. A. Bathker, "DSN 70-Meter Antenna X-Band Gain, Phase, and Pointing Performance, With Particular Application for Voyager 2 Neptune Encounter," *TDA Progress Report 42-95*, Jet Propulsion Laboratory, Pasadena, California, pp. 237-245, November 15, 1988.
- [11] J. Ruze, "Antenna Tolerance Theory—A Review," *Proceedings of the IEEE*, vol. 54, no. 4, pp. 633-640, April 1966.

**Table 1. 8420-MHz sources for DSN antenna calibrations**

Source	Flux, Jy	Source Size/Beamwidth Correction, C(R)	
		64-m	70-m
3C274	Old: 46.00	Old: 1.085 <sup>a</sup> 1.089 <sup>b</sup>	Old: 1.106
	New: 45.20	New: 1.15	New: 1.18
	(+0.076 dB)	(+0.253 dB) <sup>a</sup> (+0.237 dB) <sup>b</sup>	(+0.281dB)
Others	(+0.076 dB)	Varies according to source size	

Notes:

Jy = Jansky =  $10^{-26}$  W/m<sup>2</sup>/Hz

<sup>a</sup>JPL Document D-3801, Rev. A (DSS-43/63, 64-m calibration)

<sup>b</sup>JPL Document D-3801, Rev. B (DSS-14, 64-m, and initial 70-m calibrations)

**Table 2. Summary of design expected and average measured efficiencies, 64-m antennas**

64-m Antenna/Band	Design-expected nominal efficiency	Estimated minimum-likely efficiency	Average measured efficiency
All/S-Band	0.603	0.591	0.582
DSS-14/X-Band	0.565	0.541	0.535
DSS-43/X-Band	0.565	0.533	0.490
DSS-63/X-Band	0.565	0.533	0.487

**Table 3. Summary of design-expected and average measured efficiencies, 70-m antennas**

70-m Antenna/Band	Design-expected nominal efficiency	Estimated minimum-likely efficiency	Average measured efficiency
All/S-band	0.758	0.749	0.761
All/X-band	0.708	0.686	0.686

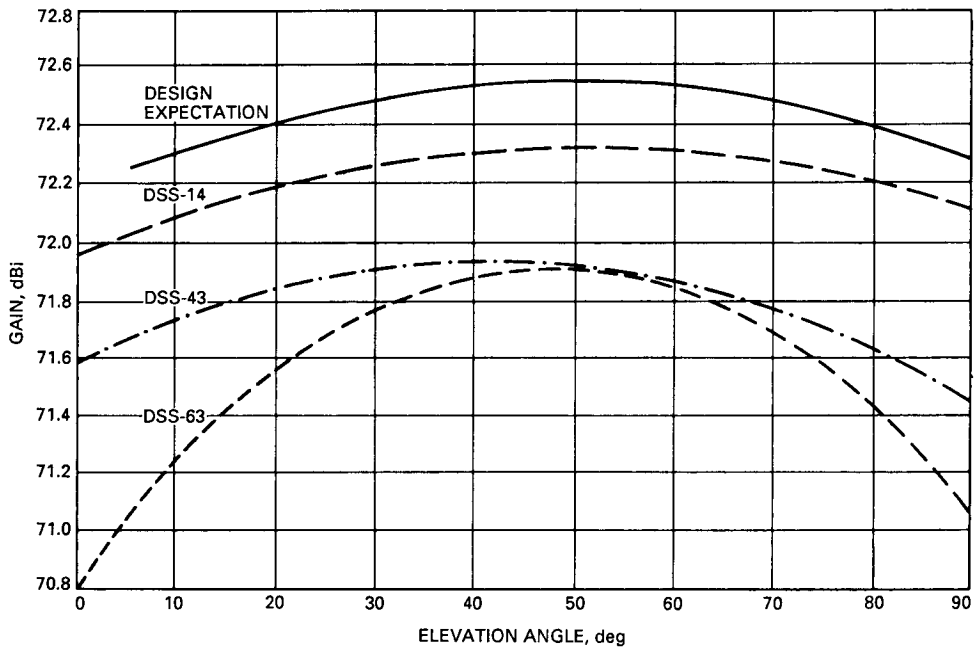
**Table 4. Summary of measured S- and X-band gain and efficiency results**

	DSS-14 X-Band	DSS-43 X-Band	DSS-63 X-Band	All Stations S-Band
<b>64-m</b>				
Gain, dBi	72.33	71.94	71.91	61.40
Efficiency	0.535	0.490	0.487	0.582
<b>70-m</b>				
Gain, dbi	74.17	74.10	74.28	63.34
Efficiency	0.684	0.673	0.701	0.761
<b>64-70-m</b>				
Gain improvement, dB	+1.84	+2.16	+2.37	+1.94

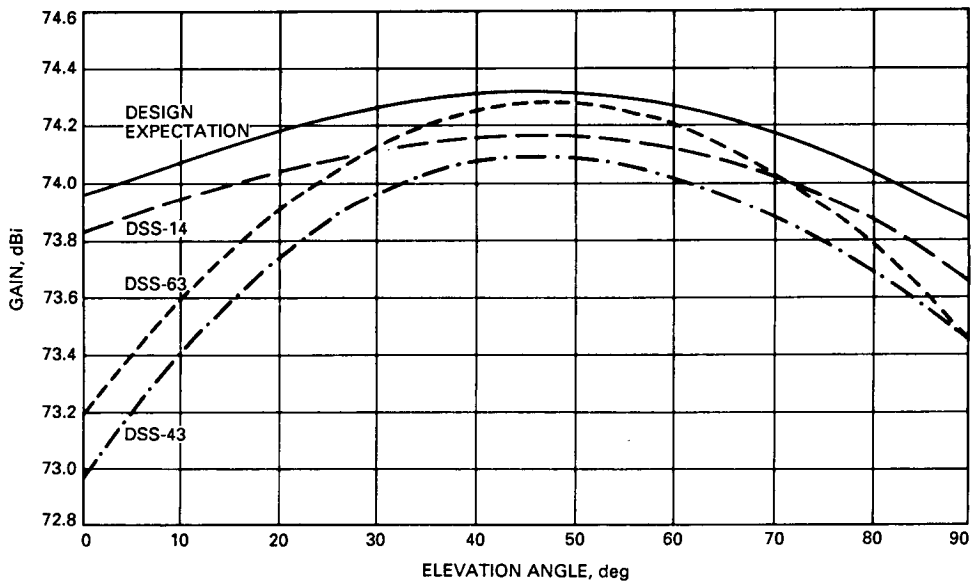
**Table 5. Summary of zenith system noise temperatures for 64/70-m antennas**

Antenna/System	S-Band Zenith noise, K	X-Band Zenith noise, K
<b>With Atmosphere</b>		
DSS-14/64-m	21.2	25.2*
DSS-14/70-m	19.5 (-1.7)	20.9 (-4.3)*
DSS-43/64-m	20.1	21.8
DSS-43/70-m	18.3 (-1.8)	21.2 (-0.6)
DSS-63/64-m	27.7*	22.0
DSS-63/70-m	23.0 (-4.7)*	20.6 (-1.4)
<b>Without Atmosphere</b>		
DSS-14/64-m	19.1	23.2*
DSS-14/70-m	17.7 (-1.4)	18.7 (-4.5)*
DSS-43/64-m	18.2	19.2
DSS-43/70-m	16.5 (-1.7)	18.7 (-0.5)
DSS-63/64-m	25.6*	19.0
DSS-63/70-m	21.2 (-4.4)*	18.1 (-0.9)

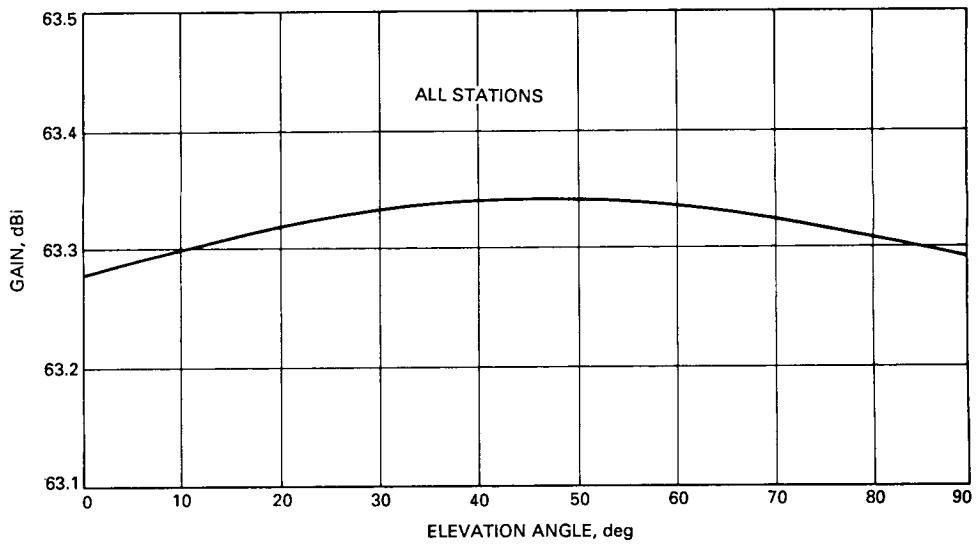
\* Considered abnormal, and change not representative of design improvement.



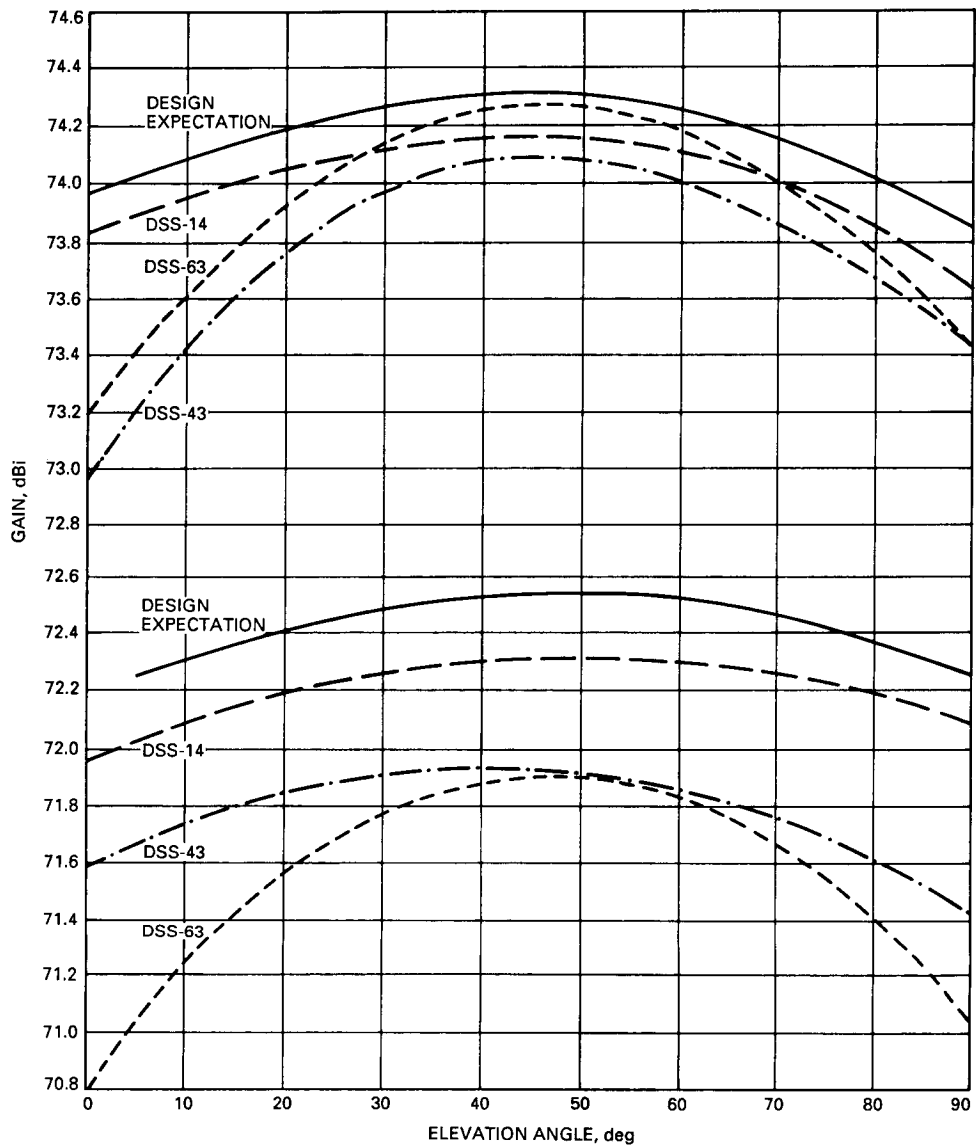
**Fig. 1. DSN 64-m antenna X-band gain performance (without atmosphere), plus design expectation.**



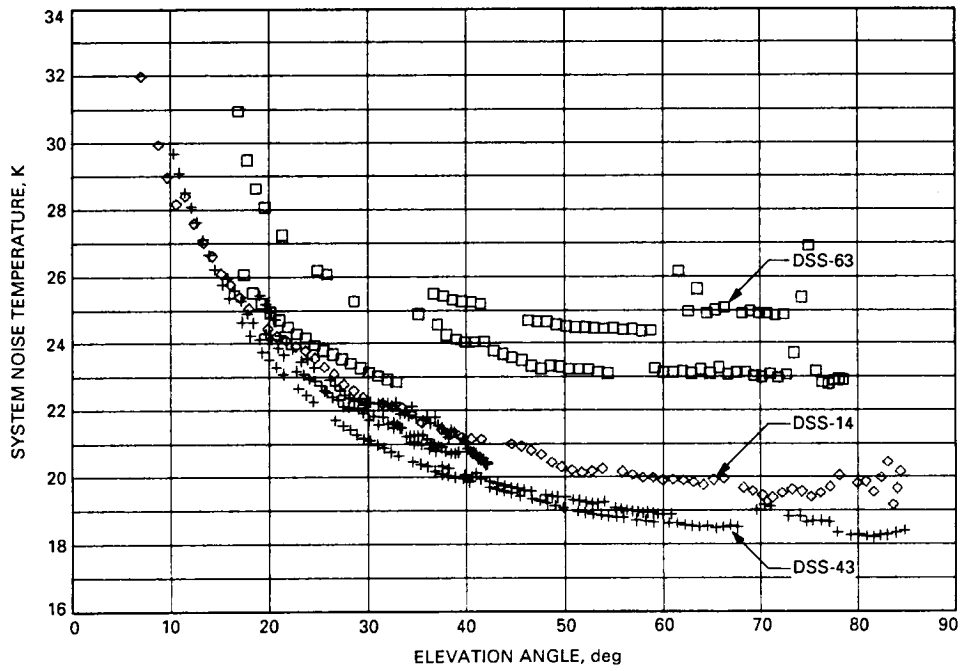
**Fig. 2. DSN 70-m antenna X-band gain performance (without atmosphere), plus design expectation.**



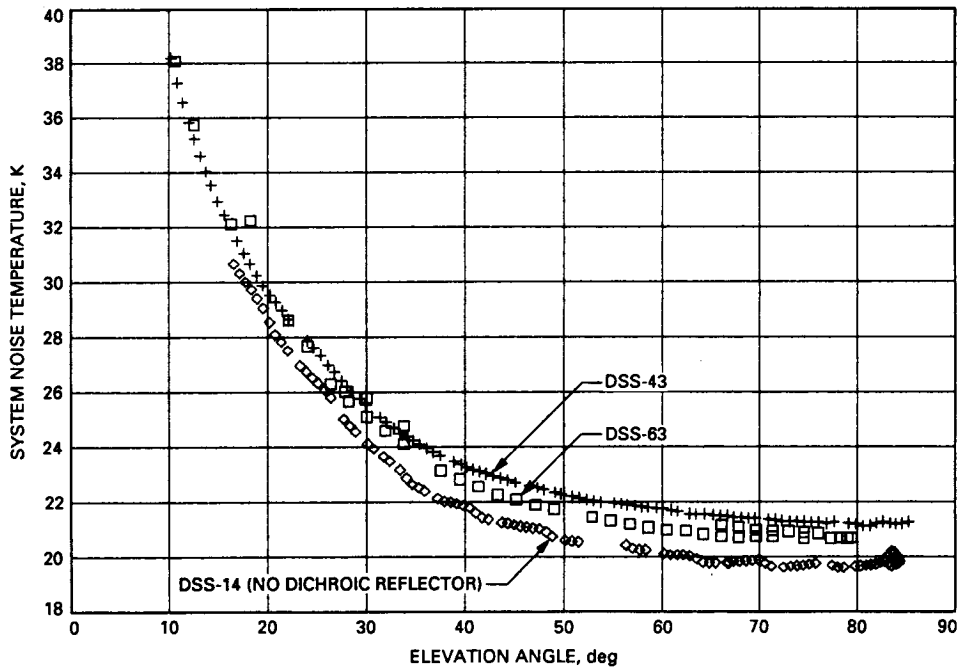
**Fig. 3. DSN 70-m antenna S-band gain performance (without atmosphere).**



