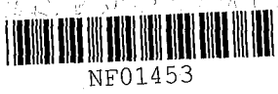
NASA-CR-173263
19840011553

The Telecommunications and Data Acquisition Progress Report 42-76

October-December 1983

E.C. Posner
Editor

February 15, 1984



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The Telecommunications and Data Acquisition Progress Report 42-76

October-December 1983

E.C. Posner
Editor

February 15, 1984

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

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

Preface

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

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

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

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

OSTDS Tasks:

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

OSSA Tasks:

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

Discretionary Funded Tasks

This Page Intentionally Left Blank

Contents

OSTDS TASKS
DSN Advanced Systems
TRACKING AND GROUND-BASED NAVIGATION

Another Look at the Optimum Frequencies for a Water Vapor Radiometer 1
G. M. Resch
NASA Code 310-10-60-55-00

Inversion Algorithms for Water Vapor Radiometers Operating at 20.7 and 31.4 GHz 12
G. M. Resch
NASA Code 310-10-60-55-00

COMMUNICATIONS, SPACECRAFT–GROUND

DSN Microwave Antenna Holography 27
D. J. Rochblatt and B. L. Seidel
NASA Code 310-20-65-54-00

Loop Gain and Circuit Parameters for Residual Carrier Tracking in the Advanced DSN Block V Receiver 43
P. H. Young
NASA Code 310-20-66-18-00

X-Band Uplink Ground Systems Development 53
C. E. Johns
NASA Code 310-20-64-07-00

Doppler System Phase Transfer Functions for a System With an X-Band Uplink and X-Band and S-Band Downlinks 58
M. A. Koerner
NASA Code 310-20-64-07-00

STATION CONTROL AND SYSTEM TECHNOLOGY

Transfer Function Bounds for Partial-Unit-Memory Convolutional Codes Based on Reduced State Diagram 70
P. J. Lee
NASA Code 310-30-71-83-02

Simulations for Full Unit-Memory and Partial Unit-Memory Convolutional Codes With Real-Time Minimal-Byte-Error Probability Decoding Algorithm 77
Q. D. Vo
NASA Code 310-30-71-83-02

Signal-to-Noise Ratio and Combiner Weight Estimation for Symbol Stream Combining 86
Q. D. Vo
NASA Code 310-30-70-84-06

A Systolic VLSI Design of a Pipeline Reed-Solomon Decoder 99
H. M. Shao, T. K. Truong, L. J. Deutsch, J. H. Yuen, and I. S. Reed
NASA Code 310-30-70-84-08

DSN Systems Implementation
CAPABILITIES FOR NEW PROJECTS

ISEE-3 Microwave Filter Requirements 114
J. L. Galvez, H. Marlin, and P. Stanton
NASA Code 314-40-51-60-01

Telemetry Simulation Assembly Implementation in the DSN 120
M. E. Alberda
NASA Code 314-30-53-29-08

Operator Workload Measurement Validation for the Mark IV-A DSCC Monitor and Control Subsystem 132
M. Le May, E. E. Hird, and B. Y. Rodriquez
NASA Code 311-03-53-29-18

First Results of GPS Time Transfer to Australia 141
J. McK. Luck, J. R. Woodger, J. E. Wells, P. N. Churchill, and P. A. Clements
NASA Code 314-30-56-56-19

Implementation of Large Antennas for Deep Space Mission Support 161
R. Stevens
NASA Code 055-40-01-00-96

NETWORK SUSTAINING

Enhanced Radio Frequency Carrier Margin Improvement for an Array of Receiving Systems with Unequal Predetection Signal-to-Noise Ratios 170
M. H. Brockman
NASA Code 311-03-41-82-11

Data and Information Systems
DATA SYSTEM STANDARDS

Optimum Cyclic Redundancy Codes for Noisier Channels 189
P. Merkey and E. C. Posner
NASA Code 314-40-31-30-01

OSSA TASKS
Search for Extraterrestrial Intelligence

SETI Investigations at Jodrell Bank, England: September Through November 1983 196
G. S. Downs and S. Gulkis
NASA Code 199-50-62-09-05

Author Index, 1983 206

Another Look at the Optimum Frequencies for a Water Vapor Radiometer

G. M. Resch
TDA Technology Development Office

A water vapor radiometer is used to estimate the columnar content of atmospheric water vapor or equivalently the line-of-sight path delay due to water vapor. Two measurement channels are used in order to separate the effects of the liquid and vapor phases of water. The efficiency of the path delay or columnar vapor estimate is dependent on the choice of on-line frequency channel. Previous analysis of this problem has suggested frequencies from 20.3 to 21 GHz. The frequency that yields the minimum error in the inversion algorithm is shown here to be both site and season dependent. Hence, the concept of an "optimum" frequency must represent an averaging process over the entire range of meteorological conditions that is expected. For a range of sites and conditions representing a cross section of the continental United States the optimum on-line frequency seems to be 20.6 GHz.

I. Introduction

Interferometic techniques at microwave frequencies are being used in the fields of geodesy and spacecraft navigation (Ref. 1). The time delay imposed by varying amounts of atmospheric water constitutes a potentially limiting error source for these techniques. This error can be reduced by estimating the line-of-sight delay due to water vapor using a device known as a water vapor radiometer (WVR).

Water vapor in the Earth's atmosphere slows the passage of an electromagnetic wave. Equivalently, the electrical path through the medium is longer than the vacuum path (Ref. 2) by the amount ΔL_v, where,

$$\Delta L_v \text{ (cm)} = 1.723 \times 10^{-3} \int \frac{\rho_v}{T} \, ds \qquad (1)$$

where ρ_v is measured in g/m^3, s in meters, and a line of sight through the entire atmosphere is assumed. Water vapor in the Earth's atmosphere emits spectral line radiation centered at a frequency of 22.235 GHz. If the opacity of the atmosphere is low, then to a good approximation the brightness temperature measured on the vapor line is proportional to the columnar vapor content along the line of sight. Hence, if atmospheric opacity or brightness temperature can be measured, Eq. (1) can be estimated. These measurements can be accomplished with passive microwave sensing techniques (Ref. 3) using a device known as a water vapor radiometer (WVR). Typically, the WVR includes a second measurement channel somewhat offset from the water vapor line that is used to eliminate the effects of liquid water in the form of clouds.

The water vapor line is pressure broadened so that the detailed line shape is a function of the vertical distribution of

vapor. Since water vapor is *not* a well-mixed constituent of the atmosphere, it will vary according to site, season, and local meteorological conditions. Both the point value of the vapor density and its distribution vertically and horizontally will vary on a variety of time and spatial scales. The full width of the vapor line at half intensity is approximately 4 GHz whereas the typical observing bandwidth of a WVR is 0.1 - 0.5 GHz. This implies that the accuracy with which Eq. (1) can be estimated will be somewhat dependent upon the exact frequency choice of the WVR. Previous workers have suggested the "optimum" frequency for the WVR is between 20.3 and 21.0 GHz. In this paper the concept of an "optimum" frequency choice is examined along with its site and seasonal variations using a meteorological data base representing a cross section of conditions of the continental United States.

II. The Problem

Section I established the problem as the estimation of the integral quantity,

$$\Delta L_v = 1.723 \times 10^{-3} \int \frac{\rho_v}{T}\, ds$$

The spectrum of atmospheric emission around the 22 GHz water vapor line is shown in Fig. 1 for two different vertical distributions of vapor. The pressure distribution in a standard atmosphere was used with a surface temperature of 30°C and the standard lapse rate. In order to calculate Fig. 1(a), a constant relative humidity between 0 and 1 km is assumed. For Fig. 1(b), a constant relative humidity between 2 to 3.8 km altitude is assumed. The columnar water vapor content was very nearly the same in both cases and equal to 2 g/cm^2. The high altitude vapor clearly shows a sharper line than the low altitude profile although the path delay is the same in both cases. This presents a tradeoff decision in regard to the operating parameters of a WVR. If we wish to maximize the signal from a given amount of vapor, then clearly we should observe at the line center, i.e., f = 22.225 GHz. Equally clear is the fact that we would also be most sensitive to variations in the vapor profile at this frequency. Signal detection is not in general a problem if standard radiometric techniques are used so we will choose to minimize the effects of variations in the vapor profile. Inspection of Fig. 1 would suggest that a single frequency measurement of the brightness temperature somewhere near the half-power point of the line profile would provide the most accurate estimate of the path delay.

Figure 2 shows the brightness temperature of the atmosphere between 10 and 40 GHz plotted for three different cases. Figure 2(a) shows the spectrum of a standard atmosphere without water vapor or liquid. In this case the emission is primarily from the wings of a complex of oxygen emission lines near 60 GHz and is called emission from the "dry" atmosphere. In Fig. 2(b) an exponential distribution of vapor has been added having a columnar content M_v = 2 g/cm^2, and the spectral line centered at 22.2 GHz is clearly evident. Figure 2(c) shows the same atmosphere as in 2(b) but with an additional M_L = 0.1 g/cm of liquid water that is assumed to exist in the form of small droplets.

Figure 2(c) shows that a relatively small amount of liquid water (i.e., such as exists in clouds) can cause a large change in the observed brightness temperature. Crane (Ref. 4) has estimated the effective dielectric constant for liquid water in the form of clouds or light precipitation to be 1.6 so that the excess path delay due to liquid water is approximately

$$\Delta L_L = 1.6\, M_L \qquad (2)$$

where M_L is the total precipitable liquid water along the line of sight in g/cm^2. Since the liquid water from most clouds translates to a precipitable liquid column that is a few tens or hundreds of microns, the path delay due to liquid water is negligible. If there is enough liquid in the atmosphere to contribute to the path delay, then the inversion algorithms probably are not applicable.

Although the liquid water found in clouds does not contribute to the path delay, it does contribute to the brightness temperature – the main observable of the WVR. Droplets of liquid water act as scattering centers removing energy from the line-of-sight path and injecting scattered power from other sources, i.e., the Earth at T = 290 K. If the water droplets are small compared to the wavelength of the radiation and the integrated liquid content is not too large, then the assumptions regarding a non-scattering, low opacity medium will not be violated. This means that the path delay can be estimated if the brightness temperature effects of the liquid and vapor state of water can be separated. A second channel is necessary in the WVR solely to separate the liquid water and vapor effects. The frequency of this second channel should be such as to provide good sensitivity to liquid and maximum contrast to the vapor measurement. Inspection of Fig. 2 suggests that a frequency between 30-32 GHz in the atmospheric "window" would be the appropriate choice for the off-line channel. The frequency of 31.4 was chosen because (a) it is on the wing of the water vapor line so that it is relatively insensitive to vapor but is sensitive to liquid and (b) that frequency band is allocated to space research.

Resch (Ref. 5) has shown that a general formulation of the inversion algorithm for the excess path delay due to water vapor ΔL_v, can be expressed as

$$\Delta L_v = \left[G_1 - \left(\frac{f_1}{f_2}\right)^2 G_2 \right]^{-1}$$

$$\times \left[\tau_1 - \left(\frac{f_1}{f_2}\right)^2 \tau_2 - (1-\beta)\,\tau_{d1} \right] \qquad (3)$$

where τ_i is the observed opacity at frequency f_i, and τ_d is the opacity due to the "dry" atmosphere, i.e., primarily oxygen, and β is the ratio of τ_d at the two observing frequencies. The observable quantity with a WVR is the brightness temperature T_{bi}, which is transformed to an opacity using the relation

$$\tau_i = -\log_e \left(\frac{T_{mi} - T_{bi}}{T_{mi} - T_c} \right) \qquad (4)$$

where T_{mi} is the mean radiating temperature of the atmosphere, and T_c is the cosmic blackbody background (2.9 K). The quantity G_i is termed the single-frequency weighting function and is defined as

$$G(f, s) = (5.803 \times 10^2) \left(\frac{T\alpha_v}{\rho_v} \right) \qquad (5)$$

where T is the physical temperature (Kelvin), ρ_v is the vapor density (g/m^2), and α_v is the vapor absorption coefficient (neper/m). In the inversion algorithm we have the quantity W defined as,

$$W = \left[G_2 - \left(\frac{f_2}{f_1}\right)^2 G_1 \right] \qquad (6)$$

and called the dual frequency weighting function, or simply the weighting function.

Previous workers have attacked the problem of optimum frequency choice by examining the behavior in either G_i or W and picking the frequency of minimum variation. Menius et al. (Ref. 6) considered the single frequency weighting function in a standard atmosphere with an assumed vapor distribution that represented "average" conditions and concluded that the operating frequency should be either 20.8 or 23.7 GHz. Westwater (Ref. 7) also examined the single frequency weighting function but used two radiosonde launches chosen at random from Miami, Florida. He concluded that G_i is relatively flat for the entire frequency interval 20.1 to 21.0 GHz and that the slope closest to zero lies between 20.6 and 20.8 GHz. Gaut (Ref. 8) in his Ph.D. thesis did not directly address the optimum frequency choice but did give an excellent discussion of the problem using data representative of yearly variations in the New England area and pointed out that with three frequencies one can have a composite weighting function that is constant to a high degree of approximation. Webster (Ref. 9) employed a simulation analysis to determine the optimum frequencies of a two channel WVR using model atmospheres to solve the equation of radiative transfer. He argued that regardless of the off-line frequency, the on-line measurement should be about 1 GHz from the line center, i.e., 21 or 23 GHz. Wu (Ref. 10) systematically searched for minimum variation in W for one radiosonde launch from Pt. Mugu, California, and concluded that the frequency pair 20.3 and 31.4 GHz was optimum. He examined data from two additional launches, representing dry and wet conditions and argued that his conclusion was consistent with these data.

One might wonder whether these different conclusions are due to different analysis, different input data, or both. The next section will show that "flattest" values for both G_i and W are site and season dependent. Hence the "optimum" frequency choice for the on-line measurement channel must be made after consideration of the entire range of meteorological conditions under which our instrumentation is expected to perform.

III. Analysis of the Problem

The quantity of fundamental importance in our calculations is the total absorption coefficient evaluated at some arbitrary point s along a given line of sight through the atmosphere. The analytical formulation for the absorption coefficient used in the following discussion is taken from Ref. 11.

In order to evaluate the integral quantities we must have the distribution functions $T = T(s)$, $P = P(s)$, $\rho_v = \rho_v(s)$, etc. Depending upon immediate objectives and/or convenience, these distributions can be generated in either of two ways. The first method is to use the "standard atmosphere" (Ref. 12) to which we add an assumed exponential distribution of water vapor specified by a surface density ρ_o, and scale height h_o, of the form

$$\rho_v = \rho_o \exp\left(-\frac{h}{h_o}\right) \qquad (7)$$

so that the total precipitable vapor content is simply

$$M_v = \rho_o h_o \qquad (8)$$

The second method is to use radiosonde data from several U.S. sites to generate a table of T, P, and ρ_v sampled at standard levels.

Using the standard atmosphere we can investigate how various quantities vary under nominal conditions. Using the radiosonde data we can investigate how these quantities vary in the real atmosphere in regard to site and season. The data base of radiosonde data is taken from five sites in the United States during the year 1976. The sites – Portland, Maine; Pittsburgh, Pennsylvania; El Paso, Texas; San Diego, California; and Oakland, California – were chosen to represent a cross section of meteorological conditions. Each radiosonde launch provides an approximately vertical profile of pressure, temperature, and relative humidity that is used to calculate the weighting functions. One launch out of eight was selected from each site so as to obtain equal amounts of data from both 0 and 12 hours Universal Time and to cover seasonal trends in the data.

Figure 3 shows the variation of the single-frequency weighting function with altitude for several frequencies in the vicinity of the water-vapor line. It is obvious that $G(f = 31.4\ \text{GHz})$ is not constant, nor for that matter is G_i for any other frequency. This observation raises several questions in light of our previous assumptions. First, how badly have we violated our assumption regarding the consistency of G_i? Second, if G_i is not a constant with regard to s, how do we pick the "best" frequency to measure the on-line brightness temperature? Third, since we know that G_i is not really a constant, what value shall we take to represent W in Eq. (3); the surface value, the value at 1 km, an average value?

Consideration of the behavior of the single-frequency weighting function provides insight in regard to systematic error sources. For a single channel WVR we would choose an operating frequency on the vapor line that yielded the "flattest" G_i for all of the sites that we intend to visit. For a two channel WVR the situation is somewhat more complicated in that we desire the best inversion accuracy; hence, we must minimize the variance of W, not G_i. If we assume that the off-line measurement is made at a frequency of 31.4 GHz, then the "best" frequency for the vapor channel is found by computing $\Delta W/\overline{W}$ versus frequency for radiosonde data and noting the frequency at which this quantity is a minimum. Figure 4 shows $\Delta W/\overline{W}$ versus frequency, calculated for a standard atmosphere containing an exponential distribution of vapor with a scale height of 2.2 km, where $\overline{W}$ is the average value of W and ΔW is the rms value. The minimum value of $\Delta W/\overline{W}$ will depend on the altitude at which we truncate the exponential distribution of the water vapor. The three curves in Fig. 4 illustrate what happens to $\Delta W/\overline{W}$ as the vapor distribution is tuncated at altitudes of 11 km, 6 km, and 4 km. Note that the frequency of the minimum variation point is a function of the cutoff altitude. Since the vapor distribution in reality tends to be both site and season dependent, it suggests that the concept of an "optimum" frequency for the vapor channel is likely to be site and season dependent. This is probably the reason why previous workers have suggested slightly different values for the optimum frequencies of a WVR. Also note in Fig. 4 that the percentage deviations represented by $\Delta W/\overline{W}$ are not very large – even at the 22.235 GHz frequency. Hence, the assumption regarding the constancy of G_i is reasonably good.