**Fig. 4. DSN 64/70-m antenna X-band gain comparison (without atmosphere), plus design expectation.**

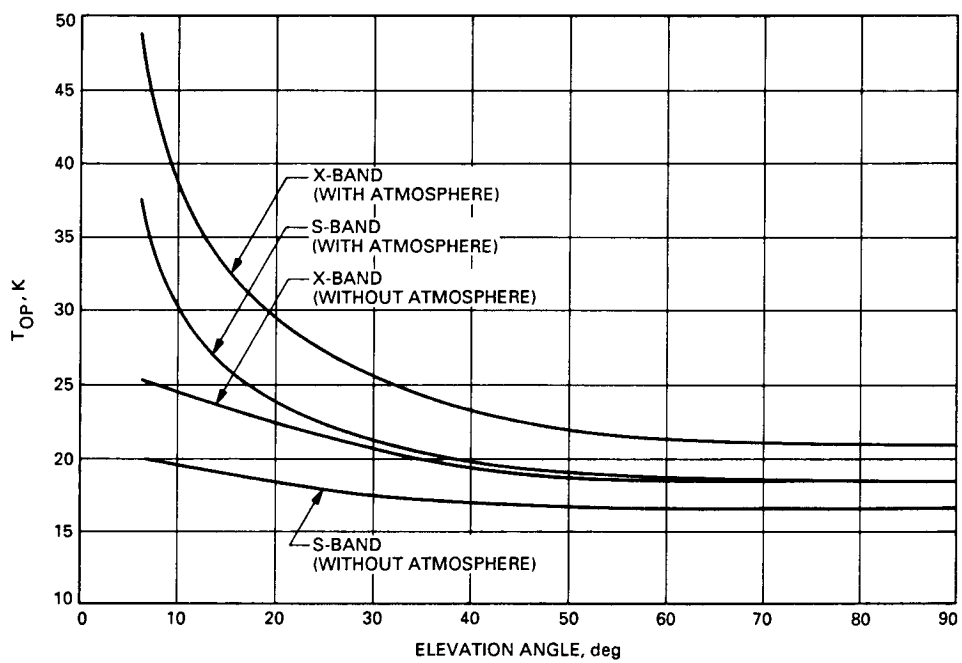


**Fig. 5. DSN 70-m antenna S-band system noise temperature (raw data, with atmosphere).**



**Fig. 6. DSN 70-m antenna X-band system noise temperatures (raw data, with atmosphere) (DSS-14: no dichroic reflector).**





**Fig. 7. DSN 70-m S- and X-band system noise temperatures (with nominal clear atmosphere, and in "vacuum" condition).**

915-33

13-725

547680 P18

N90-12801

# Stable Kalman Filters for Processing Clock Measurement Data

P. A. Clements

Communications Systems Research Section

B. P. Gibbs and J. S. Vandergraft

Computational Engineering, Inc., Laurel, Maryland

*Kalman filters have been used for some time to process clock measurement data. Due to instabilities in the standard Kalman filter algorithms, the results have been unreliable and difficult to obtain. Often, in order to obtain reasonable results, the data has had to be manually edited, the filter fine tuned, or the model adjusted by the analyst. During the past several years, stable forms of the Kalman filter have been developed, implemented, and used in many diverse applications. These algorithms, while algebraically equivalent to the standard Kalman filter, exhibit excellent numerical properties. Two of these stable algorithms, the Upper triangular-Diagonal (UD) filter and the Square Root Information Filter (SRIF), have been implemented to replace the standard Kalman filter used to process data from the DSN hydrogen maser clocks. The data are time offsets between the clocks in the DSN, the timescale at the National Institute of Standards and Technology (NIST), and two geographically intermediate clocks. The measurements are made by using the GPS navigation satellites in mutual view between clocks. The filter programs allow the user to easily modify the clock models, the GPS satellite dependent biases, and the random noise levels in order to compare different modeling assumptions.*

*The results of this study show the usefulness of such software for processing clock data. The UD filter is indeed a stable, efficient, and flexible method for obtaining optimal estimates of clock offsets, offset rates, and drift rates. A brief overview of the UD filter is also given.*

## I. Introduction

The requirements of the DSN-complex clocks are that each clock be maintained within 10 microseconds of Co-

ordinated Universal Time (UTC) as realized at the National Institute of Standards and Technology (NIST), with a knowledge of 1 microsecond and that the rate of each clock be kept within  $\pm 1E - 12df/f$  with a knowledge of

$\pm 3E - 13df/f$ . The quantity  $df/f$  is known as the fractional frequency deviation where

$$df/f = \frac{f_m - f_0}{f_0}$$

and  $f_m$  is the measured frequency and  $f_0$  is the reference frequency. Also there is a requirement that a permanent record be kept of the synchronization and syntonization of the DSN-complex clocks. In order to meet these requirements, the time and frequency offset to UTC (NIST) is monitored using the Global Positioning System (GPS) navigation space vehicles. The goal (not yet realized) is to adjust the rates of all the clocks at the same time at approximately 200-day intervals. The clock rates (frequencies) are adjusted when they are above UTC (NBS) by a certain amount to below, and then allowed to drift through zero rate offset from UTC (NIST) until they are again high enough to require adjustment. This technique, done with some care, allows one to keep the clocks within the required time and frequency offset limits and still not have any jumps in the time offset. All of the drift rates of the hydrogen masers are about the same, therefore, the complex clocks can be kept in close synchronization and syntonization.

Monitoring of the DSN-complex clocks is done by the operations analyst at JPL. Adjustments of the clocks is recommended by the operations analyst with the advice of the engineering staff. The adjustment is scheduled after a check of the users of the DSN and is made during a maintenance period at the complex. The actual adjustment is made by the engineering staff at the complex. It is assumed that the drift of the hydrogen maser rate is due to cavity drift, therefore, rate adjustments are made by adjusting the varactor, thereby changing the apparent cavity size of the hydrogen maser.

At critical times during some projects (such as the Voyager Uranus encounter), it is desirable to have the clocks synchronized and syntonized more closely than the above specifications. When this is the situation, the operations analyst will predict the time and rate of the DSN clocks 30 to 90 days prior to the encounter date and set the rate to allow the clocks to drift to the near zero time and rate offset by the critical time. At the time of the Voyager Uranus encounter, JPL engineering staff used a Kalman filter to predict the clock offset 60 days into the future with excellent results [3].

At JPL, a Kalman filter has been used to process clock offset measurements but it has never been used operationally. The present operational environment for clock management in the DSN is a personal computer with a considerable amount of manual intervention. The data is handled in weekly batches. By 1992 the process is to be transferred to the Network Frequency and Timing subsystem (NFT), which is a part of the DSN Frequency and Timing system (DFT). The flow of data and estimate updates will occur more often (perhaps as often as once per hour) so as to allow operations to be able to monitor clock performance in near real time. This will require a fairly robust processing environment with automatic elimination of outliers and other techniques to make the process as autonomous as possible. It is expected that operations-analyst intervention will be relegated to filter adjustments in response to scheduled changes of time or rate in any of the clocks being entered into the filter parameters.

The operations analysts will have the opportunity to do postanalysis of the data. In the near future, it is expected that some data will continue to be available on a weekly basis and that better ephemerides will be available several weeks after the fact. At this time, use of the filter/smoothener by the operations analyst can produce better estimates of the clock parameters for archiving. To achieve these goals, a robust filter and smoothener program is under development at JPL. The results reported in this article were obtained by running this program with data as described in the next section. No "fine tuning" adjustments or manual data editing were performed. The output, however, was excellent. The filter used in the JPL program is the numerically stable Upper triangular-Diagonal (UD) form of the Kalman filter. The UD filter has been under development for several years [1] and has earned a reputation for being accurate and efficient. The more traditional Kalman formulation has been known for some time to be numerically unstable. That is, due to rounding errors within the computer, the results produced by the filter program are sometimes completely wrong. The smoothener used in this code is a recent improvement [2] of the Rauch-Tung-Streibel (RTS) smoothener [5]. The revised RTS smoothener is also robust and accurate.

The filter and smoothener programs make use of subroutines from the Estimation Subroutine Library (ESL). This is a collection of FORTRAN77 subroutines for constructing filters and smootheners. The routines in the ESL have been carefully coded and tested to ensure their accuracy and reliability. Overall, the goal is to produce optimal estimates of clock parameters without manual intervention. The results of this first attempt are extremely encouraging.

## II. Clock Measurements

There is an NIST-designed GPS timing receiver located at each complex. A timing pulse from the complex master clock is fed to each receiver so that the data output from the receiver is the offset of the complex master clock to GPS timescale. The receivers are maintained on the mutual view schedule, which is generated by the Bureau International des Poids et Mesures (BIPM) in Paris, France. These data are kept in a database along with data from UTC (RRL) and UTC (TAO) in Japan, UTC (NIST), and the NIST cesium clock at WWVH Hawaii, U. S. The data from Japan are obtained from the GE MKIII catalogue, which is administered worldwide by the BIPM and inside the U. S. by the United States Naval Observatory (USNO). Each of the data lines is tagged to indicate the receiver from which it came and then placed in a file and stored on a disk in a personal computer. There is one file for each day, and it is named by Modified Julian Day.<sup>1</sup>

The first process is to take the first difference of the data to obtain the differences between the clocks on the ground, thus eliminating the space vehicle clocks. There are six clocks used; because of the geometry of the clock locations there are 11 possible mutual views available (Fig. 1). A program is used to produce a file which contains as entries the station pair, the satellite number, the MJD, the time (second) the data was collected, and the difference in time in tenths of nanoseconds between the first and second clock. (The above format could easily be adapted to include clock difference data from other sources such as two-way time coordination.) Data was generated by this program from MJD 47000 to 47132. There were approximately 4000 measurements in the data set.

## III. Mathematical Model

In order to describe the mathematical model used for this analysis, the following notation is introduced:

$s_i(t)$ : the clock offset at station  $i$ . That is,  $s_i = c_i - T$ , where  $c_i$  is the clock reading at station  $i$  and  $T$  is the "true" time. For this analysis, "true" time is taken to be the time given by UTC (NIST).

$r_i(t)$ : the clock offset rate, also called the frequency error. This is just the time derivative of the offset, i.e.,  $r_i = ds_i/dt$ .

$d_i(t)$ : the drift rate, that is, the second derivative of the offset,  $d_i = d^2s_i/dt^2$ .

$b_{ikl}(t)$ : the measurement biases. These depend on the two stations,  $i$  and  $k$ , and on the satellite,  $l$ , being observed. These biases will be explained in more detail later.

The dynamics model describes how the above quantities change over time. This is expressed as follows:

$$\dot{s}_i(t) = r_i(t) \quad (1)$$

$$\dot{r}_i(t) = d_i(t) \quad (2)$$

$$\dot{d}_i(t) = w_i(t) \quad (3)$$

where  $\dot{\cdot}$  denotes time differentiation, and  $w_i(t)$  is white noise, with zero mean and known power spectral density (psd)  $W$ . That is,

$$E[w_i(t)w_i(\tau)] = W\delta(t - \tau) \quad (4)$$

where  $E[\cdot]$  is the *expected value* operation, and where  $\delta$  is the Dirac delta function. Note that Eqs. (1) and (2) are just restatements of the definitions of  $r_i$  and  $d_i$ . Equation (3) says that the rates are "approximately" constant.

In order to apply a discrete filter to this problem, Eqs. (1-3) must be put into the form of a discrete dynamics equation

$$\underline{x}(t + \Delta t) = \Phi(t, \Delta t)\underline{x}(t) + w(t) \quad (5)$$

where  $\underline{x}$  denotes a vector with components that include  $s_i$ ,  $r_i$ ,  $d_i$  for each station. Note that all of these quantities are time-dependent; however, to simplify the notation, this dependence will often be suppressed. The matrix  $\Phi$ , called the transition matrix, is determined by writing out a discrete form of Eqs. (1-3). A simple way to do this, for example, at least for the offsets and rates, is to use a discrete form of Taylor's Theorem:

$$s_i(t + \Delta t) = s_i(t) + r_i(t)\Delta t + 0.5d_i(t)\Delta t^2 \quad (6)$$

$$r_i(t + \Delta t) = r_i(t) + d_i(t)\Delta t \quad (7)$$

<sup>1</sup> The Modified Julian Day (MJD) is a continuous day count with an initial epoch of 0000 hours UT on November 17, 1858.

For the drift rates, this same approach produces

$$d_i(t + \Delta t) = d_i(t) + \int_0^{\Delta t} w_i(\tau) d\tau \quad (8)$$

where  $w_i(t)$  represents the random noise. However, the effects of the noise on the drift rates are shown directly in Eq. (8), but this same noise also affects the offsets and frequency errors because of the presence of  $d_i(t)$  in Eqs. (6) and (7). As written above, however, there will be a delay, caused by the time discretization, before the noise at time  $t$  propagates into the offset and rate. To avoid this delay, noise terms are included directly in the equations for  $s_i$  and  $r_i$ . This is done by using the continuous form of the transition matrix to propagate the noise over the interval  $t$  to  $t + \Delta t$ . That is, if the noise that enters the continuous system at time  $t$  is given by the vector

$$w(t) = (0, 0, w_r(t))^T$$

where  $w_r(t)$  is white noise with zero mean and psd  $W_r$ , then the integrated effect of the noise at time  $t + \Delta t$  will be the vector

$$q(t + \Delta t) = (q_s, q_r, q_d)^T$$

where  $(q_s, q_r, q_d)^T \sim N(0, Q)$  and

$$\begin{aligned} Q &= E \left[ \int_0^{\Delta t} \Phi(\Delta t - \tau) w(\tau) d\tau \int_0^{\Delta t} w^T(\sigma) \Phi^T(\Delta t - \sigma) d\sigma \right] \\ &= E \left[ \int_0^{\Delta t} \int_0^{\Delta t} \Phi(\tau) w(\Delta t - \tau) w^T(\Delta t - \sigma) \Phi^T(\sigma) d\sigma d\tau \right] \\ &= (0.5\tau^2, \tau, 1)^T W_r (0.5\tau^2, \tau, 1) d\tau \end{aligned} \quad (9)$$

since

$$\Phi(t) = \begin{bmatrix} 1 & t & 0.5t^2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix}$$

and  $E[w_r^T w_r] = W_r$ . The integration in Eq. (9) gives

$$Q = W_r \begin{bmatrix} \frac{1}{20} \Delta t^5 & \frac{1}{8} \Delta t^4 & \frac{1}{6} \Delta t^3 \\ \frac{1}{8} \Delta t^4 & \frac{1}{3} \Delta t^3 & \frac{1}{2} \Delta t^2 \\ \frac{1}{6} \Delta t^3 & \frac{1}{2} \Delta t^2 & \Delta t \end{bmatrix} \quad (10)$$

Thus the discrete dynamics model for these filter states is

$$\begin{pmatrix} s_i(t + \Delta t) \\ r_i(t + \Delta t) \\ d_i(t + \Delta t) \end{pmatrix} = \Phi_i \begin{pmatrix} s_i(t) \\ r_i(t) \\ d_i(t) \end{pmatrix} + w_i \quad (11)$$

where

$$\Phi_i = \begin{bmatrix} 1 & \Delta t & \frac{1}{2} \Delta t^2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

$w_i \sim N(0, Q_i)$ , and  $Q_i$  is given by Eq. (10). There will be a transition, of the form of Eq. (11), for each station. Thus, the full transition matrix  $\Phi$  in Eq. (5) will have blocks of the form of Eq. (12) on its diagonal. The process noise term,  $w(t)$ , in Eq. (5) will have as its covariance matrix a block diagonal matrix, with  $3 \times 3$  matrices of the form of Eq. (10) on the diagonal.

The measurements to be used for estimating the offsets, drifts, and drift rates are the differences in clock values from the GPS times, as measured at two different stations. That is, the measurements are

$$z_{ikl}(t_j)$$

which are assumed to be of the form

$$\begin{aligned} z_{ikl}(t_j) &= (c_i(t_j) - GPS_l) - (c_k(t_j) - GPS_l) \\ &\quad + b_{ikl}(t_j) + v_{ik}(t_j) \\ &= c_i(t_j) - c_k(t_j) + b_{ikl}(t_j) + v_{ik}(t_j) \end{aligned} \quad (13)$$

where, as defined earlier,  $c_i(t_j)$  is the clock reading at station  $i$  at time  $t_j$ . The fact that  $s_i(t_j) = c_i(t_j) - NIST(t_j)$  allows one also to write this as

$$\begin{aligned}
z_{ikl}(t_j) &= (c_i(t_j) - NIST(t_j)) - (c_k(t_j) - NIST(t_j)) \\
&\quad + b_{ikl}(t_j) + v_{ik}(t_j) \\
&= s_i(t_j) - s_k(t_j) + b_{ikl}(t_j) + v_{ik}(t_j)
\end{aligned}$$

The first expression is used to obtain the measurement data as described in Section II. That is, the data actually used for the analysis are values for  $c_i(t_j) - GPS_i(t_j)$ . However, for the filter/smoothing, the measurements are assumed to be of the second form, where the biases,  $b_{ikl}(t_j)$ , satisfy the trivial dynamics equation

$$b_{ikl}(t + \Delta t) = b_{ikl}(t) \quad (14)$$

and the *measurement noise*,  $v_{ik}(t_j)$ , is, as usual, a normally distributed zero mean random variable with constant variance  $r$ . The biases are appended to the "core states" to obtain the actual state vector used in the filter:

$$\mathbf{x} = (s_1, r_1, d_1, s_2, r_2, d_2, \dots, s_m, r_m, d_m, b_{111}, b_{112}, b_{113}, \dots)$$

The transition matrix will have an identity matrix attached to it to reflect Eq. (14). Then the measurements can be written in standard form as

$$z_j = H_j \mathbf{x}_j + v_j \quad (15)$$

where  $H_j$  is a  $1 \times n$  measurement matrix, consisting, in this case, of two positive ones and a negative one, in appropriate positions. For example, a measurement based on observations from stations 1 and 3, and satellite 6, will have a measurement matrix

$$H = (1, 0, 0, 0, 0, 0, -1, 0, 0, \dots, 0, 1, 0, \dots)$$

where the final one appears in the position corresponding to bias  $b_{136}$ .