Radiosonde data shows that the altitude distribution of water vapor changes from day to day. Figure 5(a) shows f_m equals the frequency of minimum $\Delta W/\overline{W}$ vs. date for the Portland, Maine, radiosonde data. Figure 5(b) shows the same quantity calculated from the El Paso, Texas, radiosonde data. For these figures, W is calculated only at those points where the radiosonde indicates the presence of water vapor. The figures illustrate that the optimum frequency of the WVR is indeed site and season dependent. In our geodetic support applications the WVRs are expected to operate at a wide variety of sites throughout the year. The best frequency choice can only be optimum in some average sense. Figures 5(c) through 5(h) show the distribution function for f_m calculated from all five radiosonde launch sites and suggest that the best choice for the vapor frequency under continental U.S. conditions should be 20.6 GHz.

For historical reasons (Ref. 13), the operating frequencies of the WVRs recently constructed at JPL in order to support VLBI experiments are 20.7 and 31.4 GHz. This means that the error in the path delay estimates from these instruments may depend on the vapor profile a bit more strongly than if we had chosen the lower operating frequency. Figure 6(a) shows the distribution of $\overline{W}$ calculated from radiosonde data for Portland, and Fig. 6(b) shows the histogram for all sites. The time plots of this data for all sites show a clear seasonal dependence that raises the possibility that the variations in W can be reduced by modeling. Inspection of Eqs. (3) and (5) suggests that it might be possible to empirically parameterize W in the form $W = W(P, T, \rho_v)$, i.e., as a function of surface parameters, so as to further reduce these variations.

IV. Summary

The formulation of an inversion algorithm for use with a water vapor radiometer (Eq. 3) contains a term of the form,

$$W = \left[G_1 - \left(\frac{f_1}{f_2} \right)^2 G_2 \right]$$

that we call the dual frequency weighting function. The terms G_1 and G_2 are the single frequency weighting functions associated with two measurement channels of the WVR and were

assumed to be constant in the derivation of the algorithm. We have seen that quantities are not truly constant in the real atmosphere and hence introduce "noise" in the inversion process. If one wishes to minimize the level of this noise and (1) accepts the fact that the off-line channel is f_2 = 31.4 GHz, and (2) uses only the constancy of W as a figure of merit, then it is shown the "optimum" on-line frequency is 20.6 GHz for conditions representing a cross section of the continental United States. The analysis shows that the frequency choice is site and season dependent. If one desired the best possible accuracy, one might consider tailoring the on-line frequency for local meteorological conditions.

While this work is suggestive, it is by no means complete. A rigorous analysis of the optimum frequency problem would solve for both operating frequencies, f_1 and f_2, of the WVR. It would be done by including (1) a noise model of instrumental performance, (2) the finite bandwidth of each measurement channel, and (3) a realistic weighting of the liquid water retrieval. The last item is perhaps the most difficult with which to deal.

It is clear that the off-line measurement is required solely to deal with the possible presence of liquid water in the form of clouds. In clear sky conditions the off-line channel simply adds noise to the inversion. We have assumed that the spectrum of liquid water varies as frequency squared, a good assumption if the drop size is small. As the drop size grows to be an appreciable fraction of a wavelength, this assumption is violated as the frequency behavior is much more complex. The violation of this assumption is felt most strongly at the highest frequency measurement channel. The analysis must include a realistic estimate of the drop size distribution and occurrence statistics.

References

1. Resch, G. M. (1980), "Water Vapor – The Wet Blanket of Microwave Interferometry" in *Atmospheric Water Vapor*, Deepak, Wilkerson, and Ruhnke, eds., pp. 265-282, Academic Press, New York.

2. Waters, J. W. (1967), "Absorption Effects on Radiowave Phase and the Correction of Vapor-Caused Phase Fluctuations by Radiometric Measurements of Water Vapor Emission," VLA Scientific Memo #8, National Radio Astronomy Observatory.

3. Hogg, D. C., F. O. Guiraud, J. B. Snider, M. T. Decker, and E. R. Westwater (1983), "Microwave Radiometry for Measurement of Water Vapor," in *Reviews of Infrared and Millimeter Waves*, Vol. 1, K. J. Button, ed., Plenum Publ. Corp.

4. Crane, R. K. (1966), "Microwave Scattering Parameters for New England Rain," *Tech. Rept. #426*, Lincoln Lab. (MIT), Cambridge, MA.

5. Resch, G. M. (1984), "Inversion Algorithms for Water Vapor Radiometers Operating at 20.7 and 31.4 GHz." *TDA Prog. Rept. 42-76*, Jet Propulsion Laboratory, Pasadena, CA.

6. Menius, A. C., C. F. Martin, W. M. Layson, and R. S. Flagg (1964), "Tropospheric Refraction Corrections Using a Microwave Radiometer," Pan American Airways Tech. Staff Memo #19, ETV-TM-64-12.

7. Westwater, E. R. (1967), "An Analysis of the Correction of Range Errors due to Atmospheric Refraction by Microwave Radiometric Techniques," *ESSA Tech. Rept. IER 30-ITSA30*, available from National Technical Info. Service, Springfield, VA.

8. Gaut, N. E. (1968), "Studies of atmospheric water vapor by means of passive microwave techniques," *Tech. Rept. #467*, M.I.T., Research Lab. of Electronics, Cambridge, MA.

9. Webster, Jr., W. J. (1975), *On the Determination of Atmospheric Path Delay by Passive Microwave Radiometer*, Doc. #X-922-75-107, Goddard Space Flight Center, Greenbelt, MD.

10. Wu, S. C. (1979), "Optimum Frequencies of a Passive Microwave Radiometer for Tropospheric Path-length Correction," *IEEE Trans. Ant. & Prop.*, AP-27, pp. 233-239.

11. Waters, J. W. (1976), "Absorption and Emission by Atmospheric Gases" in Methods of Experimental Physics, M. L. Meeks, ed., *Astrophys.*, Part B, V12, pp. 142-175.

12. List, R. K. (1958), ed., *Smithsonian Meteorological Tables*, 6th ed. rev., The Smithsonian Institution, Washington, D.C.

13. Resch, G. M., M. Chavez, and N. I. Yamane (1982), "Description and Overview of an Instrument Designed to Measure Line-of-Sight Delay Due to Water Vapor." *TDA Prog. Rept. 42-72*, Jet Propulsion Laboratory, Pasadena, CA, pp. 1-19.

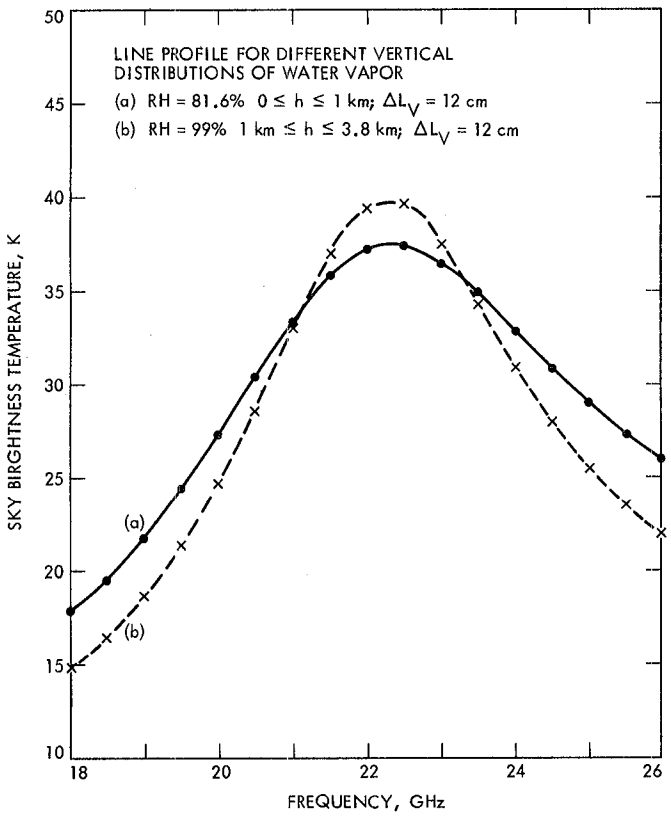


Fig. 1. Line profile of atmospheric water vapor for two different vertical distributions in a standard atmosphere: (a) *RH* = 81.6% for 0 ≤ *H* ≤ 1000 m; and (b) *RH* = 99% for 1000 ≤ *H* ≤ 3800 m. In both cases ΔL_v = 12 cm

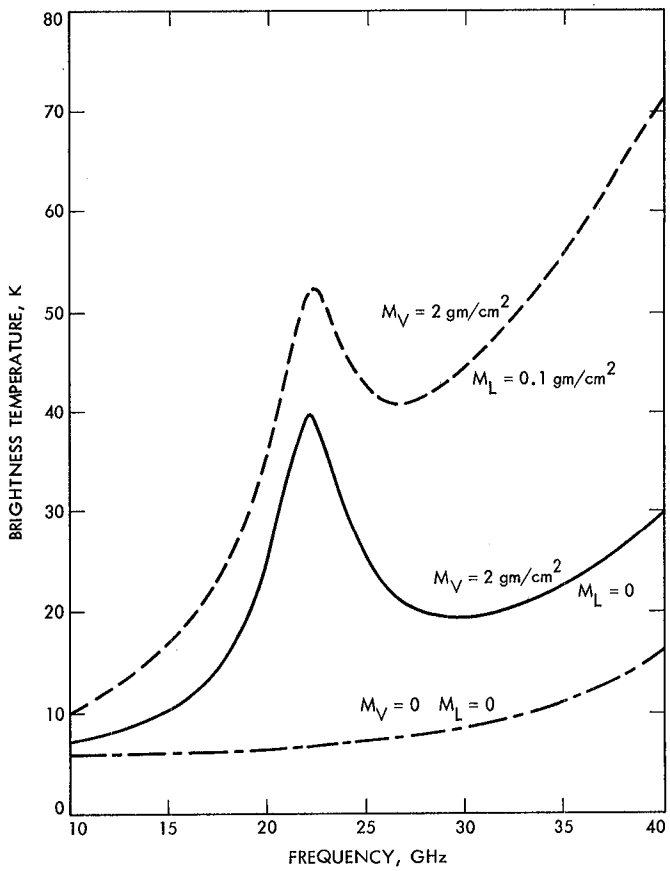


Fig. 2. Brightness temperature of the atmosphere for three cases: (a) oxygen only, no vapor and no liquid; (b) oxygen plus 2 g/cm² of precipitable vapor; (c) oxygen plus vapor plus 0.1 g/cm² of liquid (as in a cloud)

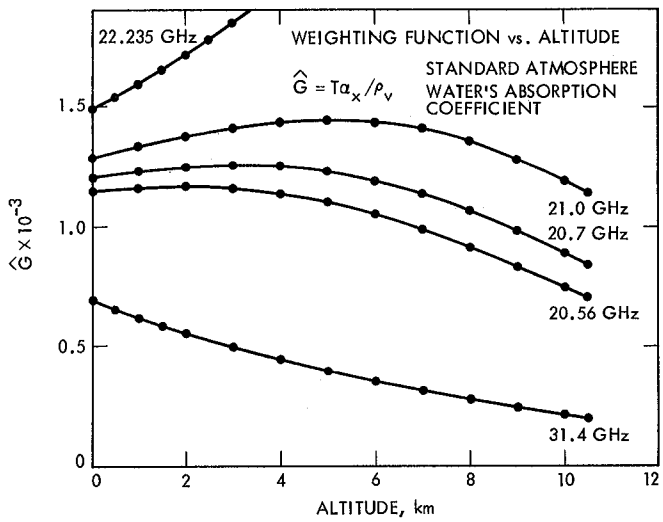


Fig. 3. Variation of the single frequency vapor weighting function versus altitude for several frequencies on or near the water vapor emission line

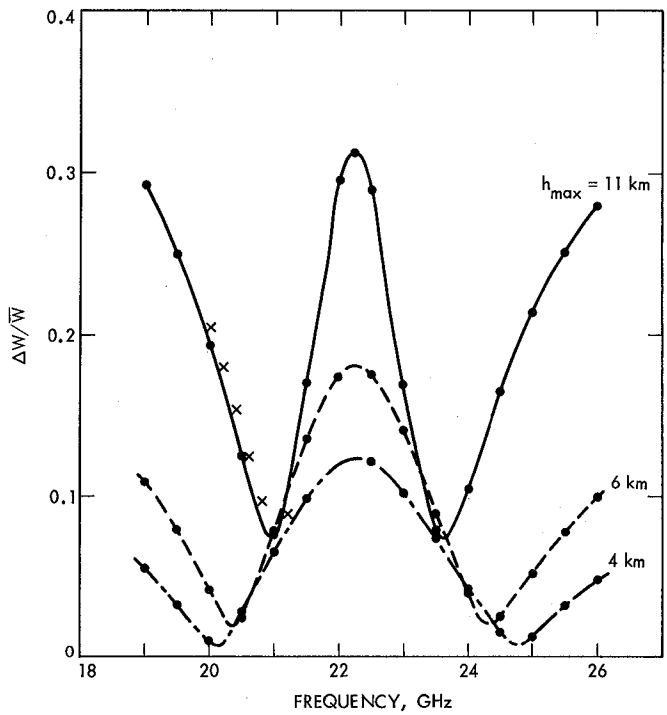


Fig. 4. Variation of the vapor weighting function in a standard atmosphere as a function of frequency for various cutoff altitudes. The off-line frequency is 31.4 GHz.

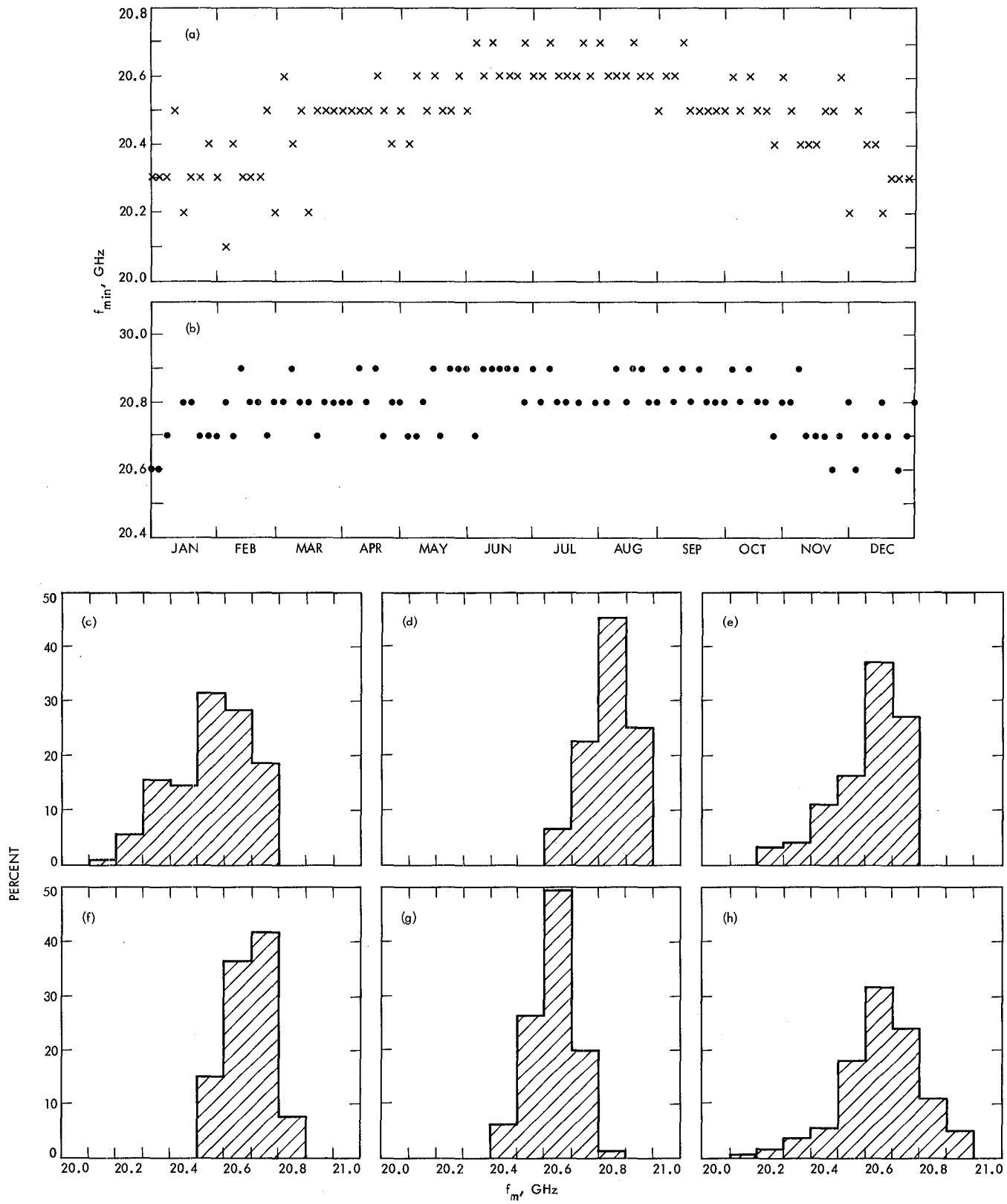


Fig. 5. f_m = the frequency (GHz) of minimum variation in the weighting function (i.e., minimum $\Delta W/\bar{W}$) versus time for (a) Portland, Maine; (b) El Paso. Also shown are the histograms for (c) Portland; (d) El Paso; (e) Pittsburg; (f) San Diego; (g) Oakland; and (h) all sites.

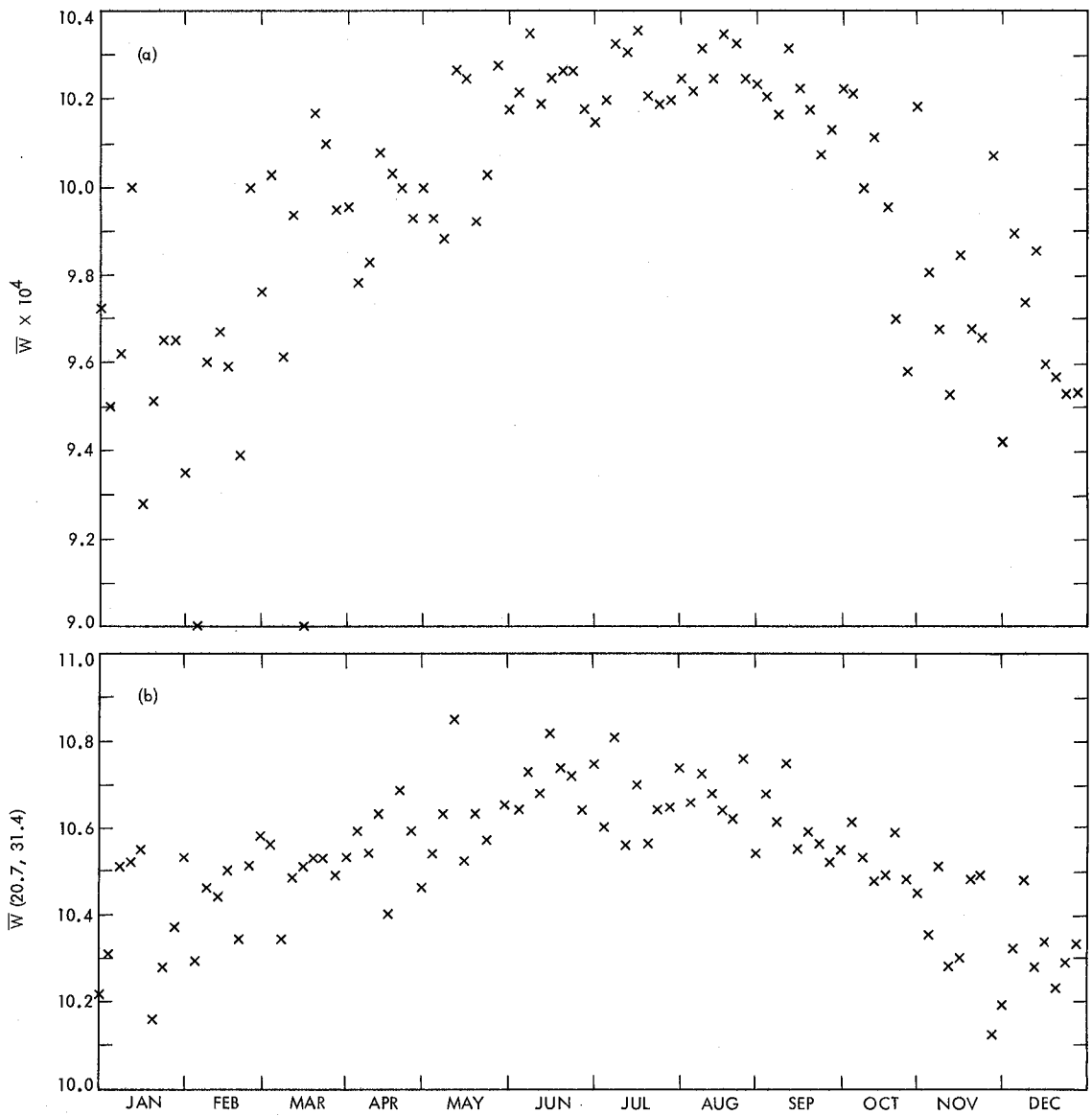


Fig. 6. W = the average value of the weighting function at 20.7 and 31.4 GHz, the frequencies of the JPL water vapor radiometers versus time for (a) Portland; (b) El Paso; and (c) histogram of *W* from all sites

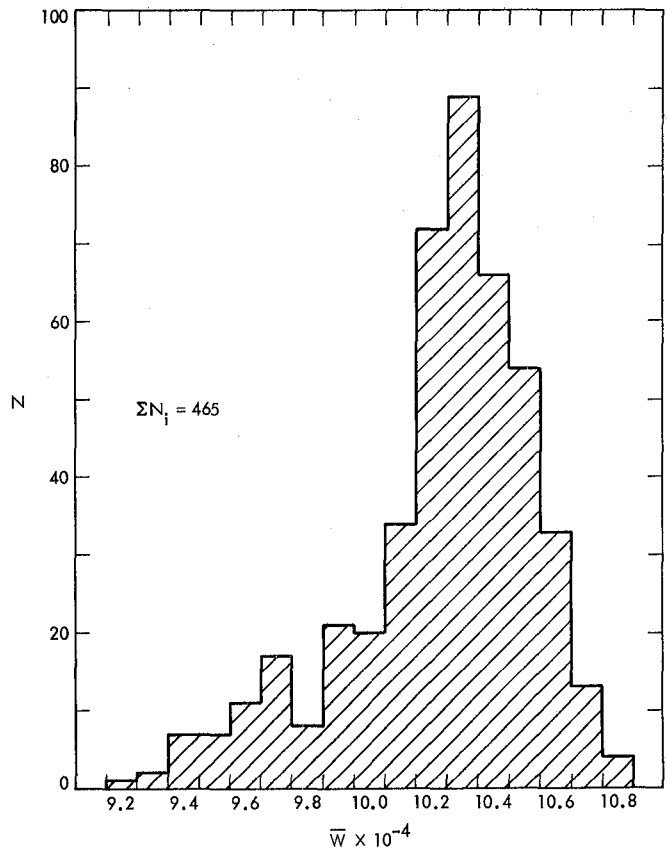

100
80
60
40
20
0
N
$\Sigma N_i = 465$
9.2
9.4
9.6
9.8
10.0
10.2
10.4
10.6
10.8
$\overline{W} \times 10^{-4}$

Fig. 6 (contd)

Inversion Algorithms for Water Vapor Radiometers Operating at 20.7 and 31.4 GHz

G. M. Resch
TDA Technology Development

Eight water vapor radiometers (WVRs) have been constructed as research and development tools to support the Advanced Systems Program in the Deep Space Network and the Crustal Dynamics Project. These instruments are intended to operate at the stations of the Deep Space Network (DSN), various radio observatories, and mobile facilities that participate in very long baseline interferometric (VLBI) experiments. It is expected that the WVRs will operate in a wide range of meteorological conditions. Several algorithms are discussed that can be used to estimate the line-of-sight path delay due to water vapor and columnar liquid water from the observed microwave brightness temperatures provided by the WVRs. In particular, systematic effects due to site and seasonal variations are examined. The accuracy of the estimation as indicated by a simulation calculation is approximately 0.3 cm for a noiseless WVR in clear and moderately cloudy weather. With a realistic noise model of WVR behavior, the inversion accuracy is approximately 0.6 cm.

I. Introduction

The applications that concern the DSN are contained in the general areas of radio geodesy and spacecraft navigation (Ref. 1). The experimental techniques utilize microwave signals from extraterrestrial radio sources to measure a differential – time of arrival, doppler, or range. Propagation effects imposed by the Earth's atmosphere are treated as unknown time delays or an increase in range and must be calibrated. In this section the water vapor problem is reviewed, and it is noted that the excess path length can be expressed as an integral of the vapor density divided by the temperature integrated along the line of sight. In Section II, the equation of radiative transfer is solved, and it is shown that, given certain assumptions, the brightness temperature or opacity of the atmosphere can be used to estimate the excess path delay due to water vapor. In Section III, a cross section of meteorological data is used to solve for the constants in several formulations of the inversion algorithm, and site and seasonal variations are discussed.

The primary atmospheric propagation effect that concerns the geodesist or navigator is refraction. The apparent or electrical path length L_e, along some atmospheric path L, is defined as

$$L_e = \int_L n(s)\, ds \tag{1}$$

where $n(s)$ is the refractive index at the position s.

As a matter of convenience we will work with the excess path length $\Delta L = L_e - L$ or

$$\Delta L = \int_L (n - 1)\, ds \tag{2}$$

Since the refractive index of the atmosphere departs from unity by only a few parts per ten thousand (or less), it is customary to use the refractivity N, defined as

$$N = (n - 1)\, 10^6 \tag{3}$$

so that the refractivity of a unit volume is characterized by this number of N units, typically on the order of 320 at sea level. Bean and Dutton (Ref. 2) discuss several formulations of N as a function of atmospheric parameters. Used here is the form given by Smith and Weintraub (Ref. 3),

$$N = 77.6 \left(\frac{P}{T}\right) + 3.73 \times 10^5 \left(\frac{e}{T^2}\right) \tag{4}$$

where P = total pressure (mb), T = temperature (K), and e = partial pressure of water vapor (mb). This expression is considered accurate to 0.5% for frequencies less than 30 GHz in normal ranges of temperature, pressure, and relative humidity. Note that the total refractivity can be written as the sum of a "dry" term $N_d = 77.6\ (P/T)$ and a "wet" term $N_v = 3.73 \times 10^5\ (e/T^2)$ that yields both a dry, ΔL_d, and wet, ΔL_v, path delay correction:

$$\Delta L_d = 77.6 \times 10^{-6} \int \frac{P}{T}\, ds \tag{5}$$

$$\Delta L_v = 3.73 \times 10^{-1} \int \frac{e}{T^2}\, ds \tag{6}$$

The concern here is with the wet term, and elementary definitions (e.g., see Ref. 4) can be used to express the vapor partial pressure as the vapor density ρ_v to get the expression

$$\Delta L_v \text{ (cm)} = 1.723 \times 10^{-3} \int_L \frac{\rho_v}{T}\, ds \tag{7}$$

where ρ_v is measured in units g/m^3, s is in meters, and a line of sight through the entire atmosphere is assumed. Since water vapor is *not* a well mixed constituent of the atmosphere, ΔL_v will vary according to site, season, and local meteorological conditions. Both the point value of the vapor density and its distribution vertically and horizontally will vary on a variety of time and spatial scales. Zenith values of ΔL_v vary from 3 to 20 cm and scale approximately as the cosecant of the elevation angle for other line-of-sight paths. In applications, observations are not usually made at the zenith but over a wide range of elevation angles for which El = 30° might be an average value. The path delay might therefore vary between 3 and 40 cm as extremes. Clearly, if accuracy goals are on the order of 3 m, water vapor effects can be ignored. If the accuracy goal is 30 cm, somehow the water vapor effect must be estimated but the estimate of ΔL_v does not have to be very good, e.g., if Eq. (7) could be estimated with an accuracy of 50%, the system accuracy requirement might well be satisfied. In order to make this estimate a nominal model of the vapor distribution $\rho_v(s)$ might be used and vertical and horizontal fluctuations might simply be ignored. If the goal is a 3 cm system accuracy, then the temporal and spatial variation of ρ_v along the line of sight is no longer ignorable. The integral in Eq. (7) must be estimated while observing with the geodetic system. Accurate estimates of the line-of-sight path delay must be made from direct measurements or by use of the techniques of remote sensing. In this paper, a microwave technique of passive remote sensing that utilizes a water vapor radiometer (WVR) is described.

When a radio wave from an extraterrestrial source impinges upon the Earth's atmosphere at other than normal incidence, Snell's law tells us that the direction of propagation will be changed. At the Earth's surface, $n \sim 1.0003$ and approaches unity with increasing height so that in general the ray path has a curvature that is concave downward. To an observer on the Earth's surface, an extraterrestrial radio source will appear at a slightly higher elevation than he would calculate from an ephemeris. The difference between the observed elevation of the source and the true elevation is called the angular refraction and represents a rough measure of the curvature in the ray path. At high elevations the curvature is negligible and the path L in Eq. (1) is equivalent to the geometric line of sight to the source. However, at low elevations the path curvature becomes appreciable; Eq. (1) must be expressed in spherical coordinates and is solved using ray tracing methods (e.g., see Ref. 2) which require full specification of the total refractivity in both vertical and horizontal planes. Intuitively, we can see that at low elevations the actual propagation path becomes very dependent upon gradients in the vertical refractivity profile. Most of the refraction occurs in the lowest 10 km of the atmosphere so that a ray entering at say 5° elevation will travel a path whose projected length on the Earth's surface is about 115 km. Even if we knew the distribution of refractivity above the observing site, it is highly unlikely that this distribution would apply along the entire projected path. Depending on our system accuracy requirements, there will be some elevation angle below which these effects defy our current calibration ability. Our current

approach is to avoid these problems by restricting observations to elevation angles greater than 15°.

II. Formulation of the Algorithm

The previous section established the problem as the estimation of the integral quantity

$$\Delta L_v = 1.723 \times 10^{-3} \int \frac{\rho_v}{T}\, ds$$

A suggestion of how this quantity might be estimated can be gotten from consideration of Fig. 1. In this figure is plotted the calculated brightness temperature of a model atmosphere in the frequency range 10 to 40 GHz for three cases. The first case is for a standard atmosphere containing no water, i.e., $M_v = M_L = 0$ where M_v and M_L are the precipitable vapor and liquid, respectively, in units of g/cm^2. The precipitable vapor or liquid is defined as that mass of vapor or liquid that would be precipitated from a column extending through the entire atmosphere with a cross section of 1 cm^2. The second case shows the spectrum of an atmosphere containing $M_v = 2$ g/cm^2 distributed exponentially with a scale height of 2 km. The third case shows the same atmosphere with an additional liquid $M_L = 0.1$ g/cm^2 that is assumed to exist in small droplets. The "bump" in the curves for cases (2) and (3) is due to emission from the water vapor molecule, and it is apparent that the brightness temperatures is a strong function of the amount of water vapor. Thus, a measurement of the brightness temperature is effectively an estimate of the integrated vapor density which constitutes a major portion of the integral in Eq. (7). The problem then is to find the explicit form of the relationship between brightness temperature and vapor path delay and to subtract out the effect of liquid water. In order to do this, the techniques of passive remote sensing will be used.