## IV. The UD Form of the Kalman Filter

The Kalman Filter has become the standard tool for computing optimal estimates of a state vector,  $\mathbf{x}(t)$ , governed by a mathematical model of the form of Eqs. (5) and (15). To describe the form of the filter that is of interest here, the following notation will be used:

$\mathbf{x}_{k|j}$ : estimate of  $\mathbf{x}(t_k)$  using data up to time  $t_j$

$P_{k|j}$ : covariance matrix for the error in  $\mathbf{x}_{k|j}$  as an approximation to  $\mathbf{x}(t_k)$

The Kalman Filter time update, also called the prediction step is

$$\mathbf{x}_{k|k-1} = \Phi_k \mathbf{x}_{k-1|k-1} \quad (16)$$

$$P_{k|k-1} = \Phi_k P_{k-1|k-1} \Phi_k^T + Q_k \quad (17)$$

The measurement update is accomplished via the following set of equations:

$$B_k = H_k P_{k|k-1} H_k^T + R_k \quad (18)$$

$$K_k = P_{k|k-1} H_k^T B_k^{-1} \quad (\text{Kalman gain}) \quad (19)$$

$$P_{k|k} = (I - K_k H_k) P_{k|k-1} \quad (20)$$

$$\nu_k = z_k - H_k \mathbf{x}_{k|k-1} \quad (\text{innovations}) \quad (21)$$

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + K_k \nu_k \quad (22)$$

The numerical instability of the Kalman equations, as noted in Section I, is due to the structure of these equations. The most common indication that something has gone wrong is when rounding error causes  $P_{k|k}$  to be not positive definite. Note that it is not clear from these equations that  $P_{k|k}$  is even symmetric.

The UD form of the Kalman Filter is algebraically equivalent to the above equations, but has much better numerical properties. It is based on the fact that any positive definite matrix  $P$  can be written as a product of matrices

$$P = U D U^T \quad (23)$$

where  $U$  is an upper triangular matrix with ones on the diagonal, and  $D$  is a diagonal matrix, with positive diagonal elements. The idea behind the UD formulation is to start with the UD factors of  $P_{0|0}$ , then compute the factors of all the remaining covariances, without ever computing the covariances themselves. It can be shown that by avoiding the covariances, the accuracy of the results is greatly improved. On the other hand, if the covariances are needed for output they can be easily obtained by simply multiplying the factors together.

To describe the time update in UD form, assume that the factors of  $P_{k|k}$  are known from the previous step. That is,  $U_{k|k}$  and  $D_{k|k}$  have already been computed, where

$$P_{k|k} = U_{k|k} D_{k|k} U_{k|k}^T \quad (24)$$

Then  $\underline{x}_{k+1|k}$  is computed just as in the Kalman formulation; but the equation for  $P_{k+1|k}$  becomes

$$\begin{aligned} P_{k+1|k} &= \Phi_k P_{k|k} \Phi_k^T + Q_k \\ &= \Phi_k U_{k|k} D_{k|k} U_{k|k}^T \Phi_k^T + Q_k \end{aligned} \quad (25)$$

Since  $\Phi_k$  is upper triangular, so is the product  $\Phi_k U_{k|k}$ . Hence, if  $Q_k$  were zero, then the factors of  $P_{k+1|k}$  would be just

$$R_k D_k R_k^T$$

where

$$R_k = \Phi_k U_{k|k} \quad (26)$$

In order to account for the process noise,  $Q_k$  is factored into its UD factors

$$Q_k = U_q D_q U_q^T$$

which can also be written as

$$Q_k = \sum_{i=1}^n d_i q_i q_i^T \quad (27)$$

where  $q_i$  is the  $i$ th column of  $U_q$ . Then Eq. (25), using Eqs. (26) and (27) becomes

$$P_{k+1|k} = R_k D_k R_k + \sum_{i=1}^n d_i q_i q_i^T \quad (28)$$

Now the time update step is completed by finding the UD factors of  $P_{k+1|k}$ . This is done by a series of "rank one" adjustments to previous factors. That is, consider the general problem: given factors  $R$ ,  $D$ , compute the factors of the matrix

$$RDR^T + dq q^T$$

where  $d$  is a positive scalar, and  $q$  is a vector.

The matrix  $dqq^T$  is a matrix of rank one. Hence  $dqq^T$  is called a "rank one" adjustment to  $RDR^T$ . There is a very accurate and efficient algorithm for computing the factors of this new matrix, given the factors  $R$ ,  $D$ , and the quantities  $d$  and  $q$ . It is important here that the factors are not multiplied together, as this would destroy the numerical accuracy. This rank-one update algorithm, called the Turner-Agee Algorithm [1], is applied repeatedly to Eq. (28) to complete the time update step of the filter.

Next, consider the measurement update, as given in the Kalman form by Eqs. (18-22). The Kalman equation can be used to process vector measurements. However, the UD form is more efficient if the measurements are scalars. In this case, in the measurement update equations,  $B_k$  is a scalar, as is  $v_k$ , and  $K$  is a vector. The measurement update is done several times, once for each scalar measurement for each time update. After the previous time update, the factors  $U_{k|k-1}$  and  $D_{k|k-1}$  of  $P_{k|k-1}$  are available. The (scalar)  $B_k$  as defined by Eq. (18) is computed as follows:

$$\begin{aligned} B_k &= H_k U_{k|k-1} D_{k|k-1} U_{k|k-1}^T H_k^T + R_k \\ &= y_k^T D_{k|k-1} y_k + R_k \end{aligned}$$

where

$$y_k = U_{k|k-1}^T H_k^T$$

is a vector. Next, Eqs. (19) and (20) are written as

$$\begin{aligned} P_{k|k} &= (I - K_k H_k) P_{k|k-1} \\ &= P_{k|k-1} - \frac{1}{B_k} P_{k|k-1} H_k^T H_k P_{k|k-1} \\ &= U_{k|k-1} D_{k|k-1} U_{k|k-1}^T \\ &\quad - \frac{1}{B_k} U_{k|k-1} D_{k|k-1} U_{k|k-1}^T H_k^T H_k U_{k|k-1} D_{k|k-1} U_{k|k-1}^T \\ &= U_{k|k-1} \left[ D_{k|k-1} - \frac{1}{B_k} V V^T \right] U_{k|k-1}^T \end{aligned} \quad (29)$$

where  $V$  is the vector defined by

$$V = D_{k|k-1} U_{k|k-1}^T H_k^T$$

All that remains is to compute the factors  $U_v$ ,  $D_v$  of

$$D_{k|k-1} - \frac{1}{B_k} VV^T \quad (30)$$

Then it follows that

$$P_{k|k} = \underbrace{U_{k|k-1} U_v}_{U_{k|k}} \underbrace{D_v}_{D_{k|k}} \underbrace{U_v^T U_{k|k-1}^T}_{U_{k|k}^T} \quad (31)$$

Notice that Eq. 30 is a rank-one modification of a diagonal matrix. Hence its factors can be computed accurately and efficiently, just as was done in the time update.

While this UD formulation may seem more complicated than the relatively simple matrix equations of the Kalman form (Eqs. 16–22), a computer program to implement the UD filter is very straightforward. Two special subroutines are needed:

- (1) A subroutine to efficiently compute the product of two upper triangular matrices, such as  $\Phi_k U_k$
- (2) A subroutine to adjust the factors of a matrix by a rank-one addition: given factors  $U$ ,  $D$ , a positive scalar  $c$ , and a vector  $y$ , compute the factors of  $UDU^T + cyy^T$

Then the time update is done by the following steps:

- (1) Compute  $R = \Phi_k U$  (sometimes called the deterministic update)
- (2) For each process noise term  $d_i q_i q_i^T$ , update the factors using the rank-one adjustment routine

The measurement update is also broken down into a series of simple steps, one for each scalar measurement:

- (1) Compute the scalar  $B = y^T D y + r$  where  $y = U^T H^T$
- (2) Use the rank-one adjustment routine to compute the factors of  $D - \frac{1}{B} VV^T$  where  $V = DU^T H^T$
- (3) Combine these factors with the previous factors, according to Eq. 31

## V. Filter Implementation

In addition to the clock measurement data, as described in Section II, the filter routine must be given the following information:

- (1) Prior estimates of the state vector
- (2) Uncertainties associated with the state estimates
- (3) Process-noise standard deviations
- (4) Measurement-noise standard deviations

Generally, the filter is somewhat insensitive to prior estimates of the state; hence, they are usually initialized at zero. The uncertainties must be large enough so that the filter will accept the measurement data and use it to adjust the state estimates, except that the offsets for the NIST station should remain close to zero. The process noise can depend on the station; however, for the results of Section VI, all stations were given the same amount of noise. Similarly, the measurement noise can be made a function of the stations and the space vehicle. For simplicity during this first analysis, however, the measurement noise uncertainty was assumed to be the same for all measurements.

In order to run the smoother, certain information must be saved at each measurement update time. Thus, to limit the amount of storage required to run the filter and smoother, the program was set up to process measurements in so-called "mini-batches." Each mini-batch consists of all measurements obtained during a fixed timespan, say one day, or a half a day. The midpoint of this timespan is chosen as the time of the update (also called the "epoch time"). The measurements within the mini-batch are translated to the epoch time by using the same transition matrix,  $\Phi(t_1, t_2)$ , as is used in the dynamics equation. More precisely, suppose that the actual measurement is

$$z(t_1) = Hx(t_1) + w \quad (32)$$

where  $t_1$  denotes the time at which the measurement was taken. Let  $t_2$  be the epoch time for this mini-batch. Then, if it is assumed that the timespan for this mini-batch is sufficiently short, so that the integrated effect of the noise is negligible, one can write

$$x(t_1) = \Phi(t_2, t_1)x(t_2) \quad (33)$$

so that Eq. (32) becomes

$$z(t_1) = H\Phi(t_2, t_1)x(t_2) + w \quad (34)$$



That is, the measurement  $z(t_1)$  becomes a measurement of the state at the epoch time, but with the measurement matrix given by  $H\Phi$  instead of just  $H$ . It is important to realize that this use of mini-batches can change the filter results because it assumes that the effect of the process noise over the timespan of the mini-batch is negligible. The code that was developed allows the user to select one or more mini-batches per day, or to process each measurement as it is received.

The filter/smoothen was implemented in FORTRAN-77, making use of routines from the Estimation Subroutine Library (ESL) to do most of the numerical computations. The UD form of the Kalman filter, as described in Section IV was used for the filter; the smoothen made use of Bierman's modification of the RTS smoothen [2]. The general structure of the program is as follows:

- (1) Read a namelist file to initialize the filter:
  - Start and stop times
  - Maximum number of measurements to process
  - Initial error covariances and state estimate
- (2) For each measurement:
  - If new minibatch,
    - (a) Do time update
    - (b) Print estimates and estimate error standard deviations
    - (c) Translate measurements to epoch time and do the measurement update; write the smoothen gains to a file for later use
- (3) After all measurements have been processed, print out the correlation coefficients
- (4) If the smoothen has been requested, run the smoothen back to the start time

The results of Section VI were obtained by running the program with 6 stations, 11 station pairs, and 6 satellites. Hence the filter state vector has 84 components. An important feature of the filter implementation is the data editing capability. Before each measurement is actually included into the estimate, a test of the measurement's consistency with past measurements is performed. This is done as follows. First the residual (or innovations) as defined by Eq. (21) is computed, as is the variance of this residual which is just the quantity  $B_k$  defined by Eq. (18).

If the square of the residual divided by its variance is less than a user-specified tolerance, then the measurement is accepted and used to update the current estimates. For the results described in Section VI, a tolerance of 400 was used. This means that a measurement was rejected only if its residual was more than 20 standard deviations away from its mean value of zero.

## VI. Numerical Results

The numerical results described in this section were obtained by running the filter and smoothen on the data described in Section II. Approximately 10 data points were rejected by the data editor that is built into the filter. Figure 2 gives an overview of the results of the smoothen. This figure shows the offsets from UTC (NIST) of the three clocks in the DSN. This is a useful display for the operations analyst. It can be used to tell when adjustments to the rates are in order. Notice that a rate adjustment would have been appropriate around MJD 47080; however an adjustment was not made until MJD 47130. The Allan variances of the smoothen results, shown in Fig. 3, are consistent with those reported in [4], but are somewhat better. It is felt that these variances can be further improved with better modeling.

In order to contrast the output from the smoothen with the input to the filter, see Fig. 4(a). This shows the raw data for the California clock, with the smoothen results superimposed. A rate of  $1.2E-13$  was removed from both plots. The smoothen nicely bridges a gap in the data at about MJD 47050. During this gap, the standard deviations computed by the smoothen increased from about 4 nanoseconds, on MJD 47046, to 40 nanoseconds on MJD 47052, then dropped back to 4.4 nanoseconds on MJD 47053. Other data gaps were also handled well. This ability of the filter/smoothen to bridge data gaps will be important for operational uses. Figure 4(b) shows a superposition of the results produced by the JPL smoothen and the NIST filtered estimates. Again, the  $1.2E-13$  rate has been removed from both results. The agreement is quite good, with less than 8 nanoseconds difference. The consistent bias, with the JPL results slightly lower, may be a result of the way in which the space vehicle biases are handled.

The offset rate of the complex clocks is the most important parameter from the operational point of view. Presently, the rate is determined by taking the mean of all the GPS measurements for ten days, then doing a least-squares linear fit and using the slope of that fit as an estimate of the rate.

The results of these "hand calculations" are shown as dots in Fig. 5. Generally, there is good agreement between these results and those produced by the filter/smoothen. A careful examination of the times at which the hand calculations differ noticeably from the smoothen results revealed two reasons for these discrepancies. For example, in the vicinity of MJD 47060-47070, there was very little data for the clock in Spain. This caused the smoothen to report a large uncertainty in its rates over this interval, and explains the difference between the smoothen output and the hand calculation shown in Fig 5(c). The point obtained by the hand calculations for the clock in Australia on MJD 47070 was found, on closer examination, to have been calculated incorrectly. The correctly calculated value is very close to that produced by the smoothen.

## VII. Conclusions and Future Work

The UD form of the Kalman Filter appears to solve many of the problems encountered in the past with processing clock data on a routine basis. The numerical stability of the filter/smoothen provides an accurate and reliable method for computing optimal estimates of the offsets, offset rates, and drift rates. The automatic data editing feature removes bad data points appropriately.

The first several days worth of the data that were used for this study seemed to be very poor. For this reason, data corresponding to the first four days were skipped. As a comparison, this same data was also processed by using a Square Root Information Filter (SRIF). The SRIF is an even more numerically reliable form of the Kalman Filter. This increased reliability, however, is attained at the cost of an increase in computational time. The SRIF was able to handle the bad data at the beginning of the data set with no problems. Since computer run-time is of little concern for this application, it may be advisable to replace the UD filter with a SRIF.

Possible improvements to the mathematical model used in the filter would be of interest. For example, it is clear that the assumption about the biases being constant is incorrect. An improved model would allow the biases to be time varying. The rate at which they should vary, however, would have to be determined by statistical parameter estimation techniques, such as maximum likelihood estimation. Also, in the current model, all clocks are treated as if they were identical. In fact there are three different types of clocks in the DSN. Special characteristics of the clocks could be taken into account to improve the accuracy of the model.

## References

- [1] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, New York: Academic Press, 1977.
- [2] G. J. Bierman, "A Reformulation of the Rauch-Tung-Streibel Discrete Time Fixed Interval Smoother," *IEEE Conference on Decision and Control*, Austin, Texas, pp. 840-844, December 1988.
- [3] P. A. Clements, A. Kirk, and R. Unglaub, "Results of Using the Global Positioning System to Maintain the Time and Frequency Synchronization in the Deep Space Network," *Proceedings of the Eighteenth Annual Precise Time and Time Interval Applications and Planning Meeting*, pp. 395-405, December 1986, (also published in *TDA Progress Report 42-89*, Jet Propulsion Laboratory, Pasadena, California, pp. 67-72, May 15, 1987).
- [4] D. A. Howe, "Progress Toward One-Nanosecond Two-way Time Transfer Accuracy Using Ku-Band Geostationary Satellites," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. UFFC-34, no. 6, pp. 639-646, November 1987.
- [5] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum Likelihood Estimates of Linear Dynamic Systems," *AIAA Journal*, vol. 3, no. 8, pp. 1445-1450, August 1965.