We will consider the emission and absorption properties of the atmosphere in terms of a gaseous medium in local thermodynamic equilibrium. For a nonscattering, nonrefractive medium, the equation of radiative transfer given by Chandrasekhar (ref. 5) can be transformed using the Rayleigh-Jeans approximation of Planck's law of radiation to the form,

$$T_b(s) = T_b(0) \exp\left[-\tau(s, o)\right] + \int T(s)\, \alpha(f, s) \exp\left[-\tau(s, s')\right]\, ds \qquad (8)$$

which is shown schematically in Fig. 2. Radiation at a frequency f, of apparent blackbody temperature $T_b(s)$ is detected at position s from a medium that both emits and absorbs. Radiation $T_b(0)$ is incident on the medium at $s = 0$ and is attenuated by the factor $\exp\left[-\tau(s, o)\right]$. A unit volume of the medium at a physical temperature $T(s)$, characterized by an absorption coefficient $\alpha(f, s)$ will emit radiation at frequency f, which is attenuated by the factor $\exp\left[-\tau(s, s')\right]$. The optical thickness τ (or opacity) is defined as

$$\tau(s, s') = \int_s^{s'} \alpha(f, s'')\, ds'' \qquad (9)$$

By the mean value theorem of integral calculus the integral in Eq. (8) can be written

$$\int T \alpha \exp(-\tau)\, ds = T_M \int \alpha \exp(-\tau)\, ds \qquad (10)$$

where T_M is termed the mean radiating temperature of the atmosphere. Using this definition the solution to Eq. (8) is

$$T_b = T_c \exp(-\tau) + T_M \left[1 - \exp(-\tau)\right] \qquad (11)$$

For low values of the opacity ($\tau \ll 1$) $T_b \sim T_c$, i.e., the medium is transparent and we simply "see" the incident radiation which in this case is T_c, the cosmic blackbody background at an apparent temperature of 2.9 K. For large values of the opacity ($\tau \gg 1$) $T_b \sim T_M$, the medium is opaque and we "see" the gas radiating at its effective temperature. It is convenient to solve Eq. (11) for the opacity in the form

$$\tau = -\log_e\left(\frac{T_M - T_b}{T_M - T_c}\right) \qquad (12)$$

We can express the total atmospheric absorption in Eq. (9) as the sum of its three primary contributors, a water vapor term, α_v, a liquid term, α_L, and a "dry" term, α_d, that describes the background radiation primarily from the wings of a series of oxygen resonance lines near 60 GHz:

$$\tau = \int (\alpha_v + \alpha_L + \alpha_d)\, ds = \tau_v + \tau_L + \tau_d \qquad (13)$$

Explicit formulations of the absorption coefficients can be found in the literature. This study uses Waters' form of the water vapor absorption coefficient (Ref. 6), Snider and Westwater's form for oxygen (Ref. 7), and Staelin's form for liquid water (Ref. 8). Since the water vapor absorption is a

linear function of the vapor density we can write the vapor opacity term as

$$\tau_v = \int \left(\frac{\rho_v}{T}\right)\left(\frac{\alpha_v T}{\rho_v}\right) ds \qquad (14)$$

Thus, the opacity due to water vapor can be expressed in a form that contains the functional form (i.e., Eq. [7]) of the path delay integral.

To express the vapor opacity as a linear function of the path delay, $\tau_v = G(\Delta L_v)$ where G should be constant with respect to s but may be a function of frequency f, implies

$$G(f, s) = (5.803 \times 10^2)\left(\frac{\alpha_v T}{\rho_v}\right) \qquad (15)$$

Let us suppose that we measure the brightness temperature at two frequencies f_1 and f_2 and transform the observables T_{b1}, T_{b2} using Eq. (12) to estimate the total opacity. At each frequency Eq. (12) can be written

$$\tau_1 = G_1(\Delta L_v) + \tau_{L1} + \tau_{d1} \qquad (16)$$

$$\tau_2 = G_2(\Delta L_v) + \tau_{L2} + \tau_{d2} \qquad (17)$$

Staelin (Ref. 8) has given an expression for the absorption coefficient for liquid water which varies as frequency squared so that $\tau_L = kf^2$, or $\tau_{L2} = (f_2/f_1)^2\,\tau_{L1}$. In a similar manner the dry opacities scale by frequency according to $\tau_{d2} = \beta(f_2/f_1)^2\,\tau_{d1}$, where the parameter β depends on f_1 and f_2. Thus, Eq. (17) is transformed to

$$\tau_2 = G_2(\Delta L_v) + \left(\frac{f_2}{f_1}\right)^2\left(\tau_{L1} + \beta\tau_{d1}\right) \qquad (18)$$

which, together with Eq. (16), can be solved to give

$$\Delta L_v = \left[G_1 - \left(\frac{f_1}{f_2}\right)^2 G_2\right]^{-1} \times \left[\tau_1 - \left(\frac{f_1}{f_2}\right)^2 \tau_2 - (1-\beta)\,\tau_{d1}\right] \qquad (19)$$

$$\tau_{L1} = \left[G_2 - \left(\frac{f_2}{f_1}\right)^2 G_1\right]^{-1} \times \left[G_2\,\tau_1 - G_1\,\tau_2 - \left(G_2 - \beta\left[\frac{f_2}{f_1}\right]^2 G_1\right)\tau_{d1}\right] \qquad (20)$$

As we did with the vapor term, we can assume that $\tau_L = Z_L M_L$, where Z_L is the weighting function for liquid water, and M_L is the precipitable liquid. Thus,

$$M_L = \left[G_2 - \left(\frac{f_2}{f_1}\right)^2 G_1\right]^{-1} \times \left[G_2\,\tau_1 - G_1\,\tau_2 - \left(G_2 - \beta\left[\frac{f_2}{f_1}\right]^2 G_1\right)\tau_{d1}\right] Z_L \qquad (21)$$

Equations (19) and (21) represent the formal solutions for the excess path delay due to water vapor ΔL_v, and the integrated liquid content M_L, in terms of the transformed observables τ_1 and τ_2 defined by Eq. (12).

If we are able to measure the brightness temperatures T_{b1} and T_{b2}, then in principle we can estimate the vapor path delay ΔL_v and integrated liquid content M_L. The accuracy of these estimates is limited by our ability to measure the brightness temperatures and the quality of our assumptions, namely, (1) that the quantities T_M, G_1, G_2, τ_{d1}, and Z_L are constant and (2) that the radiation from liquid water varies as f^2. Accurate measurement of T_b is primarily an instrumental calibration problem. In the next section the overall quality of these assumptions is examined directly.

III. Determination of Constants in the Algorithm

In order to use the inversion algorithms given in Eqs. (19) and (21), we must determine the mean radiating temperature of the atmosphere, the vapor weighting functions, the liquid weighting function, and the opacity due to the dry component of the atmosphere. These quantities have been assumed to be constant in the derivation of the algorithm, but in a real atmosphere they exhibit variation and correlation with other atmospheric parameters. While these variations are not so

large as to invalidate the basic assumptions, they clearly indicate some level of error in the algorithm. This implies that we must always expect some level of "algorithm noise," and our efforts must be directed towards its reduction. Furthermore, some fraction of this algorithm noise is likely to represent systematic variations that are a function of site and season. If we were given a meteorological history of each WVR site, we might be able to develop the constants in our algorithms in such a way that would be optimized for that site and hopefully reduce the seasonal variations. However, this procedure presents potential operational problems: (1) The meteorological history may not be available or easy to obtain; (2) it is a fair amount of work if there are many sites (e.g., as with mobile VLBI); and (3) someone must keep careful track of which algorithm goes with a particular site, and if that someone mixes the algorithms, the error could be compounded. A far simpler procedure would be to derive a single formulation of the algorithm in which the "constants" were no longer constant but instead some simple function of site and seasonal parameters, i.e., in effect we would use a model to reduce the site and seasonal systematic errors in the algorithm. Any WVR user must decide on which approach to use based on his accuracy requirements as well as cost and operational reliability. The interpretation adopted here is to use a single algorithm for all sites and to demonstrate that the residual site and season dependencies are less that the currently required level of accuracy.

We can best evaluate the level of algorithm noise as well as site and seasonal variations by determining the constants in a given formulation of the algorithm by using a regression analysis. That is, we will use meteorological data (e.g., radiosonde data) to compute the path delay and to solve the equation of radiative transfer for the associated brightness temperatures at 20.7 and 31.4 GHz for a relatively large number of cases. This data base will then be used to determine the value of the constants in an inversion algorithm that will minimize the residuals in a least squares sense. Of course, as a practical matter, we will still be left with the problem of relating the temperature scale of our WVRs to the temperature scale defined in Eq. (8). Since our knowledge of fundamental quantities like the absorption coefficients as well as our calibration of the radiometers will always be less than perfect, we must accept the fact that ultimately the WVRs must be compared directly with some independent method of measuring path delay if we wish an absolute calibration.

A data base of radiosonde data was assembled from five sites in the United States during the year 1976. The sites – Portland, Maine; Pittsburgh, Pennsylvania; El Paso, Texas; San Diego, California; and Oakland, California – were chosen to represent a cross section of meteorological conditions. Each radiosonde launch provides an approximately vertical profile of pressure, temperature, and relative humidity that is used to calculate the vapor path delay and to numerically solve the equation of radiative transfer for the brightness temperature, mean radiating temperature, and the opacity. One launch out of eight was selected from each site so as to obtain equal amounts of data from both 0 and 12 hours Universal Time and to cover seasonal trends in the data. Thus, for each radiosonde site we have 92 values of the vapor path delay ΔL_{vi} and 92 pairs of brightness temperatures $T1_i$ (20.7 GHz) and $T2_i$ (31.4 GHz). We use the latter to estimate the path delay $\langle \Delta L_v \rangle_i = f(T1_i, T2_i)$ and in a straightforward manner solve for the constants in any given functional form $f(T1_i, T2_i)$ that minimizes the difference $\Delta L_{vi} - \langle \Delta L_v \rangle_i$, in a least squares sense.

The radiosonde data does not provide any direct indication of the presence or amount of liquid water. As we noted previously, the presence of even small amounts of liquid (as in clouds) has a pronounced effect on the brightness temperatures measured by the WVR. Hence, it is essential that we evaluate the performance of the vapor retrieval algorithm in the presence of liquid. In order to simulate the presence of clouds the data from each radiosonde launch are scanned for an indication that the relative humidity is greater than 95%, which we assume indicates an equilibrium condition with liquid. The points at which the relative humidity falls below 94% define the top and bottom of the cloud, and the altitude of these points is calculated by simple linear interpolation. Given the cloud thickness and altitude, we use all three of the models for the cloud liquid density given by Decker *et al.* (Ref. 9) which we denote by CMODEL = 1, 2, or 3 (CMODEL = 0 denotes no liquid). The temperature of the liquid was taken to be the interpolated temperature of the radiosonde in the cloud, and the absorption coefficient of liquid water given by Staelin (Ref. 8) is used to calculate the brightness temperature. Using this criterion, a total of 121 radiosonde launches from the 5 sites was found that suggests the presence of liquid water. Thus, for each cloud model there were 121 values of the precipitable liquid M_{Li}, and 121 pairs of observables $T1_i$ and $T2_i$. The regression analysis for the liquid water retrieval then proceeds in a manner that is completely analogous to analysis used for the vapor algorithm. For the vapor algorithm, the constants are derived in the regression analysis using clear-sky data and tested with the cloud data. For the liquid algorithm, the constants are derived with the cloud data and tested with the clear-sky data.

Generally, the zenith brightness temperature in clear-sky conditions at both of our frequencies will be less than 50 K and the corresponding opacity less than 0.2 neper. This suggests that we could expand the logarithm in Eq. (20) and keep only the first order term in T_b, i.e., a low opacity approx-

imation. The estimated path delay would then be of the form

$$\langle \Delta L_v \rangle_i = A_0 + A_1 \, [T1 - 0.4346 \times T2]_i \qquad (22)$$

where $T1$ and $T2$ are the brightness temperatures at 20.7 and 31.4 GHz, respectively. Table 1 summarizes the best fit parameters for this estimate. First, note that since we are using noise free data, the RMS of the fit represents the quality of the assumptions that have gone into the estimate, i.e., "algorithm noise." We would expect a larger RMS than shown if we actually compared WVR data with real radiosondes for in that case we would be comparing two noisy observables and the RMS would represent the quadratic sum of the radiosonde error and WVR error. Second, note that the values of the "constants" A_0 and A_1 vary from site to site by more than their standard errors, clearly indicating systematic effects are present in the data. This is further emphasized by the fact that the RMS of the fit for all sites is larger than the RMS from any single site. Progressive degradation of the algorithm can be seen in the increasing RMS under cloudy conditions, i.e., increasing opacity. Analysis of the residuals indicates that they correlate with surface values of the pressure, temperature, and the opacity. Of these correlations the opacity is by far the most important. Since our data base represents only zenith values, both the variation and the absolute values of the opacity tend to be small. In a real experiment, the WVR may be pointed down to an elevation angle of ~15° and the range of opacities will vary accordingly. Still, the RMS of the fit for all sites is not too bad so that Eq. (22), since it is particularly simple, is adequate for a quick estimate of the delay.

In order to obtain some idea as to the performance of our algorithms in actual operation, we must know the instrumental noise spectrum imposed on the observables and must include an estimate of its magnitude in the regression analysis. One method that can be used to estimate the instrumental stability is to have two side-by-side radiometers observe the same target, e.g., the sky, and note the difference between the two brightness temperatures. In principal, this difference should appear to be Gaussian noise with an RMS equal to $\sqrt{2}$ times the RMS fluctuations of a single radiometer and can be reduced by simply increasing the integration time. In reality, the integration time can only be increased to the point where the inevitable systematic errors begin to predominate. Data taken during the testing and calibration of the WVRs indicate that the noise spectrum is "white" on time scales less than ~3 h and therefore can be reduced by averaging. For time scales greater than 3 h, flicker noise seems to predominate and appears as a slow drift of the antenna temperature about some nominal value with an amplitude of ±1 K. Since the geodetic experiments that we expect to support are normally longer than 3 h, we will model the radiometer noise with a uniformly distributed random variable drawn from the interval ±1 K, added to the brightness temperature. The regression analysis then proceeds as in the noise-free case. Differences in the RMS of the fit between the noisy and noise-free data indicate the relative importance on instrumental noise and systematics in the algorithm.

The simple formulation in Eq. (22) does not take into account that radiation from distant vapor along the line of sight is attenuated by intervening vapor. This correction is done explicitly by using the opacity as the transformed observable. Table 2 summarizes the best fit parameters for the path delay estimate that now includes the transformation to opacity given by Eq. (12) (where T_m = 275 K). Except for the El Paso data set, we see that the constants are reasonably consistent from site to site. Although there is a small bias for the cloud liquid data, the RMS of the estimate in the presence of liquid is consistent with the clear sky data – a definite improvement over the previous algorithm. When a uniformly distributed ±1 K of noise is added to the brightness temperatures, the RMS for all sites rises to 0.55 cm, indicating roughly equal contributions from algorithm noise and instrumental systematics. This form of the algorithm is useful in circumstances where measurements of surface temperature and pressure are not readily available.

When values of the surface pressure and temperature are available, we can use them to further refine our algorithm. The next most obvious parameter to model is the mean radiating temperature T_M. Figure 3 illustrates the frequency dependence of T_M in a plot of T_M versus frequency for a standard atmosphere containing an exponential distribution of water vapor with total columnar content M_v = 2 g/cm^2. Figure 4 shows how T_M varies as a function of M_v, again in a standard atmosphere. The mean radiating temperature is also a function of the physical temperature distribution in the atmosphere and will exhibit site and seasonal variation. Figure 5 illustrates this variation for the Portland, Maine, data for the year 1976, and Table 3 summarizes the statistics of both T_M and the surface temperature T_S. The RMS variations in T_M indicated in Table 3 suggest that this could be a significant error source in the inversion algorithm, and some effort is warranted to reduce this variation. If we assume a simple linear relationship between T_M and the surface temperature T_s, then the estimates

$$T_{M1} = 50.3 + 0.786 \, T_s \qquad (f = 20.7 \text{ GHz}) \qquad (23)$$

$$T_{M2} = T_{M1} + 3.4 \qquad (f = 31.4 \text{ GHz}) \qquad (24)$$

reduce the RMS variation of T_M from all sites by a factor of two. Table 4 summarizes the best linear fit between T_M and the surface temperature T_S for each site. Note that the RMS for all sites in Table 4 is larger than the RMS for any individual site. This strongly suggests that this simple linear fit does not completely remove all site-to-site and/or seasonal variations. However, these equations do reduce the RMS to an acceptable level and have the virtue of being simple to use – a significant consideration if one must deal with data from many sites.

For the next formulation of an algorithm, we will model the mean radiating temperatures T_{M1} and T_{M2} and assume that the dry opacity scales as the surface pressure squared times the surface temperature to the -2.86 power. The multiplier of the dry opacity term will be chosen to force the bias term, i.e., A_0, to be zero. The algorithm that we shall now fit then takes the form

$$\langle \Delta L_v \rangle = A_0 + A_1 \left[\tau_1 - 0.4346\,\tau_2 - \left(\frac{A_2}{A_1}\right)\left(\frac{P_s}{1013}\right)^2 \times \left(\frac{293}{T_s}\right)^{2.86} AM \right] \quad (25)$$

where we have included the air mass scaling for the dry term, i.e., AM = cosecant (elevation). The opacity at frequency F_i is

$$\tau_i = -\log_e \left(\frac{T_{mi} - T_{bi}}{T_{mi} - T_c}\right) \quad (12)$$

where T_{m1} is given by Eq. (23) and T_{m2} by Eq. (24). Table 5 summarizes the regression analysis for this algorithm. We see that the RMS and the site-to-site consistency is a bit better than the previous algorithm and the performance in the presence of liquid is about as good. If surface measurements are available, we would prefer this algorithm to the previous formulation, but note that the accuracy of either meets the calibration requirements of the Crustal Dynamics Project. When noise is added to the RMS, the fit for all sites rises to 0.48 cm and the small site-to-site differences are blurred by larger sigmas on each of the constants.

Since the path delay due to liquid is $\Delta L_L \sim 1.6\,M_L$ (for both ΔL_L and M_L in cm) and M_L is rarely larger than a few millimeters, we see that the liquid delay is considerably smaller than the errors in the vapor estimate. The primary reason for a liquid estimate is not for direct geodetic calibration so much as it is an indicator of WVR performance. We can rewrite Eq. (21) in the form

$$M_L = A_0 + A_1 \left[\tau_2 + \left(\frac{A_2}{A_1}\right)\tau_1 + \left(\frac{A_3}{A_1}\right)\tau_d \right] \quad (26)$$

The value of (A_2/A_1) is found either from calculation of the weighting functions or a multiple regression analysis on the radiosonde data base to be (A_2/A_1) = -0.366. We will use the same functional dependence for the dry term as we used in the vapor algorithm and require the value of (A_3/A_1) to be such as to minimize the value of A_0, i.e., we will minimize the bias term. Table 6 shows the parameters for the best fit solution for this formulation. As we required, the bias term A_0 is less than its standard error for each cloud model and can be taken as effectively zero. When this is done and the liquid water algorithm is used with the main radiosonde data base (where we assumed there was no liquid water), the average residual at each site is comfortably small with an RMS value comparable to the retrieval accuracy in the liquid data set. Note that units have switched to micrometers for the liquid measure. The increasing RMS of the retrieval with cloud model suggests that the accuracy of the liquid retrieval is a function of the liquid density. Figure 6 shows the average precipitable liquid versus the RMS of the retrieval for each site and each cloud model. The error in the liquid water estimate suggested by this data is

$$\text{RMS}\,(M_L) = 0.32 \times M_L \quad (27)$$

When we include the WVR error model in the regression analysis for liquid retrieval, the RMS of fit changes relatively little and suggests that the retrieval accuracy is limited by the algorithm. In fact, the algorithm for the liquid water estimate contains relatively more error than the vapor algorithm due to the fact that the liquid weighting function contains an exponential dependence on the liquid temperature (Ref. 8), which is not an easily modeled quantity. The presence of liquid water during a tip-curve calibration observation is an immediate indication that the tip-curve data will be noisy and should be weighted accordingly. If we are observing in clear sky conditions and the WVR reports a nonzero (i.e., greater than the RMS value) amount of liquid water, it suggests that either the radiometric temperature scales needs recalibration or the vapor weighting functions are very different than the "average" weighting function determined by earlier analysis. The latter possibility could be due to an unusual vertical profile of vapor although we have chosen the vapor sensitive frequency to minimize this type of error. Finally, the presence of large amounts of liquid or equivalently, large values of opacity at 31.4 GHz, indicate that an important assumption in our derivation may be violated. In large concentrations of liquid, the drop size tends to grow and scattering plays an increasingly important part in the apparent radiation spectrum. When the effective diameter

of the drop is on the order of the observing wavelength, the scattering process is termed Mie scattering, and the spectrum is more complex than the simple Rayleigh scattering that we have assumed. The radiation spectrum of large drops can no longer be characterized by a simple power-law-type behavior, which means that both our vapor and liquid algorithms break down. Since the breakdown is primarily a function of the drop size distribution of the liquid, there is no clear criterion that distinguishes the operating from the nonoperating regimes of the algorithm. Westwater (Ref. 10) and Westwater and Guiraud (Ref. 11) estimate that these remote sensing techniques are applicable up to opacities of 3 db = 0.7 neper (at 31 GHz). Given the considerable observing experience of these experimenters, this is probably the best operating/nonoperating criterion that can presently be stated.

IV. Summary

Several reasonably simple algorithms have been derived that relate observables, i.e., the brightness temperatures measured with a two-channel WVR, to the line-of-sight path delay that is required to correct various types of radiometric data for the effects of atmospheric water vapor. The three formulations that can be used to estimate the excess path delay due to atmospheric water vapor ΔL_v, are

$$\langle \Delta L_v \rangle = -1.6 + 0.65\,(T_1 - 0.435\,T_2) \tag{28}$$

for convenience. A better algorithm if no surface measurements are available would be

$$\langle \Delta L_v \rangle = 158\,(\tau_1 + 0.435\,\tau_2) \tag{29}$$

where the opacity τ, is

$$\tau_i = -\log_e \left[\frac{275 - T_i}{272}\right] \tag{30}$$

When surface measurements are available, use

$$\langle \Delta L_v \rangle = 164 \left[\tau_1 - 0.435\,\tau_2 - 0.0016 \times \left(\frac{P_s}{1013}\right)^2 \left(\frac{293}{T_s}\right)^{2.86}\right] \tag{31}$$

where the opacity is now

$$\tau_i = -\log_e \left[\frac{T_{Mi} - T_i}{T_{Mi} - T_c}\right] \tag{32}$$

and the quantities T_{Mi} are given in Eq. (23) and (24).

The regression analysis indicates that the ultimate accuracy of the two-channel technique is about 0.2 to 0.3 cm, and with a realistic noise model of the current generation of instruments a 0.5 cm accuracy is possible. Note that the question of radiosonde accuracy is irrelevant in this analysis: It simply represents the "truth." However, if the WVRs are to be compared directly to radiosondes, then the question of radiosonde accuracy is crucially important. Indeed, the WVR's must eventually be calibrated on an absolute scale, and radiosondes seem to be the most cost-effective way to accomplish this at the present time. Consider that such a direct comparison is made at each of the radiosonde launch sites that were used in this analysis. If we assume an average zenith path delay of 10 cm and the measurement accuracy of the radiosonde is 10%, then the RMS of the radiosonde/WVR comparison would be,

$$(1.0)^2 + (0.5)^2 = 1.12 \text{ cm}$$

While it is important to do such a comparison so as to investigate the possibility of a bias term in the inversion, the analysis suggests that the WVR is more accurate than the radiosonde. Hence, the comparison must be done with great caution when attempting improvements in the WVR instrumentation or refinements in the algorithm. We are faced with a recurring problem – how to demonstrate a new measurement technique is as good as we think it is when it is better than any existing technique.

References

1. Resch, G. M. (1980), "Water vapor: The wet blanket of microwave interferometry," eds., Deepak, Wilkerson, and Ruhnke, *Atmospheric Water Vapor*, Academic Press, New York.

2. Bean, B. R., and E. J. Dutton (1968), *Radio Meteorology*, Dover Publ., New York, N.Y.

3. Smith Jr., E. K., and S. Weintraub (1953), "The constants in the equation for atmospheric refractive index at radio frequencies," *Proc. I.R.E.*, Vol. 41, pp. 1035-1057.

4. Hess, S. L. (1959), *Introduction to Theoretical Meteorology*, Holt, Rinehart & Winston, New York.

5. Chandrasekhar, S. (1950), *Radiative Transfer*, Oxford Univ. Press, New Jersey.

6. Waters, J. W. (1976), "Absorption and emission by atmospheric gases," ed., M. L. Meeks, *Methods of Experimental Physics: Astrophys.*, Part B, 12, pp. 142-175.

7. Snider, J. B., and E. R. Westwater (1969), "Atmospheric attenuation at 15, 31, and 53 GHz," *ESSA Tech. Rept. ERL 156-WPL 11*, Dec. 1969.

8. Staelin, D. H. (1966), "Measurements and interpretation of the microwave spectrum of the terrestrial atmosphere near 1-cm wavelength," *J. Geophys. Res., 71*, pp. 2875-2882.

9. Decker, M. T., E. R. Westwater, and F. O. Guiraud (1978), "Experimental evaluation of ground-based microwave radiometric sensing of atmospheric temperature and water vapor profiles," *J. Appl. Meteor., 17*, pp. 1788-1795.

10. Westwater, E. R. (1978), "The accuracy of water vapor and cloud liquid determinations by dual-frequency ground based microwave radiometry," *Radio Science, 13*, pp. 677-685.

11. Westwater, E. R., and F. O. Guiraud (1980), "Ground-based microwave radiometric retrieval of precipitable water vapor in the presence of clouds with high liquid content," *Radio Science, 15*, pp. 947-957.

Table 1. Summary of best fit parameters for a vapor algorithm involving brightness temperature

$$\langle \Delta L_v \rangle = A_0 + A_1 \{ T1 - 0.4346\, T2 \}$$

(no measurement noise)

Cloud Model = 0					
Site	A_0	σ	A_1	σ	RMS, cm
Portland	−1.57	0.05	0.662	0.003	0.28
Pittsburgh	−1.45	0.05	0.649	0.003	0.25
El Paso	−1.39	0.06	0.610	0.004	0.26
San Diego	−1.71	0.13	0.642	0.007	0.38
Oakland	−1.63	0.13	0.644	0.007	0.35
All sites	−1.62	0.05	0.646	0.002	0.41

Cloud Model	Average Residual	RMS
1	0.29	0.45
2	0.20	0.58
3	−0.26	1.55

Table 2. Summary of best fit parameters for a vapor algorithm involving the opacities

$$\langle \Delta L_v \rangle = A_o + A_i \{ \tau_1 - 0.4346\, \tau_2 \}$$

$$\tau_i = -\log_e \left(\frac{275 - TA_i}{272} \right)$$

Cloud Model = 0					
Site	A_0	σ	A_1	σ	RMS[a], cm
Portland	0.01	0.04	160.5	0.6	0.24 (.49)
Pittsburgh	0.08	0.05	158.3	0.7	0.24 (.45)
El Paso	−0.03	0.05	151.4	0.9	0.25 (.45)
San Diego	−0.07	0.11	156.9	1.6	0.37 (.54)
Oakland	−0.07	0.10	158.6	1.7	0.32 (.53)
All sites	−0.06	0.03	157.9	0.5	0.36 (.55)

Cloud Model	Average Residual	RMS[a]
1	0.18	0.35 (.56)
2	0.17	0.35 (.51)
3	0.14	0.37 (.60)

[a]Numbers in parenthesis indicate the RMS with an added noise of ±1 K in each channel.

Table 3. The average mean radiating temperature, the surface temperature, and their RMS values (f = 20.7 GHz)

Site	T_M	RMS	T_S	RMS
Portland	269.4	11.4	279.1	10.8
Pittsburgh	270.3	10.9	281.5	11.3
El Paso	275.2	7.0	289.5	6.0
San Diego	280.1	5.1	290.6	5.7
Oakland	278.0	5.3	287.7	6.0
All sites	274.7	9.4	285.7	10.5

Table 4. Estimates of the mean radiating temperature from the surface temperature ($T_{m1} = A_0 + A_1 T_S$, f = 20.7 GHz)

Site	A_0	A_1	RMS
Portland	6.6	0.94	4.2
Pittsburgh	15.3	0.91	3.7
El Paso	121.8	0.53	3.9
San Diego	86.1	0.67	3.4
Oakland	110.7	0.58	3.9
All sites	50.2	0.786	4.5

Table 5. Summary of best fit parameters for a vapor algorithm involving opacities and surface data

$$\langle \Delta L_v \rangle = A_0 + A_1 \left\{ \tau_1 - 0.4346\, \tau_2 - 0.0016\, \tau_d \right\}$$

where

$$\tau_i = -\log_e \left(\frac{T_m - TA_i}{T_m - T_c} \right)$$

$$T_{m1} = 50.3 + 0.786 \times T_s$$

$$T_{m2} = T_{m1} - 3.4$$

$$\tau_d = \left(\frac{P_s}{1013} \right)^2 \left(\frac{293}{T_s} \right)^{2.86}$$

Cloud Model = 0					
Site	A_0	σ	A_1	σ	RMS[a], cm
Portland	0.07	0.03	165.4	0.5	0.18 (.45)
Pittsburgh	0.11	0.03	165.1	0.5	0.16 (.44)
El Paso	−0.05	0.03	159.4	0.5	0.14 (.39)
San Diego	−0.03	0.09	163.7	1.3	0.30 (.47)
Oakland	0.05	0.09	163.9	1.5	0.27 (.44)
All sites	−0.001	0.03	163.9	0.4	0.28 (.48)

Cloud Model	Average Residual	RMS[a]
1	0.13	0.30 (.46)
2	0.19	0.34 (.54)
3	0.29	0.50 (.63)

[a]Numbers in parenthesis indicate RMS with an added noise of ±1 K to each channel.

Table 6. Summary of best fit parameters (μm) for a liquid water algorithm

$$\langle M_L \rangle = A_0 + A_1 (\tau_2 - 0.366, \tau_1 - 0.022\,\tau_d)$$

All Sites Cloud Model	A_0	σ	A_1	σ	RMS[a]
1	−0.3	2.8	5368	107	24 (26)
2	+8.4	12.1	5352	119	105 (108)
3	+26.7	25.7	5279	125	224 (223)

Cloud Model = 0 (no liquid)

Site	Average Residual	RMS[a]
Portland	0.6	33.0 (34)
Pittsburgh	0.3	25.0 (34)
El Paso	11.8	9.6 (15)
San Diego	−7.2	15.6 (22)
Oakland	−3.6	12.7 (17)
All sites	0.3	21.8 (25)

[a]Numbers in parenthesis indicate RMS with an added noise of ±1 K to each channel.

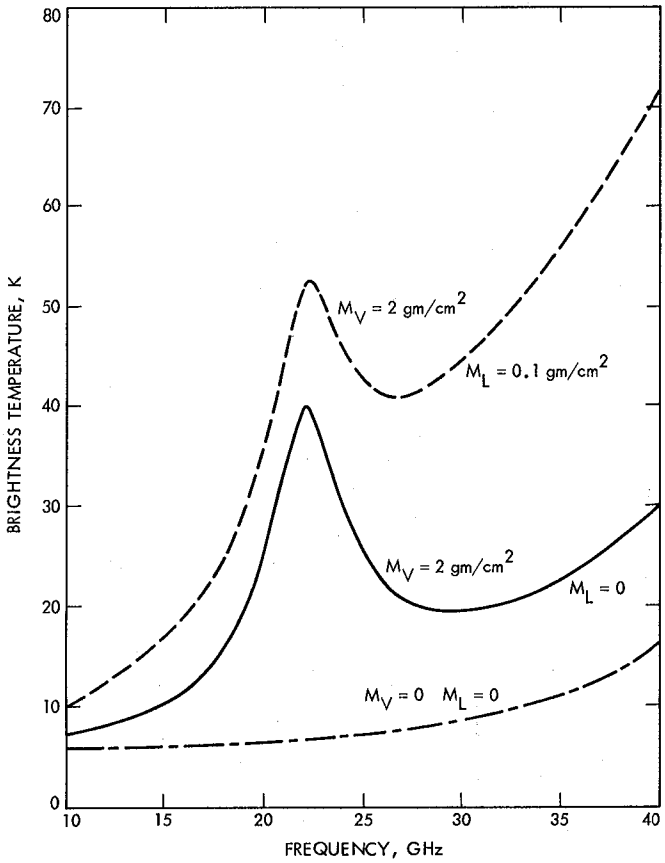


Fig. 1. The zenith brightness temperature of a standard atmosphere for three cases: (1) no water vapor and no liquid water, (2) 2 g/cm^2 of vapor and no liquid, and (3) 2 g/cm^2 of vapor and 0.1 g/cm^2 of liquid

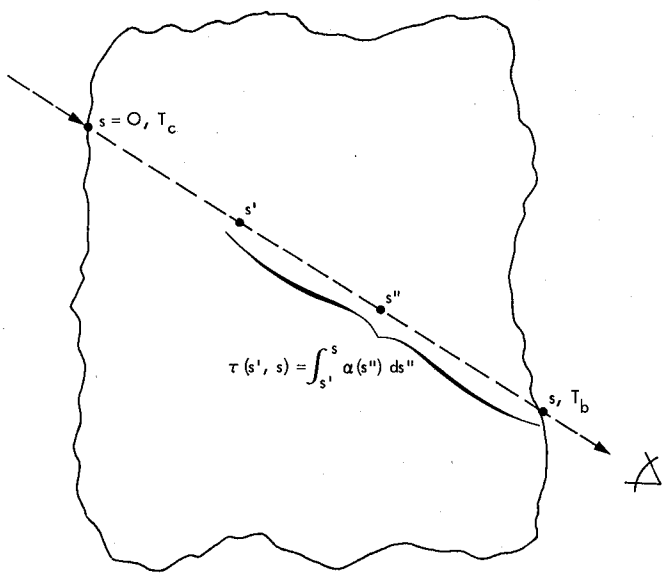


Fig. 2. Schematic representation of the equation of radiative transfer

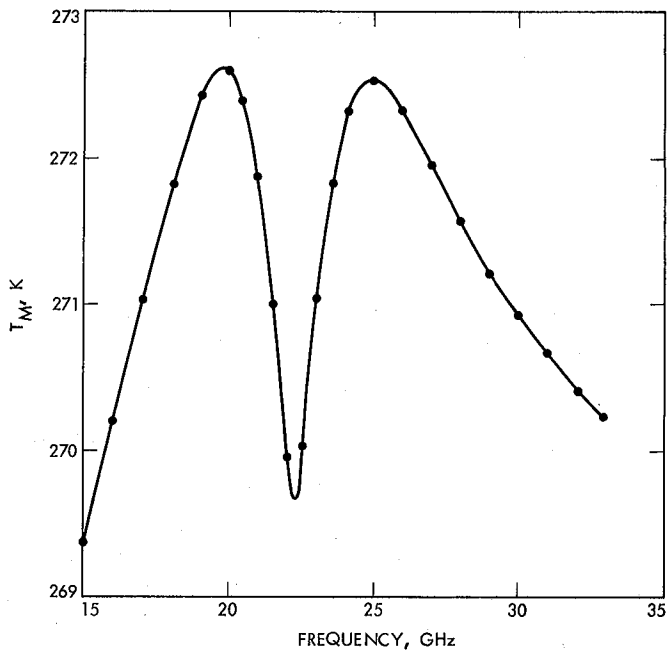


Fig. 3. The mean radiating temperature of a standard atmosphere T_M vs. frequency for M_v = 2 g/cm^2 exponentially distributed in a standard atmosphere