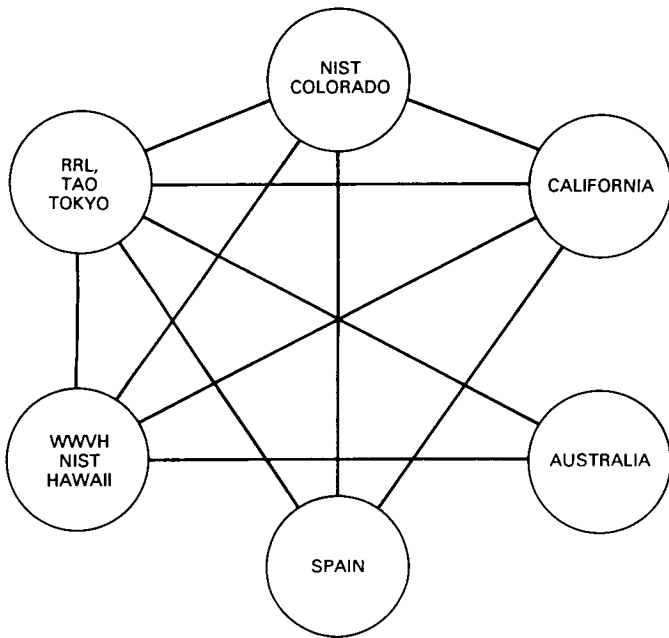


Fig. 1. Eleven possible mutual views used in the DSN time and frequency coordinates.

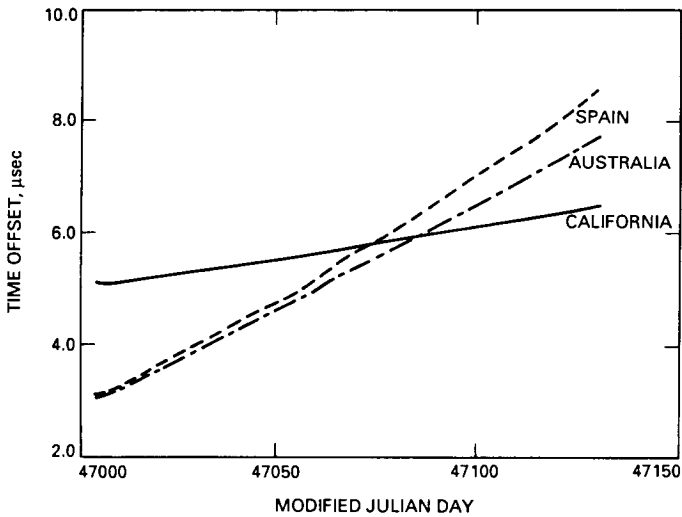


Fig. 2. Time-offset outputs of the UD smoother: DSN clock time minus VTL (NIST) clock time.

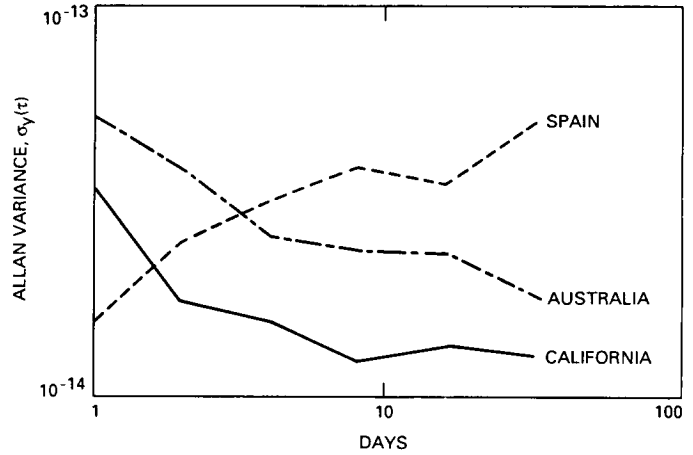


Fig. 3. The Allan variance of the time offsets from the UD smoother.

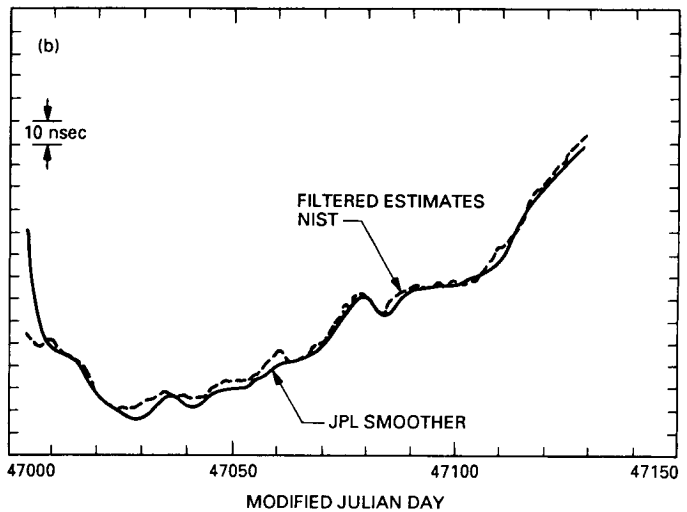
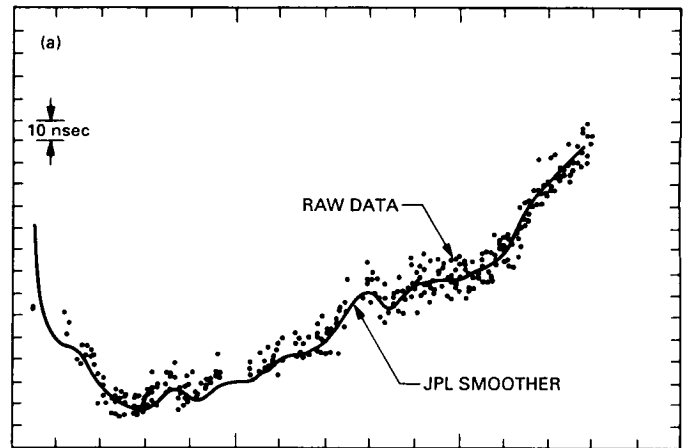
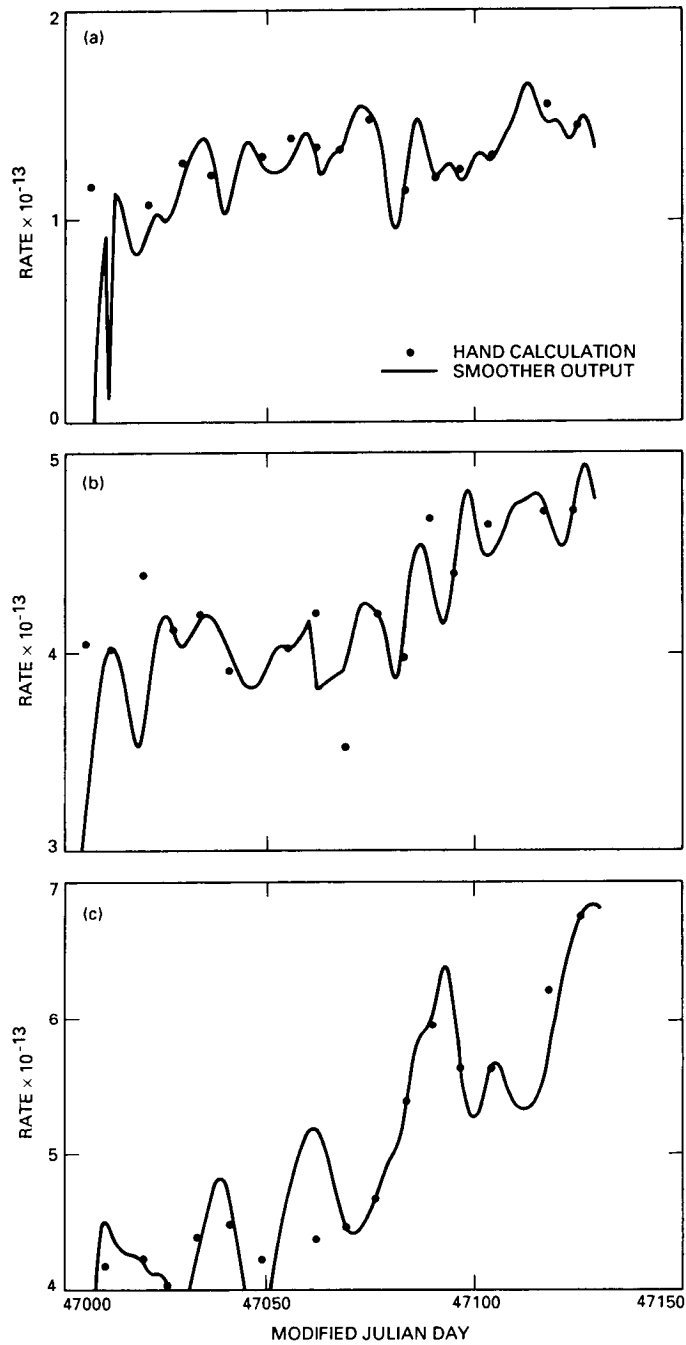


Fig. 4. Smoother output with a rate of  $1.2 \times 10^{-13}$  removed: (a) the smoothed estimate and the raw data; (b) the smoothed estimates compared to NIST estimates.



**Fig. 5. Kalman smoother output and hand calculation of rate using a linear least-squares fit over 10 days of the time offset: (a) California, (b) Australia, and (c) Spain.**

516-38  
237826

547682  
PS

N90-12802

TDA Progress Report 42-98

August 15, 1989

68.

# A Simple Model for DSS-14 Outage Times

H. C. Rumsey,<sup>1</sup> R. Stevens, and E. C. Posner  
Office of Telecommunications and Data Acquisition

*A model is proposed to describe DSS-14 outage times. Discrepancy Reporting System outage data for the period from January 1986 through September 1988 are used to estimate the parameters of the model. The model provides a probability distribution for the duration of outages, which agrees well with observed data. The model depends only on a small number of parameters, and has some heuristic justification. This shows that the Discrepancy Reporting System in the DSN can be used to estimate the probability of extended outages in spite of the discrepancy reports ending when the pass ends. The probability of an outage extending beyond the end of a pass is estimated as around 5 percent.*

## I. Introduction

A model is proposed to describe DSS-14 outage times. Outage data for the period from January 1986 through September 1988 are used to estimate the parameters of the model. The model provides a probability distribution for the duration of the outages. The model does not address questions about the mean time between outages. However, it does allow estimation of the probability of major outages even though the Discrepancy Reporting (DR) system stops recording them at the end of a pass. The philosophy has been that the best model is the simplest one that fits well enough.

The nature of the model, as above, is affected by the way outages are reported. If an extended outage occurs,

the time to restore service is not reported. Only the time lost for that pass is recorded. This limitation makes it difficult to determine the actual outage durations from the DR data and must be accounted for in the model. In the model, it is assumed that the actual time to restore service is hidden by a "cutoff" process which corresponds to the end of the pass. A typical pass runs for 9 hours. This means that the actual time to restore service is masked by a 9-hour cutoff window. (In developing the model, 8-, 9-, and 10-hour cutoffs were tried, with 9 fitting best and being reasonable on other considerations.)

## II. Outage Distribution

Let  $R(t)$  be the distribution function for *reported* outage durations, and let  $A(t)$  be the distribution function for *actual* outage durations. Assuming the "cutoff" process and outage durations are independent, then for  $t \geq 0$ ,

---

<sup>1</sup> Consultant to the TDA Office from Laguna Niguel, California.

$$\begin{aligned} \Pr(\text{reported time} \geq t) \\ = \Pr(\text{actual time} \geq t)\Pr(\text{cutoff time} \geq t) \end{aligned}$$

or

$$1 - R(t) = (1 - A(t))(1 - t/U)_+ \quad (1)$$

where  $U$  is the duration of the pass (nominally 9 hours) and  $(1 - t/U)_+$  is 0 for  $t > U$ . Equation (1) assumes that the start time of an outage is uniformly distributed over the duration of the pass.

Equation (1) deals with the “cutoff” problem but says nothing about the actual distribution of outages  $A(t)$ . The model for  $A(t)$  is based on the reported outage data and a desire to minimize the number of parameters in the model. Figure 1 shows this measured outage data from the DR system, accurate to one minute. Thus, some minutes show multiple outages. There are 498 outages presented in this figure. The mean time to restore service is about 40 minutes.

The simplest model to fit such data would be an exponential distribution. The distribution function of an exponential random variable  $X$  of a mean  $a$  is given by

$$F(t) = \Pr(X \leq t) = 1 - e^{-t/a}$$

The fit is not accurate for short outages because Fig. 1 shows the density goes to 0 at 0-length outages, but an exponential has *maximum* density at 0-length outages. An exponential random variable with mean around 40 minutes fits the first part of the outage data fairly well beyond a few minutes up to about 100 minutes, but there are too many extreme values ( $t \geq 100$  minutes) in the outage data. This suggests that there are two (or more) classes of failures, in addition to the short failures. For the first class of long failures, service can be restored quickly, in less than 40 minutes on the average. The second class of long failure requires more time to overcome.

### III. Model

The following model has been adopted, with three parameter  $\alpha$ ,  $a$ , and  $b$  to be estimated, since  $U$  is 9 hours and not estimated:

$$\begin{aligned} \Pr(\text{reported time} \geq t) &= 1 - R(t) \\ &= ((1 - \alpha)(1 + t/a)e^{-t/a} \\ &\quad + \alpha(1 + t/b)e^{-t/b})(1 - t/U)_+ \end{aligned} \quad (2)$$

The form for this tail distribution has been taken to make the density function (essentially) zero at  $t = 0$ . Observe that if  $T_a$  is a random variable with distribution function  $1 - (1 + t/a)e^{-t/a}$ , then the expected value of  $T_a$  is  $2a$  while the maximum of the density function of  $T_a$  occurs at  $t = a$ .

The parameters  $a$  and  $b$  occur symmetrically in Eq. (2). If  $a$  is chosen to be the smaller value, then outages of class “ $a$ ” can be arranged to peak around  $t = 10$  minutes. Outages of class “ $b$ ” can be arranged to fit the tails of the observed outages. More heuristics appear in Section VI.

### IV. Parameter Estimation

Equation (2) defines the model. To complete the model, good values for the parameters  $U$ ,  $\alpha$ ,  $a$ , and  $b$  must be found. The maximum likelihood method will be used to estimate the parameters. Let  $(t_1, \dots, t_n)$  be the reported outage times. The  $t$ s are reported by the DR system to the nearest minute. Let  $t$  be one of the outage times. The models specify the probability  $p(t)$  that an outage has duration  $t$ , where  $t$  is measured in minutes. Namely,

$$p(t) = R(t + 1/2) - R(t - 1/2)$$

where  $R$  is determined by Eq. (2). The probability that the observed outages occurred is the product of the probabilities of the separate outages,

$$\prod_{j=1}^n p(t_j)$$

This product is the likelihood function of the observations. It is a function of the model parameters  $U$ ,  $\alpha$ ,  $a$ , and  $b$ . Maximum likelihood says to choose these parameters to maximize this product. The maximization is easy in this case because there are so few parameters.

There is a minor problem in determining the best value for the parameter  $U$ . Recall that  $U$  is the cutoff time for the pass. Finding  $U$  is like finding the end point of an interval  $(0, U)$  in which a uniform random variable occurs. A little thought shows that the likelihood function for such a problem is maximized by taking  $U$  to be as small as possible. In the case of interest here, this would correspond to taking  $U$  to be slightly less than 8 hours (the maximum reported outage is 462 minutes). If extreme outages were common, this "defect" in the maximum-likelihood method would be no problem. Yet, the number of extended outages is small. So rather than estimate  $U$  from the likelihood function,  $U$  has been taken to be 9 hours throughout. To see how this selection affects the results,  $U = 8$  and 10 hours for the case  $F_b \equiv 1$  were also tried. As expected,  $U = 8$  hours gives a larger value for the likelihood function, but the other parameters  $\alpha$  and  $a$  are hardly changed. The results for  $U = 10$  hours were not as good. Only  $U = 9$  hours is considered below. This is consistent with the known distribution of pass lengths.

## V. Goodness of Fit

A grid of points was used to find the maximum-likelihood values for the parameters. The maximum-likelihood values found were

$$\alpha = 0.186$$

$$a = 11.4 \text{ minutes}$$

$$b = 77.5 \text{ minutes}$$

The corresponding distribution function is then given by

$$\text{Pr}(\text{reported outage} \geq t) =$$

$$\left[ 0.814(1 + t/11.4)e^{-t/11.4} + 0.186(1 + t/77.5)e^{-t/77.5} \right] (1 - t/540)_+$$

This equation represents the model. To see how well this model fits the observed outage data, the outage data was smoothed with a 5-point smoothing filter. The same filter was applied to the model as well. The results are dis-

played in Fig. 2. Qualitatively, the model fits the observed outages very well, for short, medium, and long outages.

The mean time MTR to restore service for the model is given by

$$\text{MTR} = (1 - \alpha)aT(U/a) + \alpha bT(U/b)$$

where

$$T(x) = 2 - 3/x + (1 + 3/x)e^{-x}$$

Substituting for  $\alpha$ ,  $a$ ,  $b$ , and  $U = 9$  hours (540 minutes) as always gives

$$\text{MTR} = 40.6 \text{ minutes}$$

This is in excellent agreement with the observed MTR of 40.4 minutes.

The probability that an outage exceeds 150 minutes was computed. With 498 outages, the model predicts there should be 28.4 outages of duration 150 minutes or longer. In the actual data of 498 outages there were 24 outages of this duration. Since the extended outage statistics are expected to be Poisson with an estimated mean of 28.4 and thus a sigma of  $\sqrt{28.4} = 5.33$ , the discrepancy of 4.4 is less than one sigma. This fits as well as could be expected with only 24 events. The probability of short and medium outages is also seen to fit very well. Hence, the use of the model seems indicated.

## VI. Heuristics

It can be expected that the density of outages near zero outage time is very nearly 0, because it takes some minimum time to notice an outage and to respond to it, even by switching in a hot standby automatically. The  $(1 + t/a)e^{-t/a}$  term in Eq. (2) does just this—it has density 0 at 0, for the corresponding density is  $(t/a^2)e^{-t/a}$ . This density also covers intermediate outages, but so does the "b" class. One might think that yet another distribution should be mixed in to cover these intermediate outages that cannot be recovered merely by switching something



in. But as has been seen, adding an extra one or two parameters is not necessary—the fit is very good with just the single time parameter  $a$ , the location of the maximum density of the short and medium outages, and the additional parameter  $\alpha$  which gives the relative fraction of long outages.

The long outages are described by a similar distribution  $(1 + t/b)e^{-t/b}$  with  $b \gg a$ . Why the tail of this and the shorter-outage distribution should be exponential is less clear, but the fit is good, and hard to tell from a distribution with a long low constant tail given the amount of data. It can be observed that the form of the  $a$  and  $b$  distribution arises as a difference of pure exponentials with infinitesimally close memoryless repair rates, but this does not seem to help the heuristics. The long tail can arise from certain failures such as low-noise maser warm up that takes a certain minimum time, e.g., 12 hours to recover from, when hot standbys for switching in are not provided.

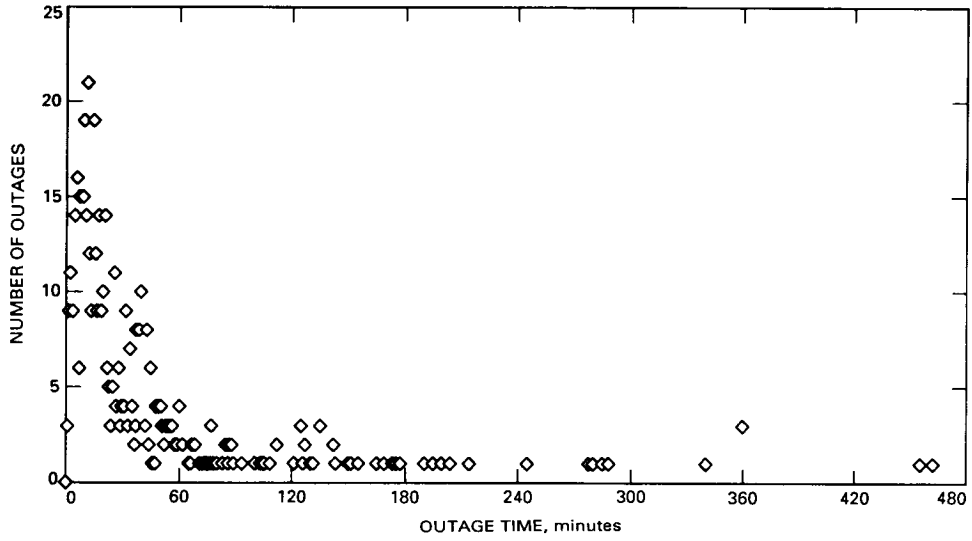
Finally, as explained, the  $(1 - t/U)_+$  multiplier term in Eq. (2) arises from the truncation of outage data at the end of a pass, where it was assumed that failures occur uniformly over the duration ( $U = 9$  hours) of a pass. It is this truncation which makes it hard to distinguish a negative exponential tail from a long flat tail. The three-parameter model has been adopted even though the heuristics are not perfect.

## VII. Summary

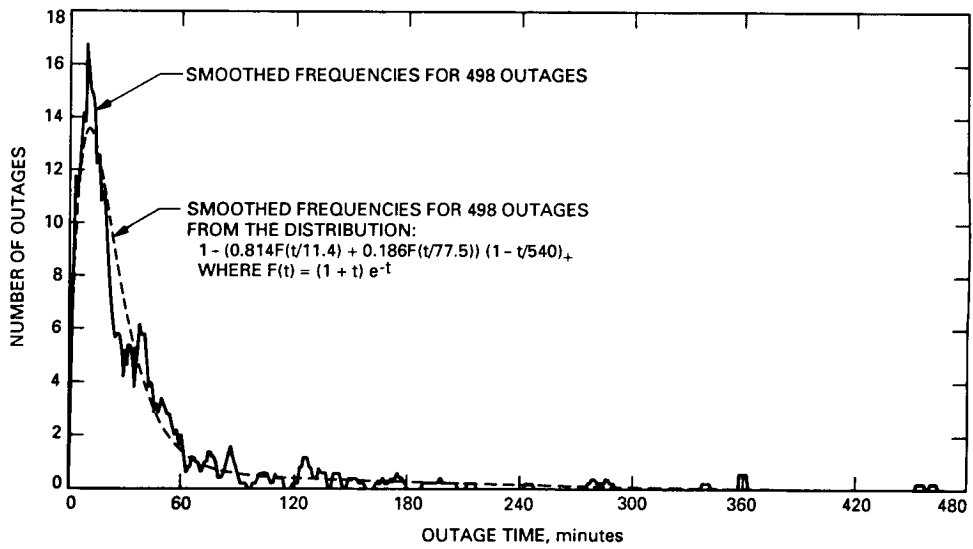
It has been shown that the Discrepancy Reporting System in the DSN can be used to give good estimates of the probability of extended outages in spite of the discrepancy reports ending when the pass ends. The probability of a major outage (one extending beyond the end of a pass) is estimated by the best-fit model as around 5 percent. The model also gives good estimates for the probability of short, medium, and long outages. It is simple and yet fits very well.

## Acknowledgments

We thank D. W. Ginavan and G. C. Smith of the Bendix Field Engineering Corp. for organizing and providing the outage data used in this study.



**Fig. 1. DSS-14 outage times, January 1986–September 1988.**



**Fig. 2. Smoothed DSS-14 outage times, January 1986–September 1988 compared with model.**

547683  
P6  
57-32  
N90-12803  
23-27  
August 15, 1989  
118

August 15, 1989

# Ka-Band MMIC Beam Steered Transmitter Array

D. L. Rascoe, A. L. Riley, J. Huang, V. Lubecke, and L. Duffy  
Spacecraft Telecommunications Equipment Section

*A 32-GHz six-element linear transmitter array utilizing Monolithic Microwave Integrated Circuit (MMIC) phase shifters and power amplifiers has been designed and tested as part of the development of a spacecraft array feed for NASA deep-space communications applications. Measurements of the performance of individual phase shifters, power amplifiers, and microstrip radiators were carried out, and electronic beam steering of the linear array was demonstrated.*

## I. Introduction

Downlink communication systems for NASA deep-space missions presently operate at X-band (8.5 GHz). However, in the mid-1990s, advanced deep-space missions may utilize Ka-band systems (32 GHz downlink, 34 GHz uplink) to achieve communications enhancement on the order of 8 dB. A receiver Monolithic Microwave Integrated Circuit (MMIC) array at Ka band has been reported [1], but a transmitter array, a critical element in these systems, has not.

At JPL, a 32-GHz solid-state transmitter is under development utilizing state-of-the-art GaAs MMIC devices. The initial goal of this work is to produce a Ka-band planar phased array with 5-W output power to feed a 4-m reflector system. The electronic beam-scanning capability of the phased array will allow the narrow beam ( $\approx 0.2$  deg) of the large reflector to perform fine beam pointing toward the Earth ( $\pm 1$  deg). As a first step in this effort, a linear array composed of six subarrays of microstrip radiators, six MMIC switched-line phase shifters, and six MMIC power amplifiers was designed and tested. The MMIC devices were developed under programs funded by

the NASA Lewis Research Center. This article reports the performance results and mathematical modeling of the JPL linear transmitter array incorporating these devices.

## II. MMIC Devices

The MMIC phase shifters were designed and fabricated on GaAs by the Honeywell Sensors and Signal Processing Laboratory [2, 3]. A photograph of a phase shifter is shown in Fig. 1. A single phase shifter contains 12 MESFETs in a three-bit switched-line configuration and 2 MESFETs in an analog loaded-line configuration. The switched-line circuit provides nominal phase changes of 45, 90, and 180 deg. The loaded-line design provides a nominal continuous phase variation from 0 to 45 deg. These devices were originally designed for operation at 30 GHz for NASA's ACTS program rather than at the slightly higher Deep Space Network frequency of 32 GHz.

The MMIC power amplifiers were designed and fabricated by the Texas Instruments Central Research Laboratory [4]. A photograph of an amplifier is shown in Fig. 2.

The devices used in the array were two-stage amplifiers utilizing GaAs MESFETs with a 0.25- $\mu\text{m}$  gate length and gate widths of 0.1 mm and 0.3 mm for the first and second stages, respectively. The design goal of this device was to produce 20 dBm of output power.

### III. Array Design

The linear phased array, shown in Fig. 3, is composed of four major "building blocks:" (1) a six-way microstrip power divider, (2) a carrier strip of six MMIC phase shifters, (3) a carrier strip of six MMIC power amplifiers, and (4) an antenna array of 6 pairs of microstrip patch radiators. All blocks were separately fabricated and tested to allow evaluation of the performance of each device and for building a detailed model of the linear array. This "building-block" design approach of the array also allows for ease of replacement of any of the major components. Thus as improved MMIC devices become available or evaluation of different antenna elements is desired, the devices or elements may be readily integrated into this test-bed.

The RF input to the array is through a WR-28 waveguide-to-microstrip transition of the Van Heuven [5] antipodal finline type. The double-sided transition pattern is etched on 0.25-mm-thick Rogers 5880 Duroid and attached to the gold-plated aluminum housing using Indalloy 121 solder.

The six-way microstrip power divider consists of a set of 90-deg branch-line couplers with compensating lengths of microstrip interconnecting transmission lines to produce equal amplitude and phase at all six outputs. A close-up of the divider is shown in Fig. 4. The divider pattern was etched on 0.12-mm-thick Rogers 5880 Duroid laminated on an aluminum carrier. Each coupler was terminated by a 50- $\Omega$  TaN thin-film chip resistor connected to an approximately quarter-wave-length radial line open-circuit stub. The power divider was designed using the EEsof Touchstone simulation and MiCAD layout systems. A table of simulated versus measured data for all six ports at 32 GHz is shown in Table 1.

The MMIC carrier strips were designed to permit measurement of each MMIC device and to interface with the array. Individual devices were measured by attaching waveguide-to-microstrip transitions to the carrier as shown in Fig. 5. The carrier strips consisted of laser cut, 0.25-mm-thick alumina substrates with etched TiW/gold circuit metalization that were soldered to gold-plated kovar carrier strips. The dc and control lines were brought

through the carrier and substrate via miniature coaxial feedthroughs mating on the underside with dual in-line pair (DIP) socket connectors. The MMICs were attached to the carriers with silver epoxy and wire bonded to the alumina substrates with 25- $\mu\text{m}$ -diameter gold wire.

The antenna array, consisting of six pairs of microstrip radiators that radiate in a direction normal to the plane of the dielectric, was designed using the multimode cavity theory [6]. This theory assumes that the electric field in the cavity underneath the patch can be precisely modeled by a series of cosine modal functions. Since the fundamental mode, as well as the higher-order modes, can all be included in the analysis, accurate prediction of the resonant frequency, copolarization and cross-polarization radiations, input impedance, and bandwidth effect can be achieved.

The antenna array was etched on a 0.25-mm Rogers 5880 Duroid substrate. Each etched radiator pair consists of two rectangular microstrip patches combined by a two-way reactive power combiner. A typical input return loss for a two-patch array, shown in Fig. 6, at 32 GHz was nominally -20 dB with 2.5 percent bandwidth. The separation between radiator pairs was 1.08 free-space wavelengths. This spacing was chosen to allow ease of assembly so that DIP sockets could be used at each MMIC position. Since only  $\pm 10$  deg of beam scan is needed for the array to achieve the required reflector fine beam pointing of less than 1 deg, the relatively wide subarray spacing did not generate any serious grating lobe problems. Figure 7 is a calculated far-field pattern [7] of the 12-patch linear array with a 10-deg beam scan and 1.08-wavelength subarray spacing.

### IV. Test Results

To ensure that the microstrip array gives adequate performance, the six pairs of patches were connected directly to the six-way power divider without the amplifiers and phase shifters. The far-field pattern was measured and compared with the calculation as shown in Fig. 8. It demonstrated excellent pattern performance. The measured array gain, excluding the power divider loss, was 15.9 dBi, which indicates an efficiency of 83 percent. Individual measurements of single and double patches preceded the 12-patch array measurement, and Table 2 shows a comparison of the radiator efficiencies in these three cases.

The MMIC devices were measured on an extended HP8510 at 32 GHz in the carriers. Phase-shift measurements were referred to the nominal zero phase state of

each device and attenuation was referred to back-to-back waveguide-to-microstrip transitions. Figure 9 is a plot of nominal phase versus measured phase for six phase shifters. The devices used in the array were from two wafers, each made with slightly different mask sets. The mean and standard deviation of the phase shift was  $7.0 \pm 1.6$  percent over the range 0 to 300 deg. Attenuation as a function of nominal phase angle is shown in Fig. 10. The mean attenuation at zero phase shift is 10.4 dB for the first group and 8.0 dB for the second. The attenuation changed at a rate of approximately 0.8 dB/100 deg and 2.8 dB/100 deg for the two groups. For phase settings larger than 250 deg, two units fail to track the other four, and their attenuations become larger. This large variation in attenuation performance is partially explained by the fact that these devices were originally designed for a lower frequency, 30 GHz, and for another application. Despite this variation in attenuation, however, reasonable array beam steering was obtained.

The gain versus input power of the MMIC power amplifiers is shown in Fig. 11. The six amplifiers were fabricated from the same wafer and have a low power-gain mean and standard deviation of  $13.7 \pm 1.7$  dB. The mean and standard deviation of output power at 1-dB gain compression is  $13.6 \pm 0.3$  dBm. Several amplifiers were externally tuned and were found to be capable of producing as much as 21.5 dBm with an efficiency of 14.5 percent at 1-dB gain compression.

The complete integrated linear phased array was tested by measuring its beam steering pattern. This pattern, shown in Fig. 12, was measured with a different technique from the conventional method. Both the receive antenna, which was a 10-cm reflector, and the test transmit array were placed in fixed positions facing each other. Then the array's phase shifters were adjusted to scan the beam at 1-deg intervals from  $-25$  deg to 25 deg. Both the

measurement system and the phase-shifter bias switching system were controlled by a PC. A program was written to set the phases based on keyboard entries of beam-steering angles. The pattern was then recorded with the received power versus the scan angles. Although the pattern was measured from  $-25$  deg to 25 deg, our primary interest is in the  $\pm 10$ -deg region. Nevertheless, excellent comparison between the calculated and measured results can be observed in Fig. 12. The relatively high sidelobes were caused by improper operation of one of the middle amplifier modules. The calculation included the effect of the failed module as well as randomized errors in both phase and amplitude. If all the six subarrays were functioning properly, the calculation predicted that the sidelobes are  $-12$  dB below the beam peak. The array had a beamwidth of 7.5 deg and demonstrated acceptable beam steering over  $\pm 8$  deg.

## V. Conclusions

A 32-GHz six-element linear transmitter array utilizing MMIC phase shifters and power amplifiers has been developed and tested as a precursor to the design of a two-dimensional array for use in a NASA spacecraft array feed. The switched-line phase shifters were accurate to within 7 percent on average and the power amplifier 1-dB compressed output power varied over 0.3 dB. The array had a beamwidth of 7.5 deg and demonstrated acceptable beam steering over  $\pm 8$  deg.

From the above results, it can be concluded that this MMIC phased array has adequate beam-scanning capability for use in the two-dimensional array. The areas that need to be improved are the efficiency of the MMIC power amplifier and the insertion loss of the MMIC phase shifter.

## Acknowledgment

The authors wish to acknowledge the contributions of C. Cruzan for his efforts in assembling the array.

## References

- [1] Honeywell Physical Sciences Center, *Electronically Steerable Patch Panel Array Antenna*, Final Report (Oct. 1, 1985–Sept. 30, 1986) AFATL-TR-87-19, USAF Armament Laboratory, Eglin AFB.
- [2] V. Sokolov et al., "A GaAs Monolithic Phase Shifter for 30 GHz Application," *1983 IEEE Microwave and Millimeter-wave Monolithic Circuits Symposium Digest*, pp. 40–44, 1983.
- [3] P. Bauhahn, C. Butter, V. Sokolov, and A. Contolatis, "30 GHz Multi-Bit Monolithic Phase Shifters," *1985 IEEE Microwave and Millimeter-wave Monolithic Circuits Symposium Digest*, pp. 4–7, 1985.
- [4] P. Sanier, H. Q. Tserng, N. Camilleri, K. Bradshaw, and H. D. Shih, "A High Efficiency Ka-Band Monolithic GaAs FET Amplifier," *GaAs IC Symposium Technical Digest 1988*, pp. 37–39, 1988.
- [5] J. H. C. Van Heuven, "A New Integrated Waveguide-Microstrip Transition," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-24, pp. 144–147, March 1976.
- [6] W. F. Richards et al., "An Improved Theory for Microstrip Antennas and Applications," *IEEE Trans. Antennas and Propag.*, vol. AP-29, pp. 38–46, January 1981.
- [7] J. Huang, "The Finite Ground Plane Effect on the Microstrip Antenna Radiation Patterns," *IEEE Trans. Antennas and Propag.*, pp. 649–653, July 1983.