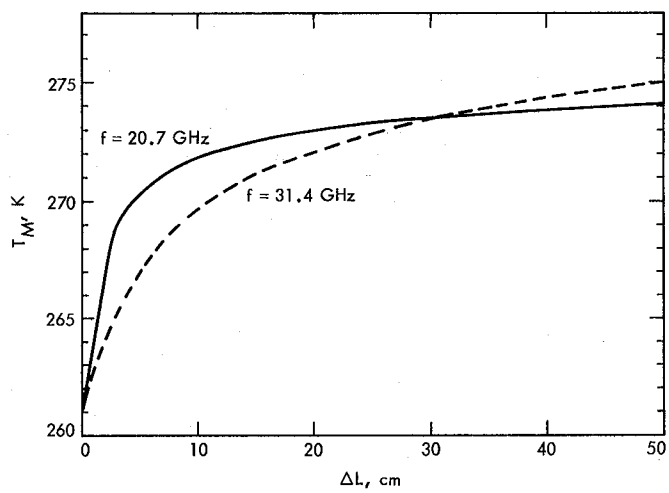


Fig. 4. T_M vs. M_v for f = 20.7 and 31.4 GHz in a standard atmosphere

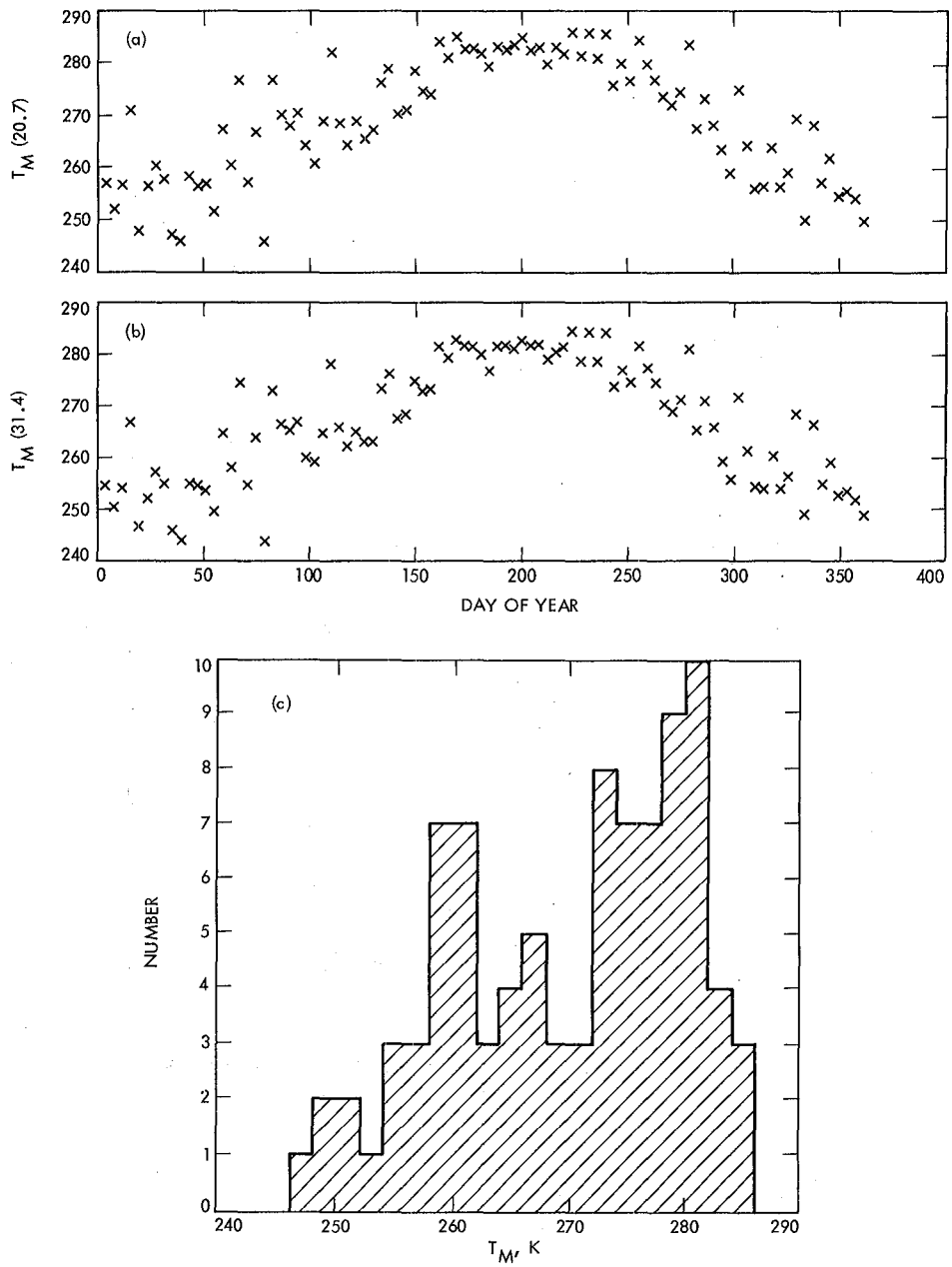


Fig. 5. Portland, Maine, radiosonde data showing (a) T_M vs. time at $f = 20.7$ GHz; (b) T_M vs. time at $f = 31.4$ GHz; and (c) the histogram of T_M at 20.7 GHz

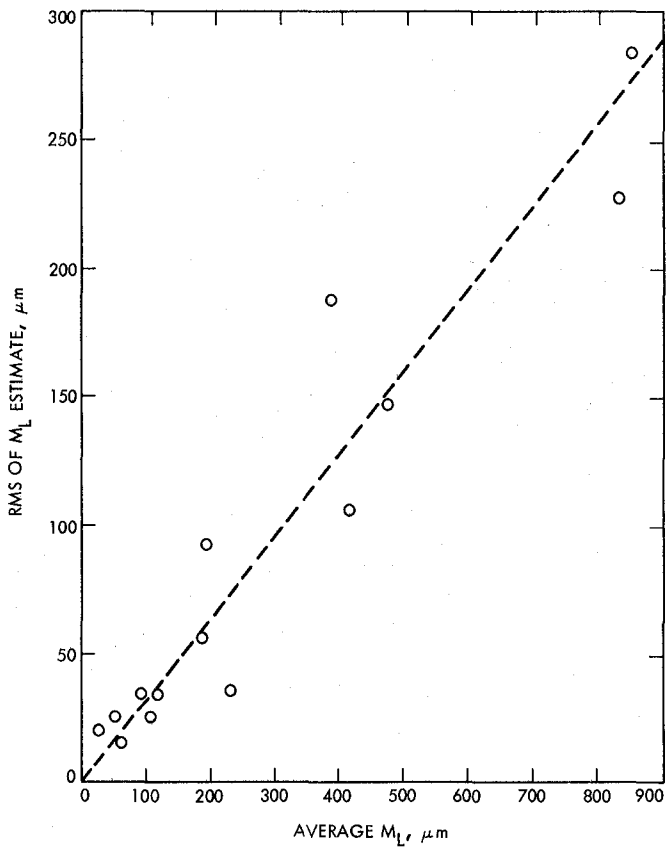


Fig. 6. The RMS of the liquid water retrieval vs. M_L, the precipitable liquid along the line of sight

DSN Microwave Antenna Holography

D. J. Rochblatt
B. L. Seidel
Radio Frequency and Microwave Subsystems Section

The DSN microwave antenna holography project will obtain three-dimensional "pictures" of the large DSN antenna surfaces. These "pictures" must be of sufficient resolution to allow adjustment of the reflector panels to an rms surface of 0.5 mm (0.25 mm, goal).

This article outlines the major parameters and equations needed to define a holographic measurement system and then discusses the proof of concept demonstration measurement that was made at DSS-43 (Australia) that resulted in contour maps with spatial resolution of 7 m in the aperture plane and resolution orthogonal to the aperture plane of 0.7 mm.

I. Introduction

Microwave holography as applied to antennas is a technique that utilizes the Fourier transform relationship between the complex far-field radiation pattern of an antenna and the complex aperture field distribution. Contour maps of a reflector antenna surface can be inferred from the phase portion of the aperture field distribution. This technique has now been successfully applied to a number of large reflector antennas around the world.

To be useful, the map must have sufficient resolution to allow the individual panels of a reflector antenna to be adjusted to an accuracy that will allow the antenna to operate at high efficiency.

Presently, optical techniques are used to measure the large DSN tracking antennas. This technique requires removal of major components of the antenna system (because of blockage) and requires up to 8 weeks of antenna down time. By contrast, the microwave holography technique leaves the antenna system intact and is estimated to require about 12 hours (of antenna time) to collect the data needed for the requisite resolution on a 64-m antenna. The antenna can then be immediately placed back in operation and the panels can be adjusted over many weeks during normal maintenance periods, a major advantage.

Early attempts at JPL to produce holographic surface maps of a large antenna were conducted using a radio star (3C273) as a far field source. These experiments were conducted using DSS-43 as the test antenna and DSS-42 as the reference antenna. Both the VLBI and the connected element interferometer (CEI) equipments were used to take data. These data, while providing the basis for our first low resolution holographic surface maps, also pointed out unanticipated structural

dynamics problems and the futility of using radio stars for an illumination source.

II. Antenna Efficiency

The ultimate performance of a large, steerable, reflector antenna is limited by imperfections of the reflecting surface. The size of the panels that form the surface of the antenna and the allowed level of losses due to surface inaccuracies dictate the required resolution of the measurements. From the size of the panels on the DSN antennas, we determine that a spatial resolution of 0.4 m is needed in the x, y (aperture) plane. For a maximum of 0.1 dB degradation in antenna efficiency due to surface imperfections, the rms surface error (ϵ) must be no greater than 0.012 λ (where λ is the wavelength):

$$\epsilon_{-0.1\ \mathrm{dB}} \leqslant 0.012\,\lambda \tag{1}$$

The relationship between the efficiency of a physical antenna and a hypothetical antenna with absence of phase errors across its surface is given (for $\epsilon/\lambda \leqslant 1/4\pi$) by (Refs. 1, 2, 3)

$$\frac{\eta_A}{\eta_0} = \exp - \left(\frac{4\pi\epsilon}{\lambda}\right)^2 + \left(\frac{2r_c}{D}\right)\left[1 - \exp - \left(\frac{4\pi\epsilon}{\lambda}\right)^2\right] \tag{2}$$

where

η_A = antenna aperture efficiency

η_0 = antenna aperture efficiency with zero phase error

ϵ = rms surface deviation

λ = wavelength

r_c = correlation radius

D = antenna diameter

For totally random surface phase errors, $r_c = 0$, Eq. (2) reduces to its first term. When r_c is significant relative to D, the second term in Eq. (2) adds to the first term to yield a higher efficiency than the case where the errors are indeed totally random.

Therefore, in a worst case, we see that

$$\frac{\eta_A}{\eta_0} = \exp - \left(\frac{4\pi\epsilon}{\lambda}\right)^2 \tag{3}$$

In general, the rms surface error (ϵ) of a perfectly focused DSN antenna at an elevation angle of 45 deg is composed of the following three major components (Ref. 4):

(1) surface panel setting errors (SPS)

(2) primary surface panel manufacturing errors (PSP)

(3) subreflector manufacturing errors (SM)

Presently, these components for the DSN 64-m antennas are

SPS = 0.48 mm

PSP = 0.89 mm

SM = 0.30 mm

for a total rms surface error of $\epsilon = 1.06$ mm. At an operating frequency of 8440 MHz this rms surface error corresponds to 0.6 dB degradation in antenna efficiency relative to an antenna with a perfect surface. Part of the 64-m antenna upgrade project has set the following goals for improvement in the antenna surface:

SPS = 0.25 mm

PSP = 0.25 mm

SM = 0.25 mm

These goals would result in a total rms surface error at an elevation angle of 45 deg of $\epsilon = 0.44$ mm. At an operating frequency of 8440 MHz, this rms surface error corresponds to 0.1 dB degradation in antenna efficiency relative to an antenna with a perfect surface. This would provide a significant 0.5 dB efficiency improvement from present capability.

III. Spatial Resolution

The first step in the process of acquiring the contour maps is to obtain the antenna complex far field pattern. To do this, two antennas are used connected to a two channel receiver-correlator (Fig. 1) to observe the incoming signal from a far field source. While the (usually smaller) reference antenna is observing the incoming signal on boresight, the test antenna is scanning the transmitting source. Samples of the signal are taken at prescribed points on a (usually convenient) rectangular grid matrix. The number of data points collected is related to the required spatial resolution in the aperture plane. The separation between each data point must be less than one beamwidth.

Consider a square grid containing N^2 sampled data points separated by less than λ/D or one beamwidth (along the scan axis). Let

$$\Delta = \frac{K\lambda}{D} \tag{4}$$

where

Δ = separation between two adjacent sampled data points

λ = wavelength

D = antenna diameter

K = a constant $0.5 < K < 1.0$

The length of each side of the grid is $(N-1)\,K(\lambda/D)$ which for large N becomes

$$L = NK\left(\frac{\lambda}{D}\right) \tag{5}$$

where

L = Length of square grid

N^2 = Total number of sampled data points

Consider the Fourier transform of a rectangular pulse extending from $(+N/2)(K\lambda/D)$ to $(-N/2)\,K(\lambda/D)$ (Fig. 2).

The function transforms from a pulse in the θ domain to a

$$\frac{\sin\left(\frac{\pi NK\lambda x}{D}\right)}{\left(\frac{\pi NK\lambda x}{D}\right)}$$

function in the X domain whose first two nulls occur at $\pm D/K\lambda N$ for a full null width of $2D/K\lambda N$.

We define spatial resolution to be at the 50% width (eliminating the λ) to obtain

$$\delta = \frac{D}{KN} \tag{6}$$

where

δ = spatial resolution in the aperture plane

For a rectangular grid matrix of $N \times N$ data points, N is given by

$$N = \frac{D}{K\delta} \tag{7}$$

For the large DSN antennas where $D = 64$ m and the required spatial resolution based on individual panel size consideration is $\delta = 0.4$ m, and we use $K = 0.8$, the number of sampled data points required is

$$N^2 = \left[\frac{D}{K\delta}\right]^2 = \left[\frac{64}{(0.8)(0.4)}\right]^2 = 40{,}000$$

or a grid of 200 X 200.

Because of symmetry requirements about the boresight, the linear dimension of the required grid must have an odd number of data points. We therefore increase the size of the required grid to 201 X 201 data points.

IV. Signal-to-Noise Ratio Requirement

It can be shown (L. E. Young, private communication) that the voltage signal-to-noise ratio (SNR) on the aperture (which is related to the surface accuracy) is related to the SNR of a complex correlation taken with both antennas receiving the signal on boresight by

$$\mathrm{SNR}(A) = \frac{\mathrm{SNR}(O)}{N} \tag{8}$$

where

SNR(A) = signal-to-noise ratio on the aperture

SNR(O) = signal-to-noise ratio of a complex correlation taken at boresight

N = the square root of the total number of sampled data points (Eq. 7)

Equation (8) is based on the Fourier transform relationship between the complex correlation of a sample in the antenna far field distribution and the complex aperture distribution elements. It is derived using basic error analysis and assuming that each complex element in the aperture distribution has the same error for each sampled data point.

The equivalent accuracy on the aperture is given by

$$\epsilon(a) = \frac{\lambda}{2\pi\,\mathrm{SNR}(A)} \tag{9}$$

We can now relate the distortion on the surface $\epsilon(s)$ to a distortion on the aperture by (Ref. 5)

$$\epsilon(s) = \frac{\epsilon(a)}{2\cos\left(\frac{\psi}{2}\right)} \tag{10}$$

where

ψ = Angle between horn to subreflector to main reflector point of distortion.

For the DSN 64-m antenna the angle ψ has a maximum value of about 60 deg where $\epsilon(s)$ assumes the largest value or worst case; hence,

$$\epsilon \equiv \epsilon(s)_{\max} = \frac{\epsilon(a)}{1.73} \tag{11}$$

Using Eqs. (8), (9), (10), and (11), we calculate for SNR(O):

$$\mathrm{SNR} \equiv \mathrm{SNR}(O) = \frac{0.29\,N\lambda}{\pi\epsilon} \tag{12}$$

Equation (12) relates the SNR requirement to the rms surface resolution and the number of sampled data points needed. We may use Eq. (7) to express the required voltage SNR as:

$$\mathrm{SNR} = \frac{0.29\,D\lambda}{\pi K\delta\epsilon} \tag{13}$$

V. Far-Field Source

The large sampled data array that must be measured in order to obtain a spatial resolution of 0.4 m and the antenna scan rate (the array size needed of 201 X 201 is independent of the frequency selected for the measurement; see Eq. 7) dictates the time required for the measurement. Because of the antenna mechanical dynamics characteristics, 20 s were needed to move the DSS-43 antenna between adjacent data points. This time is needed for the antenna movement to damp out, if the antenna is to be stopped at each sampled data point. From the above we conclude that more than 200 h would be needed for the measurement alone (independent of frequency), an unacceptable situation. The conclusion that sampled data must be taken while the antenna is continuously scanning is therefore adopted.

Strong artificial sources are needed to provide the required SNR in the short integration times necessitated by continuous scanning.

By allocation, no geosynchronous satellites are permitted in the DSN portion of X-band (8.4 - 8.5 GHz). Geosynchronous satellites are available, Defense Satellite Communications Systems (DSCS II), with transmitting sources at S-band (2.2 GHz). With some modification, the DSN antennas can be fitted to receive the DSCS II signals transmitted at 7.7 GHz.

To achieve the required rms surface precision of 0.5 mm and spatial resolution of 0.4 m in the holography maps, the following SNRs are needed (Eqs. 12, 13): 63.2 dB at 7.7 GHz or 74.0 dB at 2.2 GHz. To achieve the above SNR requirements, a geosynchronous satellite with an approximate effective isotropic radiated power (EIRP) of +40 dBm is needed in the direction of the DSN 64-m ground antenna.

VI. Elevation Angle Effects

The Nth sampled data points are taken off the Mth side lobe given by

$$M = \frac{D}{2\delta} \tag{14}$$

where

M = the outermost side lobe off boresight

For the 64-m DSN antenna,

$$M = 80$$

The elevation angle excursion can now be calculated by

$$\theta_{\max} = M\left(\frac{\lambda}{D}\right)\frac{180}{\pi}\ \mathrm{deg}$$

For the 64-m antennas this yields

$$\theta_{2.2\ \mathrm{GHz}} = \pm 9.7\ \mathrm{deg}$$

$$\theta_{7.7\ \mathrm{GHz}} = \pm 2.8\ \mathrm{deg}$$

The wide elevation angle excursions needed for high resolution holography mapping will, to a first order, not smear the "longwave"[1] gravitational distortions of the large reflector. This can be understood in a few different ways. Recall the Fourier transform relationship between the antenna aperture response and its far field response and for a moment observe the relationship:

$$\cos(\omega_0 t) \leftarrow \text{F.T.} \rightarrow \pi\left[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)\right]$$

where here

δ = Dirac delta function

The higher the frequency in the time domain, the further apart the delta functions in the frequency domain separate from the $\omega = 0$ axis, and vise versa.

[1] We use the term "longwave" to signify large portions of the diameter, say $D/4$ or larger.

This relationship is analogous to the ripple on the antenna surface (the cosine function) which transforms into delta type functions in the antenna far-field pattern. It is clear that long gravitational distortions (low frequency) will reveal themselves in the first few side lobes of the antenna far-field pattern (delta functions close to main axis) while the short-wave ripples (high frequency) will reveal themselves as delta functions in the far side lobes.

The "forgiving" nature of the large antenna holography business is this: in fact, imbedded in the central "core" array (say, 11 X 11) is the information needed to "see" the long-wave distortions, and this can be obtained with narrower elevation angle excursions. The short time required to scan 11 X 11 also minimizes possible thermal effects. Imbedded in the "outer limits" of a big array (say, 201 X 201) is the information needed to see the shortwave "ripples" and these (to a first order) do not depend on elevation angle (these originated at the "factory" and are not due to gravity changes with elevation angle). Thus, we are not much concerned with ±2.8 deg (X-band) or even ±9.7 deg (S-band) elevation angle excursions possibly "smearing" the "pictures."

VII. Data Taking

To enable the completion of a test in a reasonable time (a period of 12 h seems sensible), data must be collected while the antenna is continuously moving. The following analysis is aimed at finding relationships between dynamic range, quantizer resolution, integration time, and scan speed. The latter is limited for the large DSN antennas to a range of 0.001 to 0.25 deg/s (JPL Document 810-5, Rev. D, Volume 1, "Deep Space Network/Flight Project Interface Design Handbook").

Studying CCIR (Comite Consultatif International Radio) generalized pattern and calculated radiation pattern envelopes for $D/\lambda = 500$ (Ref. 6) reveals that the 80th side lobe may be at -70 dB to -80 dB level relative to the peak of the main lobe.

The first few side lobes of the far-field amplitude pattern for a circular aperture uniformly illuminated antenna can be approximated by

$$h(\theta) = 2\left[\frac{J_1(\theta)}{\theta}\right] \tag{15}$$

where

$J_1(\theta)$ = Bessel function of the first kind

The rate of change of the pattern with respect to θ is

$$\frac{dh(\theta)}{d\theta} = 2\theta\left[\frac{J_0(\theta) - 2J_1(\theta)/\theta}{\theta^2}\right] \tag{16}$$

where

$$\frac{dJ_1(\theta)}{d\theta} = J_0 - J_1(\theta)/\theta \tag{17}$$

At the first null of $J_1(\theta)$ the slope is

$$\left|\frac{dh(\theta)}{d\theta}\right|_{J_1(\theta) = 0} = \left|2\left[\frac{J_0(3.83171)}{3.83171}\right]\right| = 0.2102 \text{ v/rad} \tag{18}$$

where the first zero occurs at $\theta = 3.83171$. This null occurs at a different angle for different antennas and is a function of both λ and D. In our case, it occurs at $1.22\,\lambda/D$ off boresight. We can now calculate the change in voltage amplitude ΔV occurring during a small scan $\Delta\theta$ and constrain it to be smaller than one part in 2^{B-1} in its limiting case (where B is the quantizer resolution).

$$\Delta V = \Delta\theta\left[\left|\frac{dh(\theta)}{d\theta}\right|_{J_1(\theta) = 0}\right]\left[\frac{3.83171}{1.22\frac{\lambda}{D}}\right] \leqslant \frac{1}{2^{B-1}} \tag{19}$$

A 12-bit signed quantizer is needed for a voltage dynamic range of 66 dB. Using $D = 64$ m, we obtain

at S-band: $\Delta\theta s = 9.03 \times 10^{-5}$ deg

and at X-band: $\Delta\theta x = 2.58 \times 10^{-5}$ deg

The DSN 64-m antennas have a maximum scan rate of 0.25 deg/s. Let us consider the scan rates in the range of

$$0.010 \text{ deg/s} \leqslant t_{\text{scan}} \leqslant 0.25 \text{ deg/s}$$

The integration time is given by

$$\tau = \frac{\Delta\theta}{t_{\text{scan}}} \tag{20}$$

which from Eq. (19) and the scan ranges shown becomes,

for S-band: $9.0 \text{ ms} \geqslant \tau \geqslant 0.36 \text{ ms}$

for X-band: $2.6 \text{ ms} \geqslant \tau \geqslant 0.1 \text{ ms}$

We can estimate the total required measurement time, T, by dividing the scan angle by the scan rate and multiplying by the number of scan lines, adding the retrace time to that number (calculated at the maximum scan rate), and adding 20% for calibration points and antenna mechanical effects:

$$T = N\left[\frac{2\theta}{t_{\text{scan}}} + \frac{2\theta}{0.25}\right](1.2)\ \text{s}$$

We have previously found that the scan angles required are ±9.7 deg at S-band and ±2.8 deg at X-band.

The scan rates given earlier, therefore, lead to estimated measurement times, T, of

$$135\ \text{h} \geqslant T_{\text{S-band}} \geqslant 10.4\ \text{h}$$

or

$$39\ \text{h} \geqslant T_{\text{X-band}} \geqslant 3\ \text{h}$$

Clearly, the higher scan rates are desirable.

Indeed, the final integration time and scan rate selected will depend on the signal source strength, antenna system temperature, bandwidth used, and antenna diameter.

VIII. DSS-43-42 Experiment

Preliminary holographic measurements were conducted at DSS-43-42 using the Tidbinbilla connected element interferometer (C.E.I.). DSS-43, the 64-m antenna, was measured while DSS-42, the 34-m antenna, was used as the reference; 3C273 was used as the signal source at 2.28 GHz with flux density of approximately 35 Jy[2]. The system temperature at zenith is approximately 20 K. The test was done at approximately 40° elevation where the system temperature was ~25 K. The C.E.I. has a detection sensitivity of 50 mJy in 1 s. The IF bandwidth was 12 MHz. The baseline between the two antennas is 194.6 m (1485λ, north-south) and is located near Canberra (Lat. 35.4° S, Long. 149° E). A phase-coherent local-oscillator distribution system is used to make the differential phase insensitive to variation in cable lengths (Ref. 7).

For a given source flux density, the source temperature at the antenna is given by (Ref. 8)

$$T_a = \frac{10^{-26}\,\eta}{2k}\,A_g\,S \qquad (21)$$

where

T_a = Source temperature at the antenna (K)

η = Antenna efficiency

A_g = Projected geometrical area of the antenna

k = Boltzmann's constant (1.38×10^{-23} J/K)

S = Source flux density (Jy)

1/2 = A factor to account for randomly polarized source

We can now define a minimum detectable signal:

$$\Delta T = \frac{T_s}{[2\,B\,\tau]^{1/2}} \qquad (22)$$

where

ΔT = Minimum detectable signal

B = Predetection bandwidth

τ = Postdetection integration time

T_s = System temperature

For the C.E.I., A_g is the geometric mean of the areas of the test and reference antennas:

$$A_g = \pi\left[\frac{D_1\,D_2}{4}\right] \qquad (23)$$

where

D_1 = Test antenna diameter

D_2 = Reference antenna diameter

Similarly,

$$T_s = \left[T_{s_1}\ \ T_{s_2}\right]^{1/2} \qquad (24)$$

where

T_s = system temperature

T_{s_1} = Test antenna temperature

T_{s_2} = Reference antenna temperature

[2] Jy = 10^{-26} W/m^2 Hz

We can now define signal to noise ratio of the system as

$$\text{SNR} = \frac{T_a}{\Delta T} \tag{25}$$

Combining Eqs. (21), (22), (23), (24), and (25) yields:

$$\text{SNR} = 4.03 \times 10^{-4}\, \eta\, SCD_1 D_2 \left[\frac{B\tau}{T_{s_1} T_{s_2}}\right]^{1/2} \tag{26}$$

where η is the antenna efficiency and C is the correlation efficiency.

During this experiment an integration time of 10 s was used at each observed point, η is approximately 0.6, and the correlation efficiency is approximately 0.5. This yields

$$\text{SNR} = 72 \text{ dB}$$

A sampling factor $K = 0.8$ was used during the measurement of the 11-X-11 grid, to achieve a spatial resolution of (Eq. 6)

$$\delta = \frac{D}{KN} = \frac{64}{(0.8)(11)} \cong 7 \text{ m}$$

Using Eq. (12), we may calculate ϵ to be equal to 0.03 mm. The hologram presented (Fig. 3) has an rms surface error of 0.7 mm; however, it seems to be well above the noise level.

The data was collected using HA/DEC coordinates, and 20 s were allowed for movement between adjacent sampled points, and 30 s were allowed to move to the boresight for calibration. At each sampled data point, the system was sampled for 10 s (integration time). The data collected during this test were recorded on diskette and brought to JPL for software processing. This software package that produced the plots and maps shown (Figs. 5 through 11) is described in the next section (software development) (Rahmat-Samii, Y., private communication).

In a time difference of approximately 1.5 h, a second C.E.I. measurement was conducted in the exact manner as the first experiment. A presentation of the plots is shown (Fig. 4) which compares well with plots obtained during the first experiment in both amplitude and surface profile contour. The only major difference between the maps is an axial rotation of about 17.5° which results from the change in the source location during the time period between the two experiments (Rahmat-Samii, Y., private communication).

Another set of data was collected at DSS-43-42 using the VLBI block zero system. Using a bandwidth of 1.8 MHz and sampling rate of 3×10^6 samples/s, a voltage SNR of 65 dB was achieved with 10-s integration time.

$$\text{SNR} = CA\, [B\, \tau]^{1/2} \tag{27}$$

where

CA = Correlation amplitude on boresight

B = Samples/s

τ = Integration time

$$\text{SNR} = 0.35\, [3 \times 10^6 \times 10]^{1/2} \cong 1900$$

The VLBI method proved to be inadequate for the holographic measurements. It doesn't provide any verification of test results on site and the probability of recording "bad" data on tapes and hence not utilizing the time spent for the test in the most efficient way is high. The C.E.I. system provided a SNR 7 dB larger than that obtained from the VLBI system. The data collected using the VLBI technique was therefore never actually processed.

IX. Software Development

The software package that was used is composed of the following parts:

(1) Residual phase extractor. Strips the residual phase from the original data set (Figs. 5 and 6) which have nominal reference points imbedded (these reference points are measured when both test and reference antennas observe the incoming signal on boresight).

The nominal reference points are considered to be coordinates of some polynomial function (to obtain Fig. 7, second order degree polynomial was used). In the same manner, the program will optimally normalize the amplitude data set to a polynomial function (Fig. 8).

For both the amplitude and phase functions, the Chebychev polynomial may be used to yield better curve fits at the expense of hand checking the result to insure that the function doesn't drop below zero within the domain of normalized points.

(2) Plotting program. This program generated the plots and maps as shown (Figs. 5 through 11).

(3) Spiral to columnar transformation program. This program takes data arranged as a spiral configuration (this is the way data was collected thus far using the C.E.I. and VLBI experiment as DSS-43-42. This method of collecting the data will be different in the final system capable of producing high resolution maps), and transform it into a two-dimensional complex array (two values are read for each element).

(4) Far-field pattern program. This program generates the far-field amplitude and phase from the measured data.

(5) FFT program. This program computes the inverse fast Fourier transform of the far-field pattern with surface errors. A novel iteration procedure is used to improve the final solution (Ref. 5).

X. Summary

The microwave holography technique has now been successfully used to make surface maps of many of the world's large antennas. The technique allows measurement of the surface with the antenna in an operational configuration. This is contrasted to the traditional optical and/or mechanical methods that often require removal of significant portions of the antenna structure (such as cassegrain feedcone). An optical measurement and panel setting exercise on a DSN 64-m antenna typically requires that the antenna be taken out of service for about 8 weeks. When operational, microwave holography will require on the order of 12 hours of dedicated test time to collect the data for the surface map. Panel setting can then be performed during normal maintenance periods; a significant improvement in the utilization of a valuable resource.

Holography also allows for inspection of the antenna surface on a routine basis, an option that is not readily available using present optical techniques.

Early microwave holography at JPL was conducted using a natural radio source (3C273). This work pointed out the need to use artificial radio sources (geostationary satellites) in order to obtain the high SNRs needed for short integration times. Additionally, mechanical oscillations in the antenna structure at DSS-43 required us to allow 20 seconds to move the antenna from point to point in the data grid so that the antenna again became stable enough to take data. This will be avoided in the future by continuous scanning across strong artificial sources that require short integration times.

The work performed to date has demonstrated the ability of this technique to produce low-resolution surface maps of the large DSN antennas. Much remains to be done before high resolution surface maps become a reality.

Acknowledgements

The authors wish to thank many people who contributed to this work. Many helpful technical discussions were held with Dan Bathker and Larry Young. We would also like to thank Sam Gulkis, Mike Batty, Dave Jauncey and Bill Peters for supplying and operating the Connected Element Interferometer that was used to obtain the data for this experiment. Special thanks go to Yahya Rahmat-Samii for the software development and data reduction that provided many of the figures in this report.

References

1. Ruze, J., "Antenna Tolerance Theory–A Review," *Proc. IEEE 54,* 1966.

2. Vu, T. B., "Influence of Correlation Intervals and Illumination Taper in Antenna Tolerance Theory," *Proc. IEE 116,* London, 1969.

3. Baars, J. W. M., *IEEE Trans Antennas Propagation AP-21*, 1973.

4. Bathker, D. A., *Radio Frequency Performance of a 210-ft Ground Antenna, X-Band.* JPL Technical Report 32-1417, Jet Propulsion Laboratory, Pasadena, CA, Dec. 15, 1969.

5. Rahmat-Samii, Y., *Surface Diagnosis of Large Reflector Antennas Using Microwave Holography Metrology – An Iterative Approach*, 1983-URSI Symposium on Electromagnetic Theory, Santiago de Compostela, Spain, August 23-26, 1983.

6. Bathker, D. A., *Microwave Performance Characterization of Large Space Antennas.* JPL Publication 77-21, Jet Propulsion Laboratory, Pasadena, CA, May 15, 1977, pp. 4-22.