**Table 1. Simulated versus measured data for six-way power divider**

Simulated								
Out Port	1	2	3	4	5	6	Avg.	$\sigma$
MAGS21	-10.6	-10.7	-10.6	-10.5	-10.6	-10.6	-10.6	0.1
ANGS21	-11.0	-9.0	-9.0	-10.0	-8.0	-8.0		
adjANG	0.0	1.0	1.0	1.0	2.0	2.0	1.0	1.0
Measured								
Out Port	1	2	3	4	5	6	Avg.	$\sigma$
MAGS21	-11.1	-10.9	-11.3	-11.9	-11.5	-11.8	-11.4	0.4
ANGS21	71.0	52.0	77.0	77.0	54.0	75.0		
adjANG	0.0	-19.0	6.0	6.0	-17.0	3.0	-4.0	11.0
Measured-Simulated								
Out Port	1	2	3	4	5	6	Avg.	$\sigma$
MAGS21	-0.5	-0.2	-0.6	-1.4	-0.9	-1.2	-0.8	0.4
ANGS21	0.0	-20.0	4.0	5.0	-20.0	1.0	-5.0	11.0

**Table 2. Ka-band microstrip antenna gain and efficiency**

Parameter	Single Patch	Two-Patch Array	12-Patch Array
Calculated directivity, dBi	7.3	10.7	16.7
Measured gain, dBi	6.9	10.2	15.9
Efficiency, percent	91	89	83

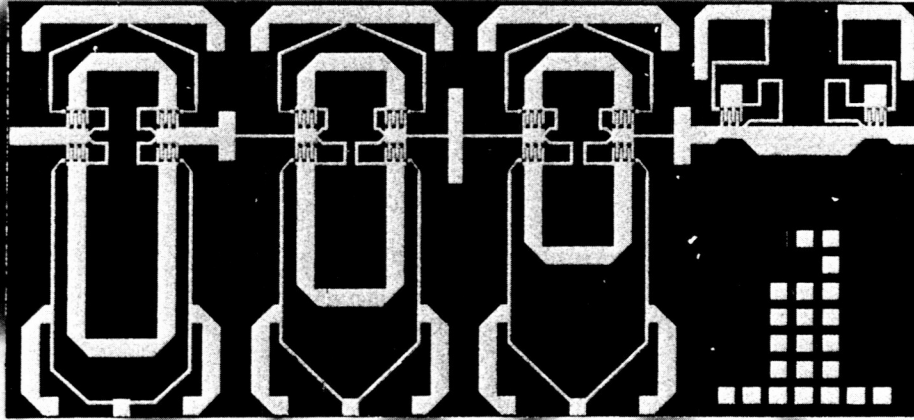


Fig. 1. Honeywell MMIC phase shifter.

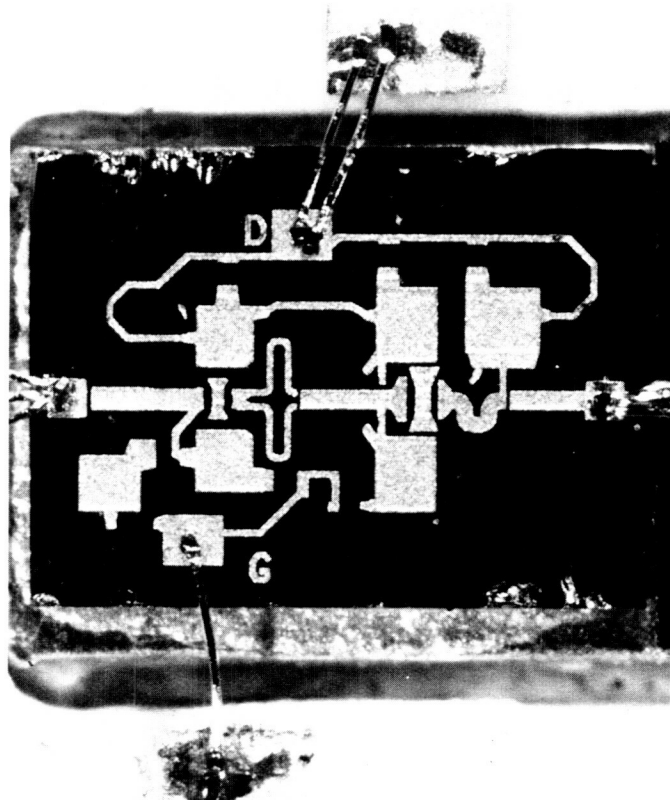


Fig. 2. Texas Instruments MMIC power amplifier.



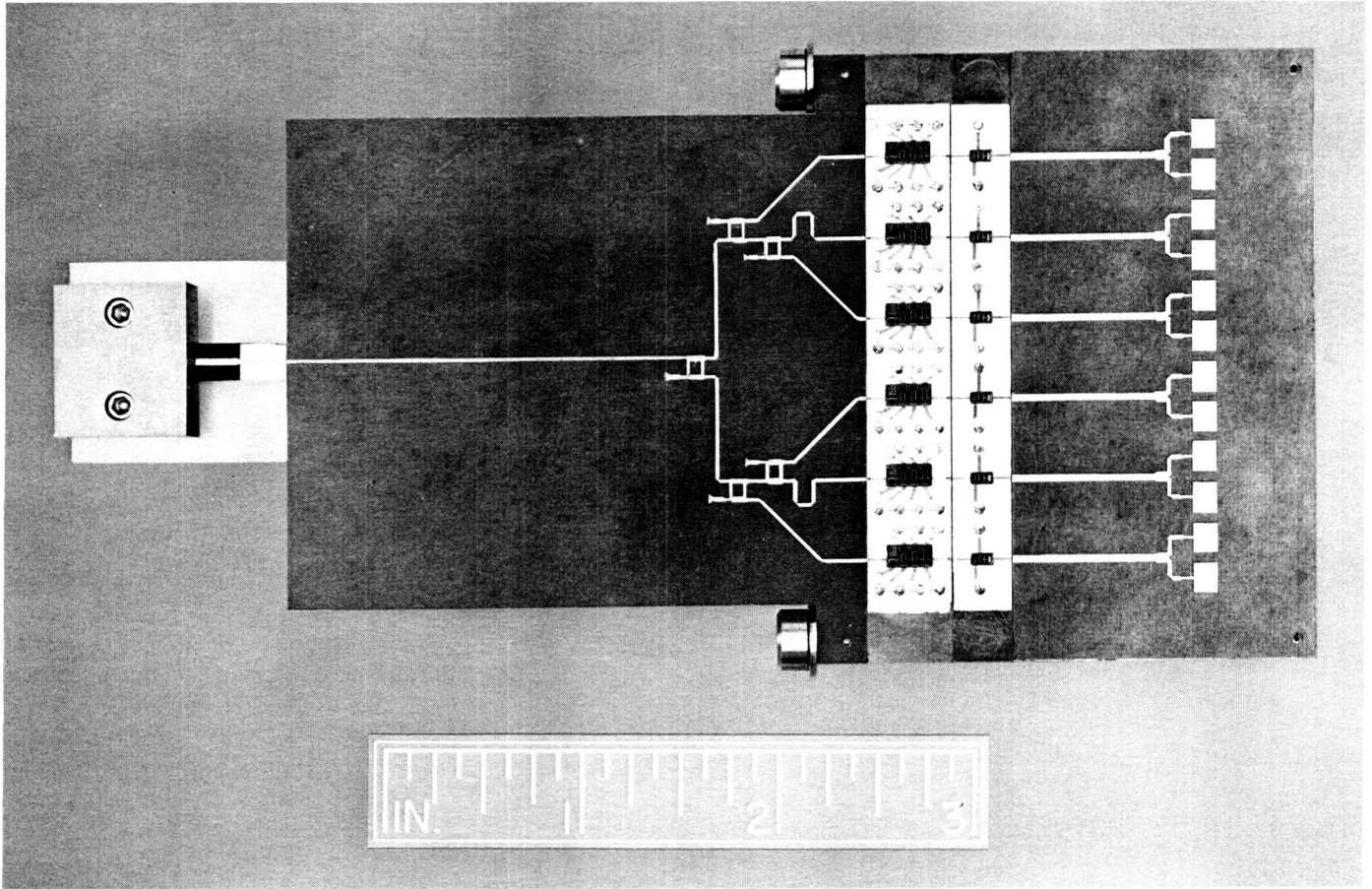


Fig. 3. Six-element linear transmitter array.

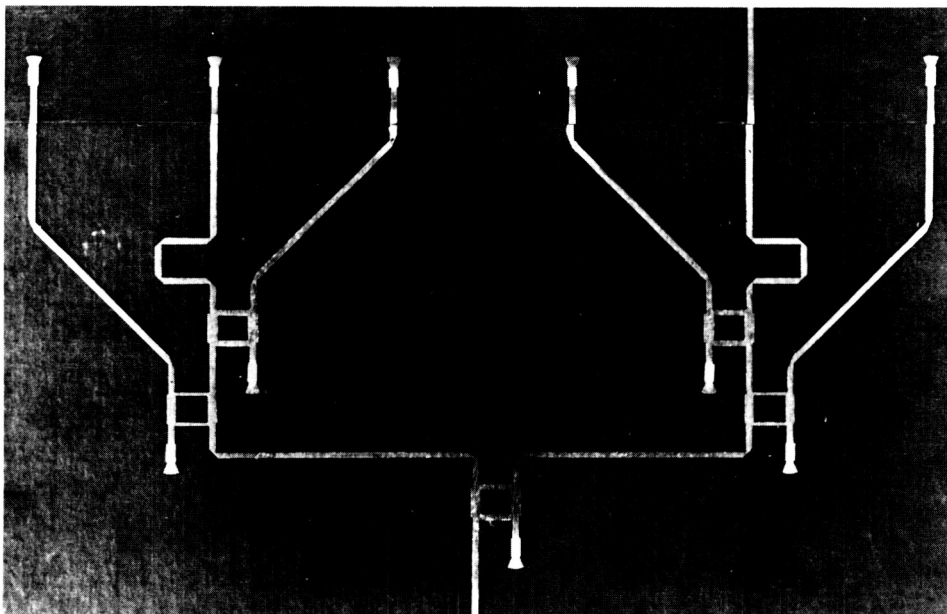


Fig. 4. Six-way microstrip power divider.

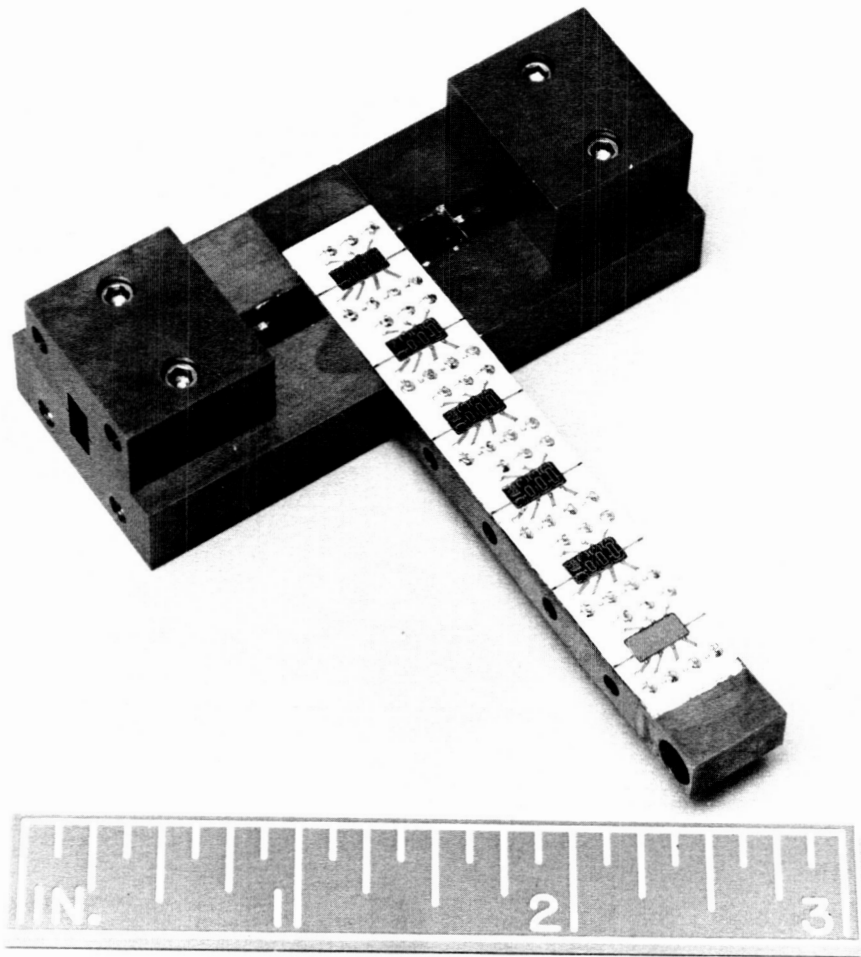


Fig. 5. MMIC phase-shifter carrier strip between two waveguide-to-microstrip transitions.

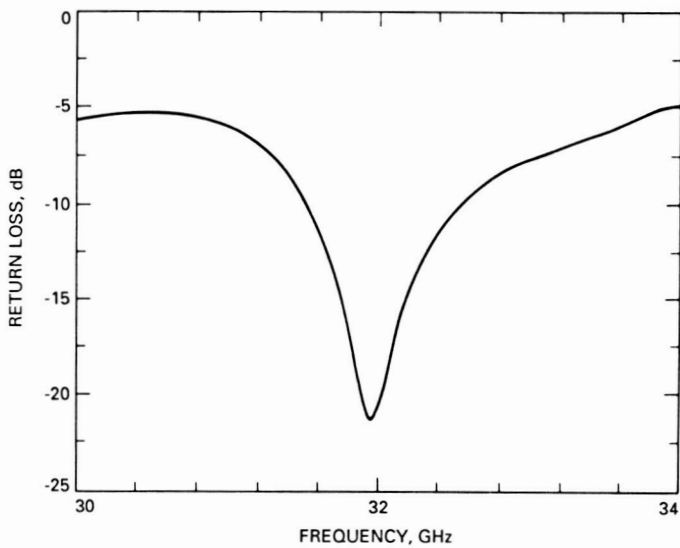


Fig. 6. Measured input return loss for a two-patch array.

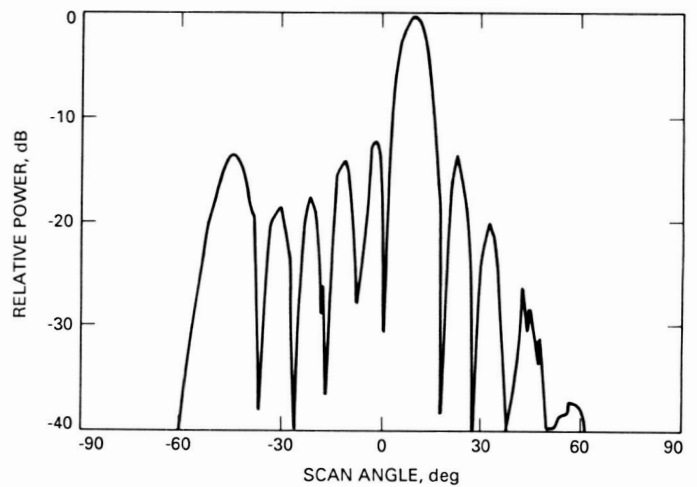
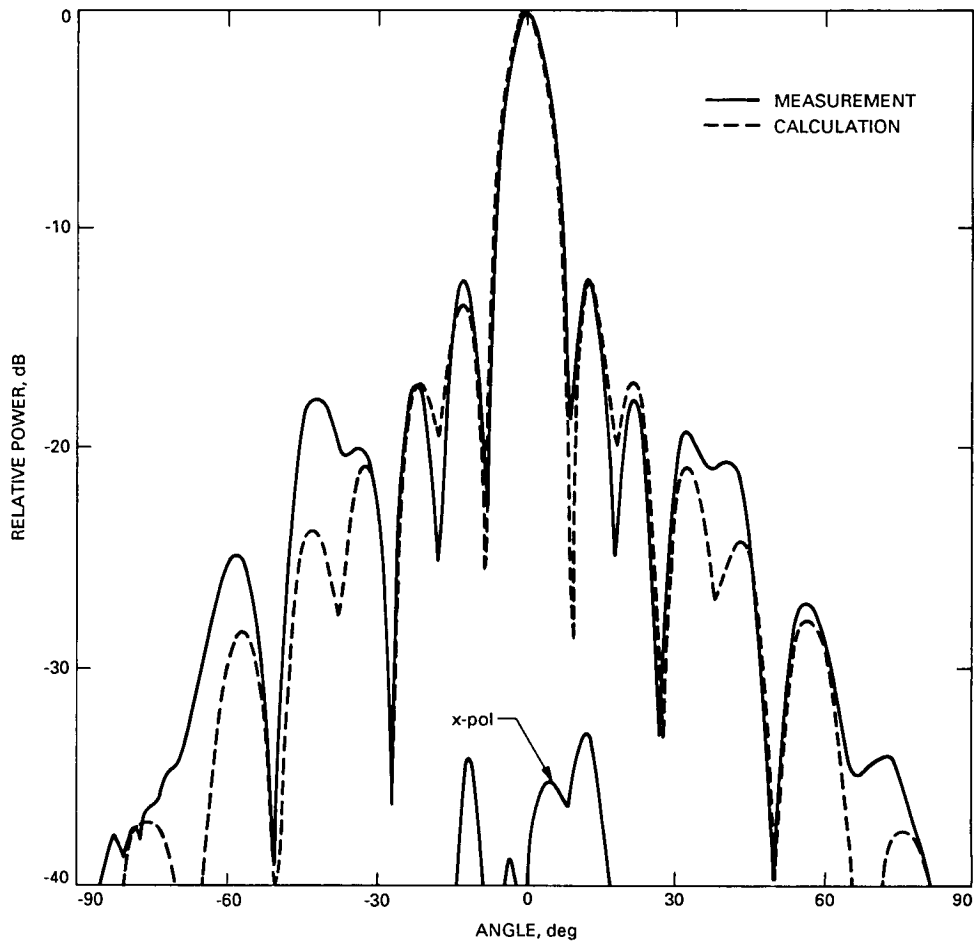
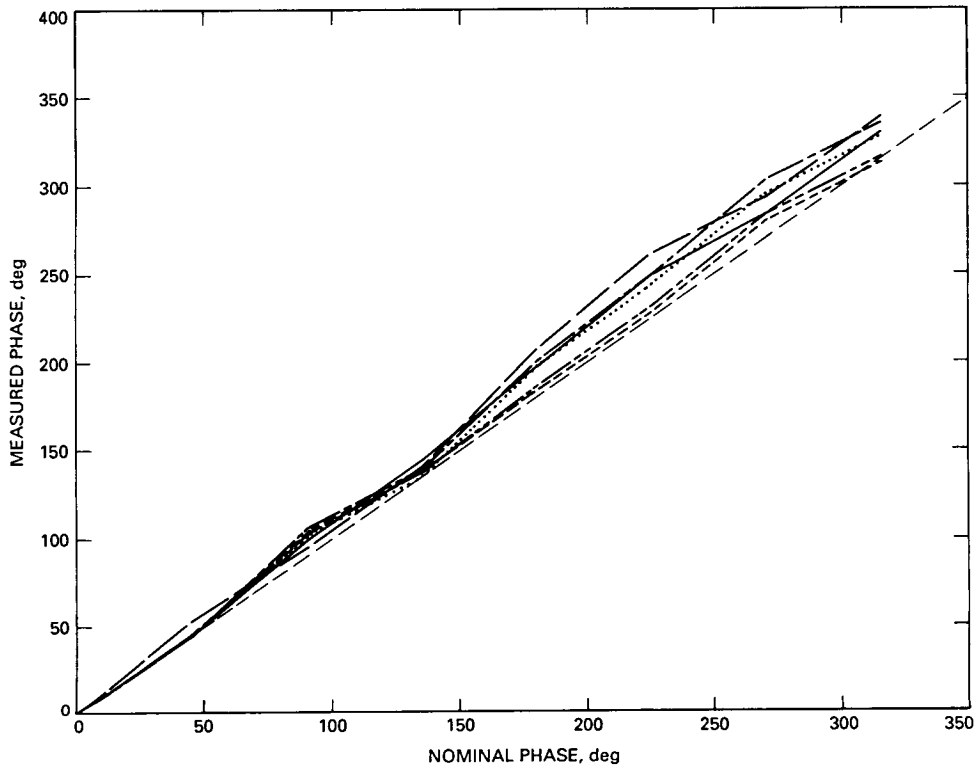


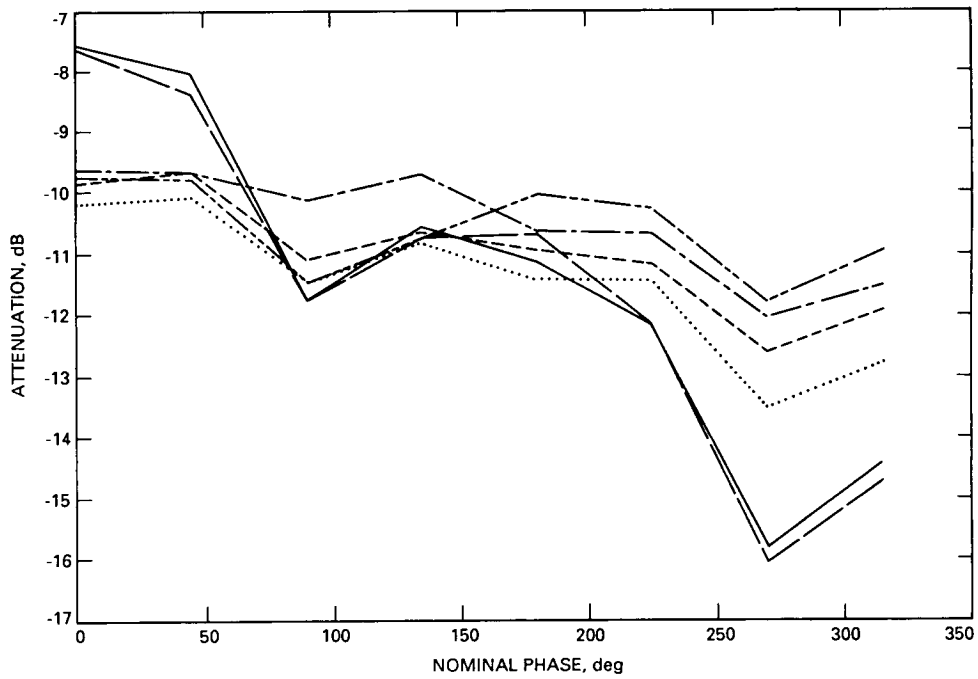
Fig. 7. Calculated far-field pattern of the 12-patch array with a 10-deg beam scan.



**Fig. 8. Comparison of calculated and measured far-field pattern for the linear array without MMIC carrier strips.**



**Fig. 9. Phase deviations measured against the nominal phase shift (dashed line) for six phase shifters.**



**Fig. 10. The spread in measured attenuation for six phase shifters.**

S18-55  
239823

547684 P10  
N90-12804

TDA Progress Report 42-98

August 15, 1989

98

# The NASA SETI Sky Survey: Recent Developments

M. J. Klein, S. Gulkis, and E. T. Olsen  
Space Physics and Astrophysics Section

N. A. Renzetti  
TDA Science Office

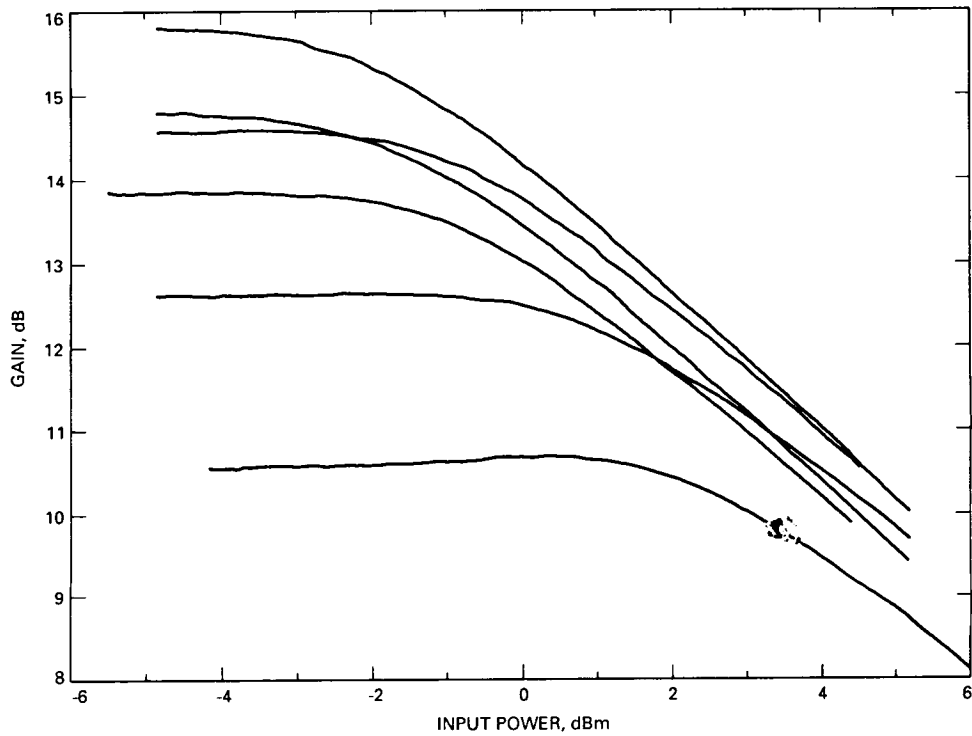
*NASA's Search for Extraterrestrial Intelligence (SETI) project utilizes two complementary search strategies: a Sky Survey and a Targeted Search. The SETI team at the Jet Propulsion Laboratory (JPL) in Pasadena, California, has primary responsibility to develop and carry out the Sky Survey part. This article describes progress that has been made developing the major elements of the survey including a 2-million channel wideband spectrum analyzer system that is being designed and constructed by JPL for the Deep Space Network (DSN). The new system will be a multiuser instrument; it will serve as a prototype for the SETI Sky Survey processor. This prototype system will be used to test the signal detection and observational strategies on DSN antennas in the near future.*

## I. Introduction

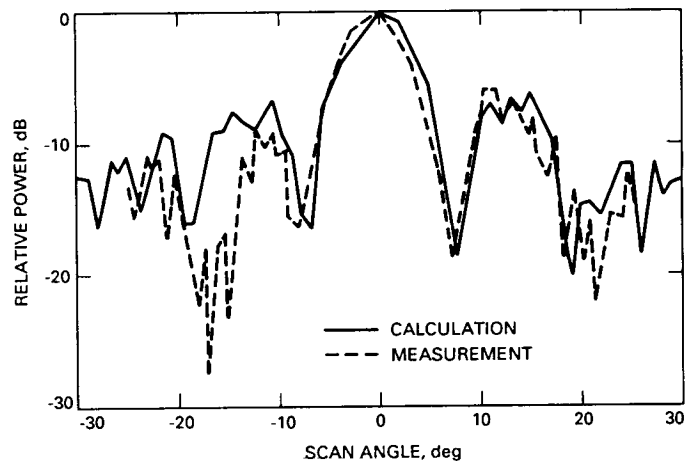
NASA's Search for Extraterrestrial Intelligence (SETI) microwave observing project is preparing to carry out two complementary search strategies: a Sky Survey and a Targeted Search. The NASA Ames Research Center has primary responsibility for developing and carrying out the Targeted Search. Similarly, the Jet Propulsion Laboratory (JPL) has primary responsibility to develop and carry out the Sky Survey. The two approaches are

complementary in that the Targeted Search stresses sensitivity to detect either pulsed or continuous signals, and the Sky Survey gives up sensitivity in order to survey the 99 percent of the sky that is not covered by the Targeted Search.

The objective of the Sky Survey is to search the entire sky over the primary frequency range from 1.0 to 10.0 GHz for evidence of narrowband signals of extraterrestrial origin. Frequency resolutions as narrow as 20 Hz will be used.



**Fig. 11. The spread in measured gain-versus-input power of six power amplifiers.**



**Fig. 12. Measured and calculated beam-scanning patterns for the transmitter array.**

Table 1 is a summary of the objectives of the survey. A significant technical challenge is to design an observational system that will meet maximum sensitivity requirements, use existing radio telescopes, and allow completion of the survey in a "reasonable" time span of 5 to 7 years. The basic strategy for conducting the survey is introduced in [1]. This article describes progress that has been made and gives a high-level description of the hardware and software of a prototype signal processing system that is currently under construction.

## II. Antenna Scan Strategy

Several scan strategies were considered for the Sky Survey and parametric studies were conducted to help in the selection process. These studies clearly showed that the antenna beam must be actively scanned across the sky at a high rate if the survey is to be completed in the allotted time. The NASA SETI plan calls for the Sky Survey to use the 34-m DSN antennas. Because these antennas will not be dedicated to SETI, the scan strategy is quantized so that blocks of time as small as 6 hours can be scheduled and efficiently used.

To accommodate the strategy, the celestial sphere is divided into 900 sky frames that are 60 deg wide by one deg high. The shape and size of the sky frame were selected to maximize the time spent scanning along the width of the sky frame and to minimize the time lost turning around at the edges. At the lower frequencies, where the antenna beamwidths are larger, it will be possible to observe a stack of several sky frames at one time; the resultant observed area of sky might be 10 deg high, for example. Careful scheduling will permit each sky frame to be observed when the width is nearly parallel with the horizon so that the antenna elevation angle will be nearly constant during each scan line. This is done to improve the uniformity of the system sensitivity within the sky frame.

A scan pattern has been selected which can be described as the "racetrack." The scan direction is oriented along a contour of nearly constant elevation and the order of progression of successive scans is chosen to minimize elevation change as time progresses. Thus, the sky frame is observed from top to bottom as it rises in the east; the same sky frame is observed bottom to top as it sets in the west. A schematic diagram of the oval ("racetrack") pattern is shown in Fig. 1.

Details of this scan pattern have been derived so as to utilize the time efficiently, to maintain maximum system sensitivity with uniformity across the sky frame, and to

facilitate the signal detection process. Finally, certain features of this pattern can be used to discriminate against some classes of radio frequency interference (RFI). This concept is explained in the following section.

## III. Receivers and Signal Processing

The SETI Sky Survey requires an end-to-end system that includes large antennas equipped with microwave feeds, polarizers, amplifiers, and receivers that will span the entire frequency band from 1.0 to 10.0 GHz. Preliminary designs for these subsystems are being studied. However, the most challenging components of the system are the spectrum-analyzer and signal-detection subsystems. Much of the challenge of the signal processor involves the impact of RFI, which is expected to be problematic, but which is poorly characterized at this time. The final design of the Sky Survey processor system will benefit if a prototype system is tested in an observational environment that includes RFI. For this reason the SETI Team at JPL decided to build and test a prototype system.

SETI is collaborating with the DSN Advanced Systems Program to design and construct an engineering development model (EDM) of a wideband spectrum analyzer system to serve as a prototype for the SETI Sky Survey processor. A functional block diagram of the EDM system is shown in Fig 2.

The principal features of the EDM system are as follows:

- (1) The Wide Band Spectrum Analyzer (WBSA) is a  $2^{21}$  channel FFT (Fast Fourier Transform) spectrum analyzer which covers an instantaneous bandpass of 40 MHz. The WBSA will operate on the 40-MHz signal at selectable resolutions of 20 Hz, 40 Hz, 80 Hz, 160 Hz, and 320 Hz. Clock-down modes will be available to increase each of these resolutions by factors of 8 or 64. Arbitrary time-domain windowing functions can be implemented in the hardware. Two independent accumulators are available so the system can function in the Dicke-switched mode commonly used for radio astronomy. Two 20-MHz wide channels of opposite polarization can also be implemented.
- (2) The Background Estimation Convolution and Threshold (BECAT) assembly is a hardware module that applies the "along scan" detection algorithm described below. It uses convolution and order statistics to determine the detected power baseline in segments

across the passband and to apply a signal threshold that reduces the data rate by a factor of  $10^3$  or  $10^4$ .

- (3) The EDM software system manager is a network-based C language environment based on the ANSI C standard. It provides observatory control, data acquisition, and analysis and it designed to be portable to any system running Berkeley UNIX.

The WBSA has been fully designed and is presently under construction. It operates at 4.5 GOPS (Giga operations per second) and uses 20-MHz IEEE standard arithmetic units. The system contains 76 Mb of dynamic random access memory (DRAM). It is contained in a single electronics equipment rack and is controlled by a separate Masscomp 5600 computer. Specifications for the BECAT module are approaching completion; design work will begin in the near future.

The EDM system hardware consists of 37 large-size ( $40 \times 40$  cm) circuit boards capable of holding 600 16-pin dual-in-line-packages (DIPs). Twenty boards are unique designs and 17 are copies. To date, 5 boards have been completely designed, fabricated and tested; 8 other boards are currently in process. A completed fixed-point Fast Fourier Transform (FFT) board is shown in Fig. 3.

The EDM software system has been designed and a framework program is currently operating at JPL. The first of three successive software "builds" is scheduled for completion by the end of 1989. The functionality of the system is increased as each "build" is completed.

## IV. Sky Survey Signal Processing

The instantaneous bandpass of the EDM is 40 MHz. Because the system gain profile will not be perfectly flat over the passband, the first step in the background estimation is to normalize the spectral data across the band. This is done by multiplying with an inverse gain matrix that is regularly updated during the calibrations performed as the antenna is turning from one scan to the next.

The resulting normalized spectrum is then divided into subspectra whose bandwidths are optimized for "spectral flatness." Each subspectrum is then examined to determine a pair of order statistics. The  $r_1$ th smallest order statistic is used to determine the mean noise temperature of the subspectrum and the  $r_2$ th smallest is a test of the validity of the assumed probability distribution function. The algorithm leads to an excellent estimate of the noise power in each subspectrum. The estimate is also robust

against strong RFI because it avoids using outliers of the probability distribution function.

For the next stages of the EDM detection process, it is convenient to describe the steps as they are applied to a single frequency bin. These steps are schematically shown in Fig. 4. As the antenna beam scans across the sky at a rapid rate, the WBSA produces a complete  $2^{21}$  point spectrum every 50 msec. For each frequency bin in the spectrum, the WBSA hardware accumulates several spectra and passes the accumulated data on to the BECAT module (Step 2). The number of accumulated spectra and the antenna scan rate are chosen so that five independent accumulations are acquired during the time that the beam has moved one half-power beam width (HPBW) on the sky. In Step 3, the accumulated data samples for each frequency bin are convolved with a 5-point filter that matches the shape of the point-source response function (output power versus time) for the scanning antenna beam. The output is compared to the mean noise power (described above) and signals that exceed a preselected threshold are compared with RFI frequency "maps" resident in the system. Those which do not match frequency patterns of previously detected RFI are stored on disk and passed on to the signal detection software (Step 4).

Next, one considers the inter-scan tests applied in the EDM software. The algorithm is based on the underlying assumption that putative extraterrestrial signals will be originating so far from earth that they will be spatially fixed on the sky; that is, they will match the sidereal motion of the stars. Therefore, a signal with sufficient strength will appear on two or more adjacent scan lines, and the lines will appear at the same place along the scan line because the sky frames are produced after the scan line data are compensated for sidereal motion. This feature makes interstellar signals distinguishable from RFI and random noise, which will generally not move with sidereal rate and direction. A schematic description of this process is shown in Fig. 5.

Another useful discriminator against RFI is the ratio test, which also utilizes the features of adjacent scan lines as well as knowledge of the beam shape in the direction orthogonal to the scan line. For this test we identify a candidate signal, then look for a companion signal on the adjacent scan lines. If the candidate signal is strong, the companion signal(s) should appear with the proper ratio unless the signal is RFI or has disappeared in the time interval between scans. The effectiveness of the ratio test diminishes as candidate signals become weak relative to the system noise.



The BECAT module is designed to provide up to 8192 reports during each spectrum accumulation. This corresponds to a data rate of 800 Kbytes per sec, which can be compared with the 84,000 Kbytes per second data stream from the WBSA accumulator. Each report consists of four numbers, as follows:

- (1) The first frequency of a contiguous run of bins that exceeds threshold
- (2) The number of bins in the contiguous run
- (3) The integrated (convolved) power contained within the run
- (4) The maximum (unconvolved) power in a single bin of the run

The first three quantities characterize the signal that appeared above threshold. The fourth quantity is used to determine if the detection was due to a signal that was present throughout the five accumulations that match the convolutional filter, or if a single accumulation contained a burst with sufficient intensity to exceed threshold even after being diluted by the four other accumulations. Events which fail this test cannot be summarily dismissed as RFI, however, because they could be caused by a pulsed signal. These events will be identified for follow-up observations.

## V. Duration of the Sky Survey

The time required to complete the Sky Survey depends upon several factors that affect the scan rate and the number of scans that must be made. Some of these parameters have already been selected: the survey will span the frequency range from 1 to 10 GHz, and it will be carried out with 34-m parabolic antennas. Preliminary designs of microwave receiving systems indicate that the instantaneous bandwidth for the 6 lowest frequency bands will be 160 MHz, and the bandwidth for the remaining 25 higher frequency bands will be 320 MHz.

The spacing of the adjacent scan tracks in the sky frame affects the spatial uniformity of the survey sensitivity. The nominal spacing is  $1/2$  HPBW. As the HPBW decreases with frequency, the number of scan lines for each sky frame will increase linearly with frequency.

The one remaining free parameter is the angular scan rate of the antenna. The tradeoff is between sensitivity and survey duration; slow scan rates improve the sensitivity and extend the time required to complete the survey. The JPL studies show that a reasonable compromise is to scan at 0.2 deg per sec and to schedule about 30 sec to complete the turnaround between scans. With these rates, one can compute the time required to scan the sky for each of the 31 frequency bands that span the 1-10 GHz range. The result is shown in Fig. 6. Note that the scan time increases almost a factor of 10 from the lowest to highest frequency surveys.

The right-hand ordinate in Fig. 6 displays the cumulative scan time as successively higher frequency surveys are included in the ensemble. For our scan rate of 0.2 deg per sec, the complete survey will require 725 days, or about 2 years of continuous round-the-clock observations. Of course, this means that the real survey will take 3 or 4 times as long when one considers that the antenna may only be scheduled one-third of the time for the survey and that even scheduled time must be allocated for maintenance, weather, and follow-up observations of "interesting" signals. Thus the observing schedule must be adhered to in order to complete the survey in the allotted time. The alternative would likely be to reduce the frequency range of the survey.

## VI. Conclusions

The EDM system is scheduled to undergo validation tests at Goldstone in 1990. These include the following:

- (1) Testing and refinement of the signal detection algorithms
- (2) Testing of alternative scan strategies and observatory control functions
- (3) Testing of RFI rejection techniques

Results from these tests need to be fed back into the design of the full-scale Sky Survey processor, which will be built in the early 1990s.

## Reference

- [1] S. Gulbis and E. T. Olsen, "The Search for Extraterrestrial Intelligence," in *Proceedings of the NRAO Workshop Number 11*, K. I. Kellerman and G. A. Seielstad, eds., pp. 161-162, 1986.

**Table 1. Objectives of the NASA SETI sky survey**

Parameters	Objectives
Spatial coverage	Entire celestial sphere
Frequency range	1.0 to 10.0 GHz inclusive and higher spot bands
Duration	$\approx 6$ years
Frequency resolution	$\approx 30$ Hz
Instantaneous bandpass	$\approx 300$ MHz
Sensitivity	$\leq 10^{-23} \sqrt{f(\text{GHz})} \text{ Wm}^{-2}$
Polarization	Simultaneous dual circular
Signals	Primarily CW with natural radio astronomy fallout

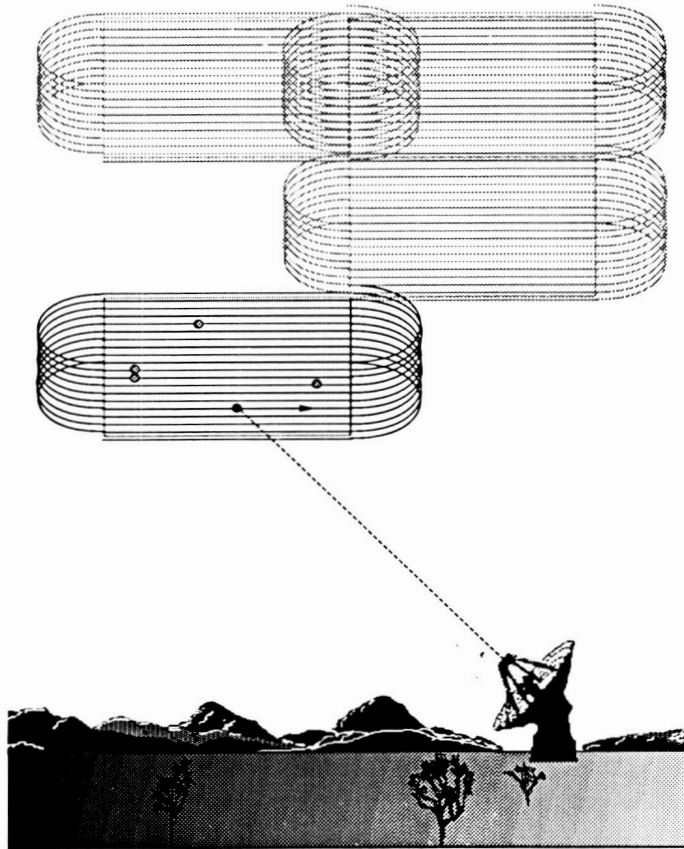


Fig. 1. Sky survey oval scan pattern.

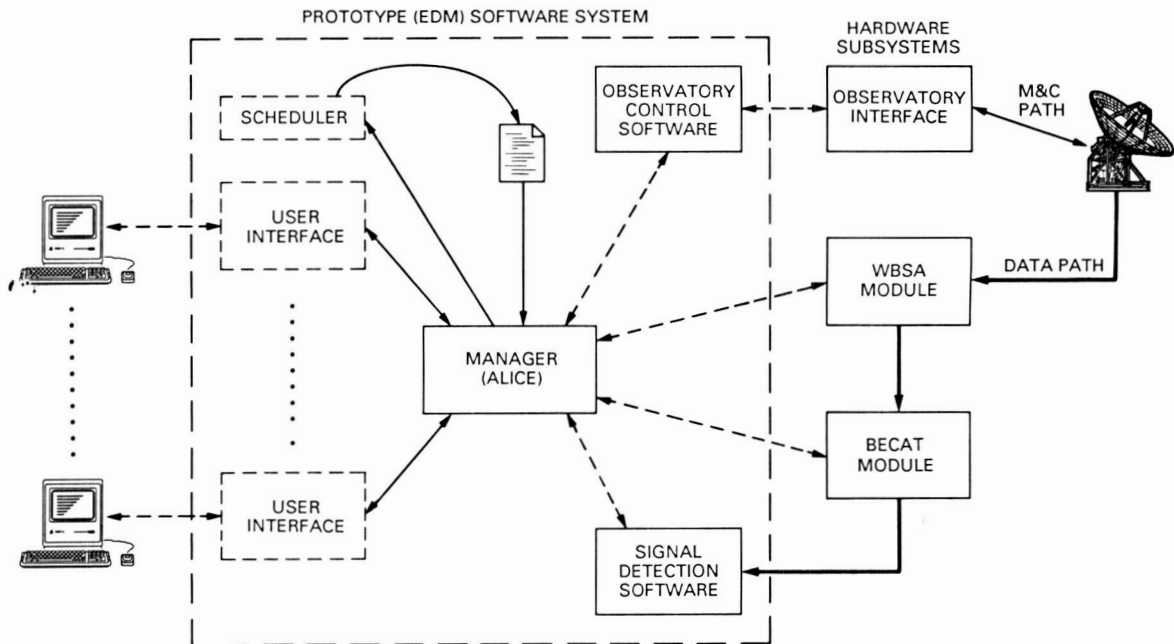
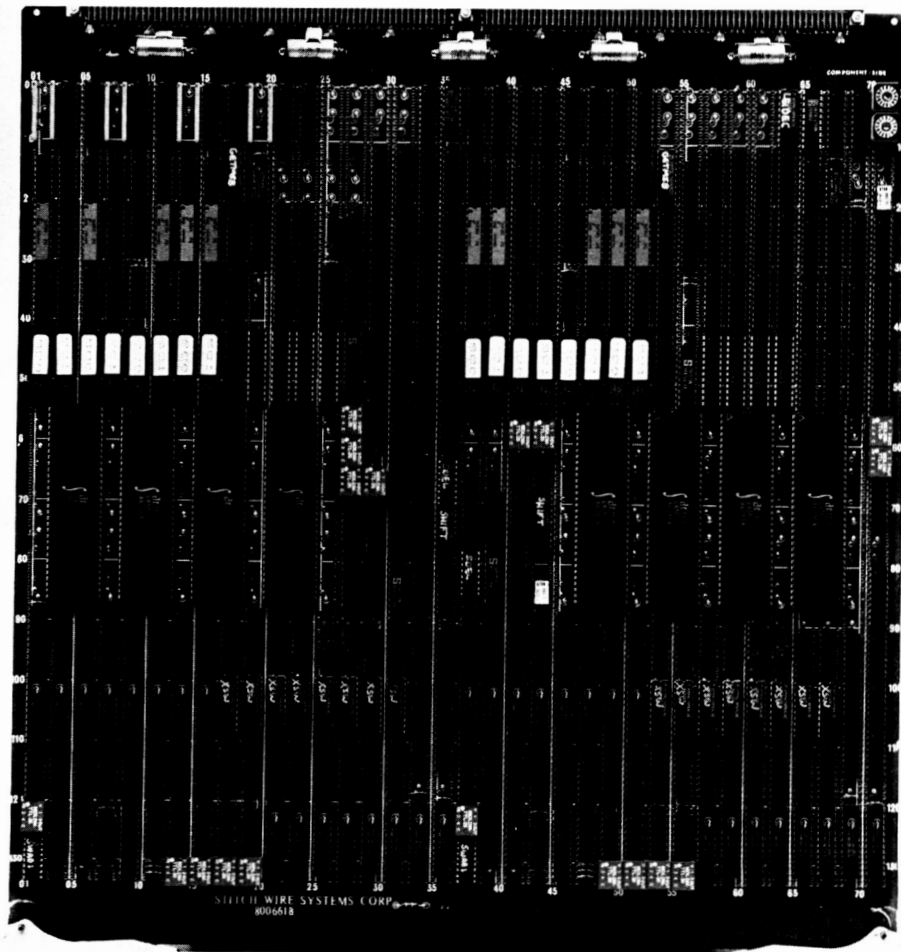
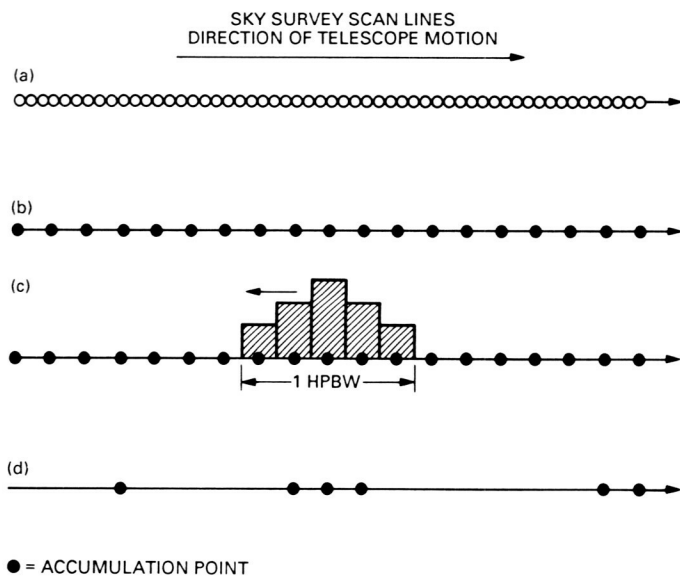


Fig. 2. Prototype system functional diagram.

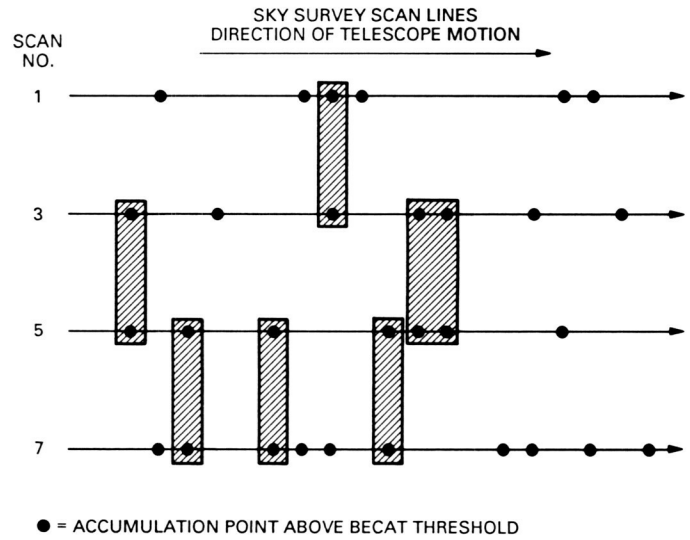


**Fig. 3. Engineering Development Model (EDM) of survey processor:  
a fixed-point FFT board.**

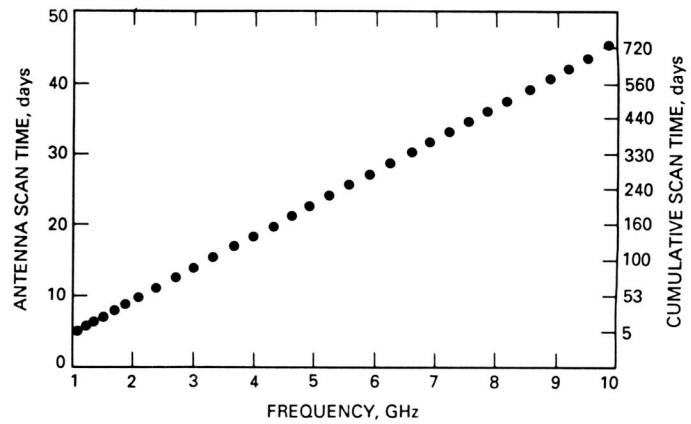
ORIGINAL PAGE  
BLACK AND WHITE PHOTOGRAPH



**Fig. 4. Signal processing in prototype hardware: (a) high-speed sampling; (b) several spectra are co-added (accumulated) by the WBSA; (c) a 5-point convolutional filter is applied by the BECAT; (d) data above BECAT threshold are passed to the signal-detection software.**



**Fig. 5. Signal processing in prototype hardware. The detection software compares candidate points aligned on adjacent spans.**



**Fig. 6. Time required to scan the sky in 31 adjacent frequency bands.**

## Referees

The following people have refereed articles for *The Telecommunications and Data Acquisition Progress Report*. By attesting to the technical and archival value of the articles, they have helped to maintain the excellence of this publication during the past year.

S. Aguirre	P. Estabrook	S. M. Lichten	J. L. Robinett
S. Arnold	N. C. Ham	K. M. Liewer	G. Rodriguez
D. Bagby	R. Hartop	R. T. Logan	B. P. Saldua
R. K. Bartman	H. Hemmati	S. T. Lowe	J. A. Schwartz
D. A. Bathker	R. L. Horttor	L. Maleki	J. Shell
W. Bertiger	L. Howard	C. McLyman	W. L. Sjogren
A. Cha	I.-S. Hsu	A. Mileant	P. Stanton
C.-C. Chen	W. Imbriale	N. Mysoor	R. Stevens
U. Cheng	G. Janik	E. M. Olsen	L. Swanson
K.-M. Cheung	C. E. Johns	T. K. Peng	C. L. Thornton
J. Collier	R. Jurgens	S. Petty	R. N. Treuhft
N. F. deGroot	P. F. Kuhnle	F. Pollara	C. C. Wang
A. C. Densmore	N. R. Kuo	E. C. Posner	M. Wert
G. J. Dick	E. R. Kursinski	M. Rayman	S.-C. Wu
S. Dolinar	R. Levy	R. L. Riggs	