7. Gulkis, S., Batty, M. J., Jauncey, D. L., and Rayner, P. T., "Tidbinbilla Two-Element Interferometer," *The Astronomical Journal,* Volume 87, No. 6, June 1982.

8. Kraus, J. D., *Radio Astronomy*, McGraw-Hill, 1966.

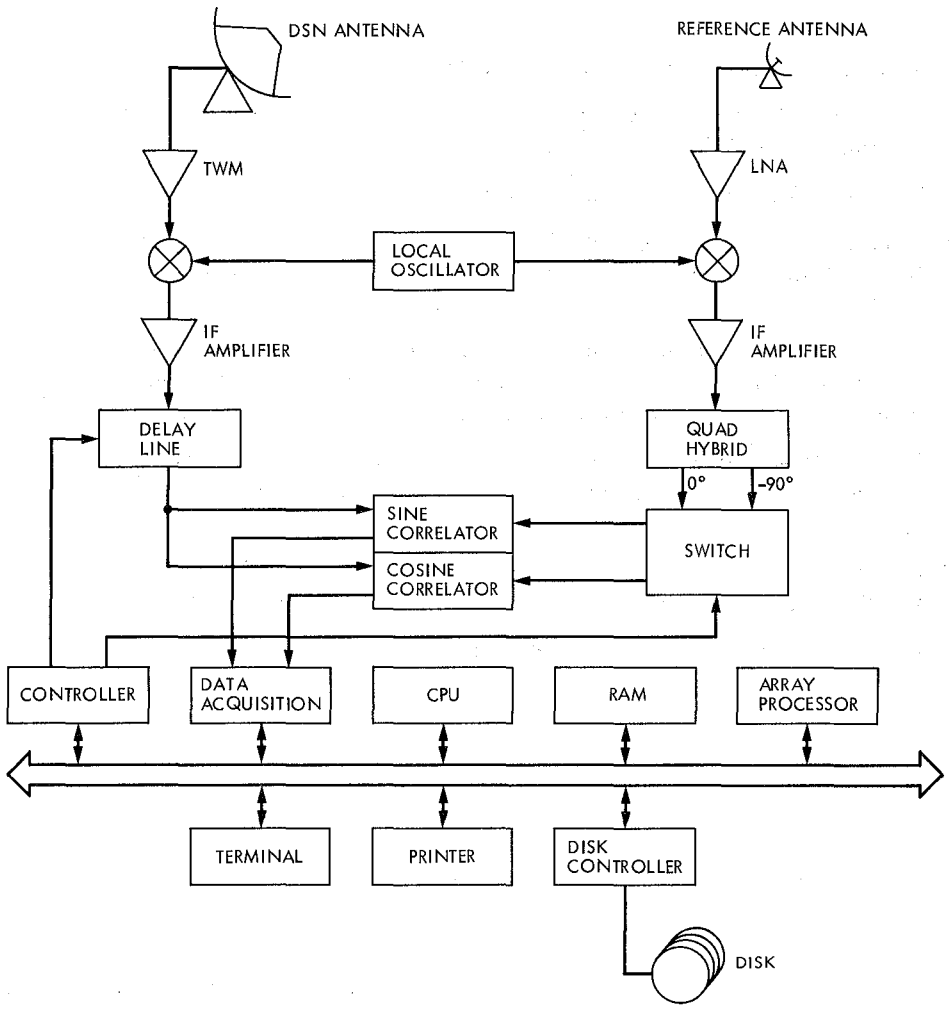


Fig. 1. Holographic mapping receiving system

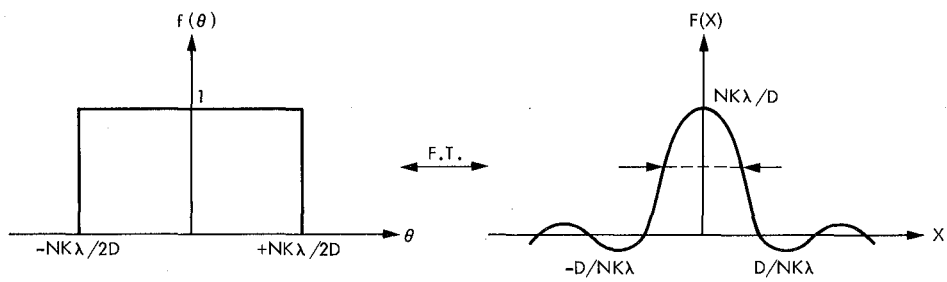


Fig. 2. Fourier transform relationship

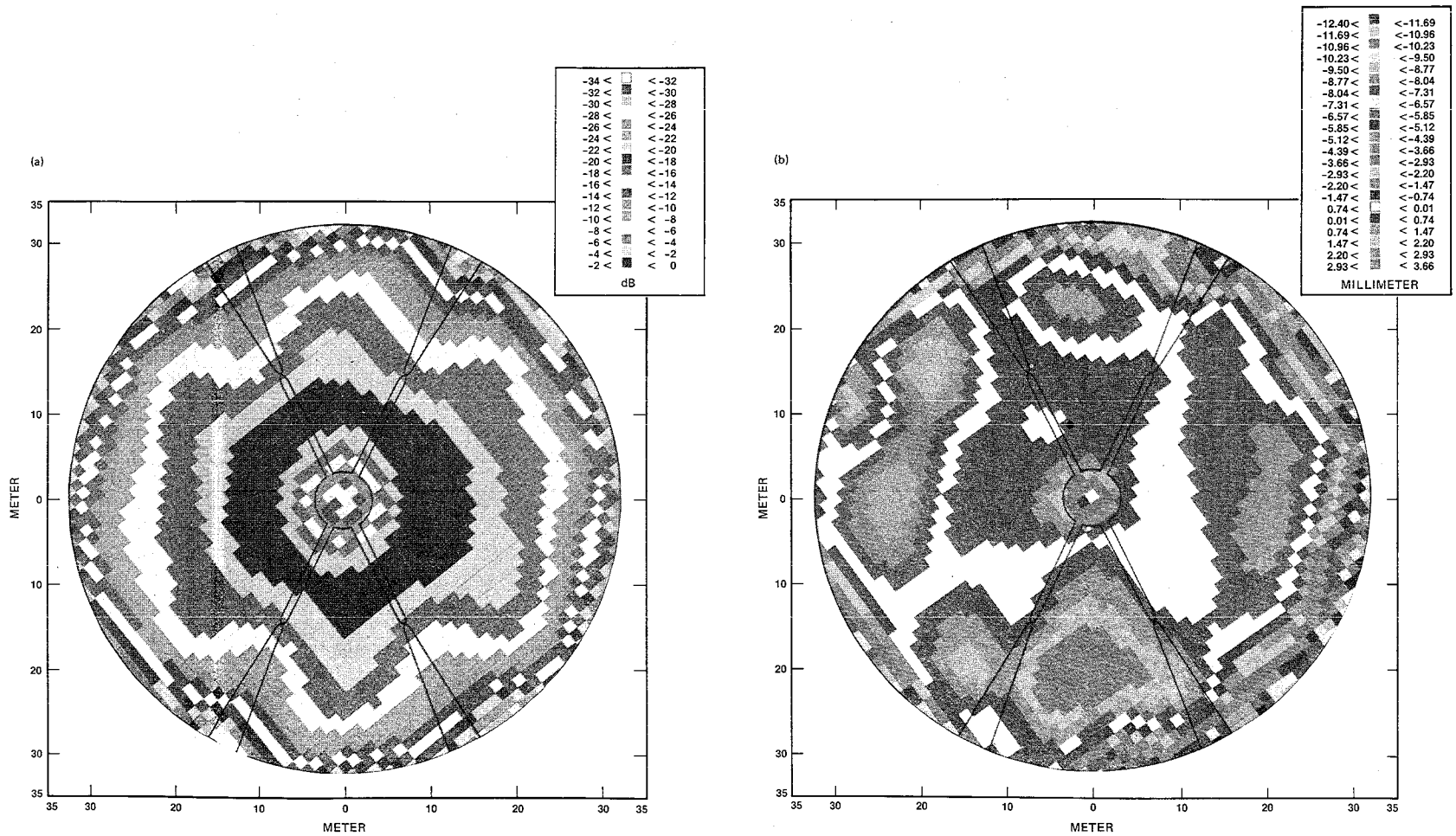


Fig. 3. Gray-level holographic maps of DSS-43: (a) amplitude profile of Tidbinbilla 64-m antenna measured at S-band; (b) surface profile of Tidbinbilla 64-m antenna measured at S-band.

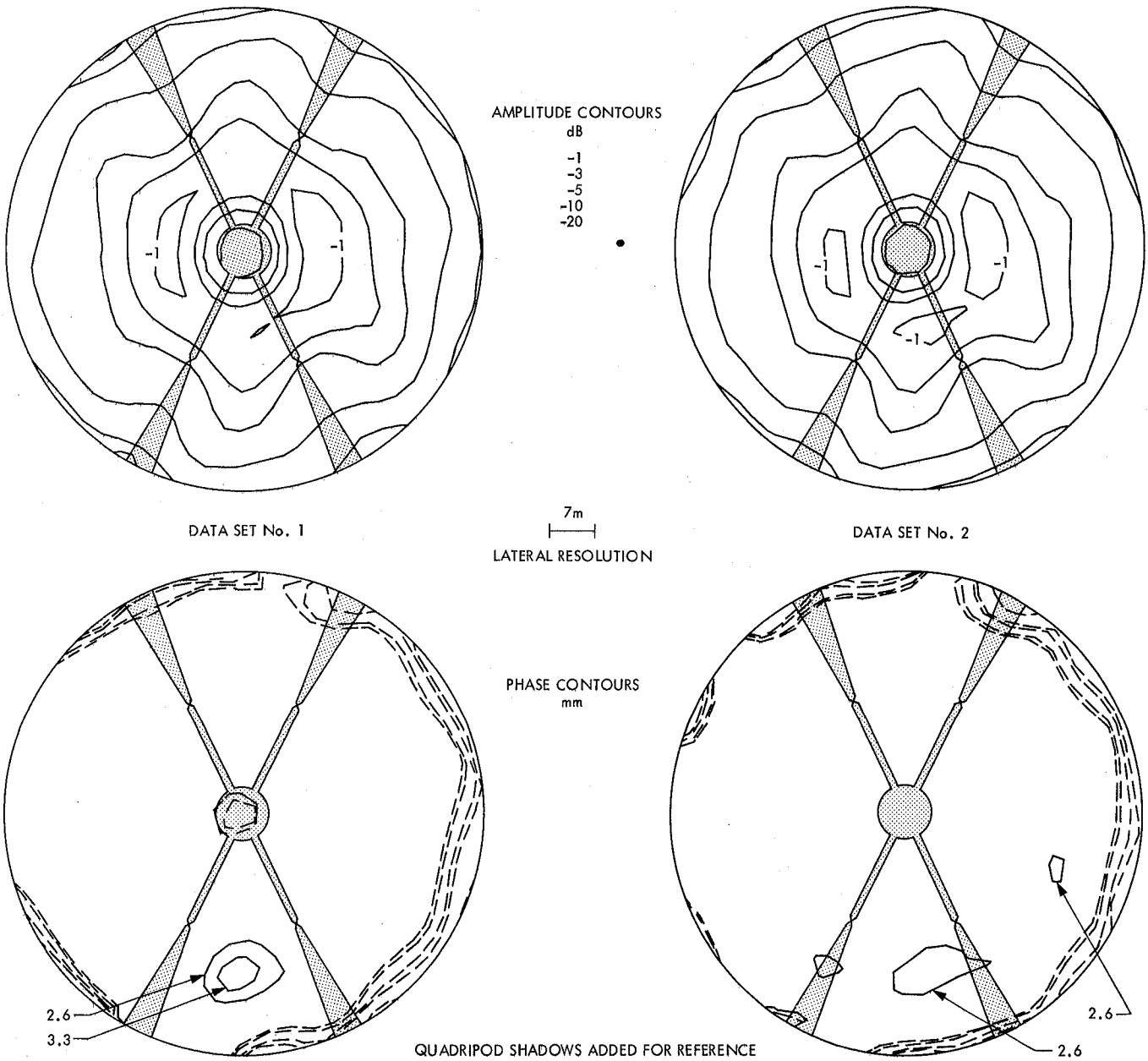


Fig. 4. Holographic surfaces, DSS-43, Feb. 1983, elevation ~ 40°

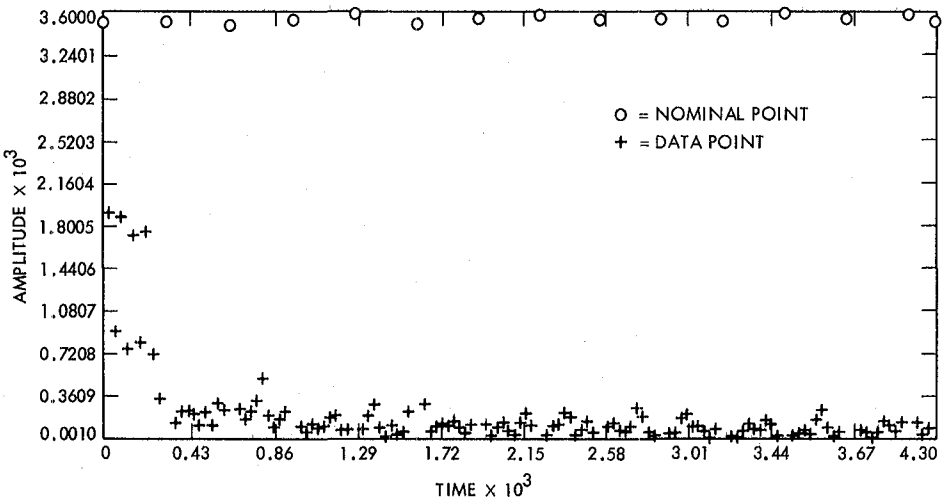


Fig. 5. Measured far-field amplitude data

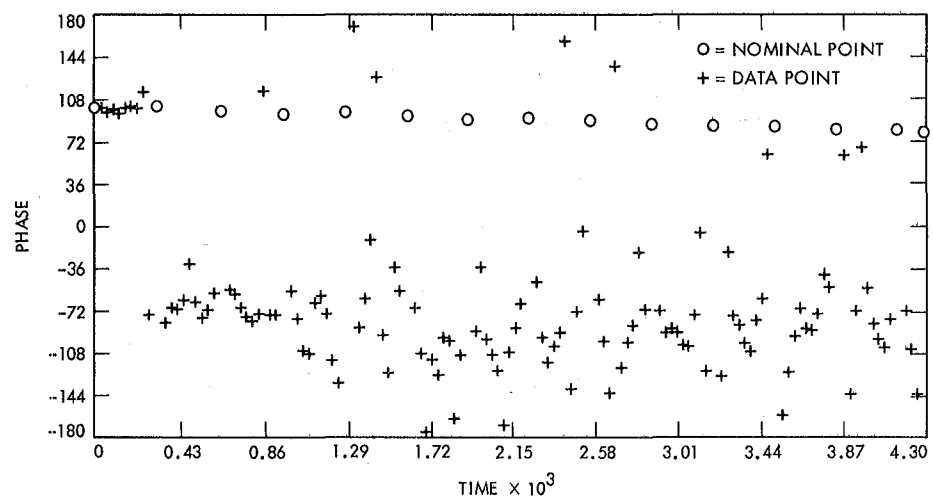


Fig. 6. Measured far-field phase data

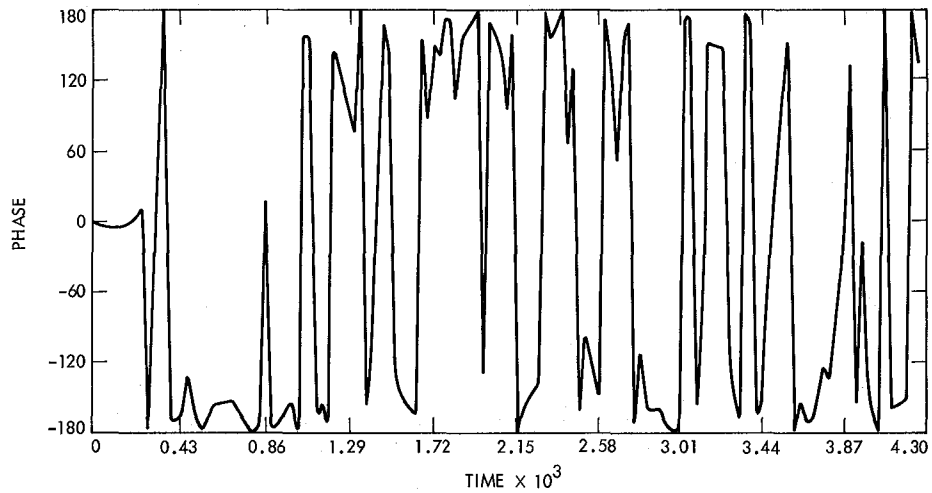


Fig. 7. Adjusted far-field phase data. Residual phase reduced to the range (−180, 180) vs. time.

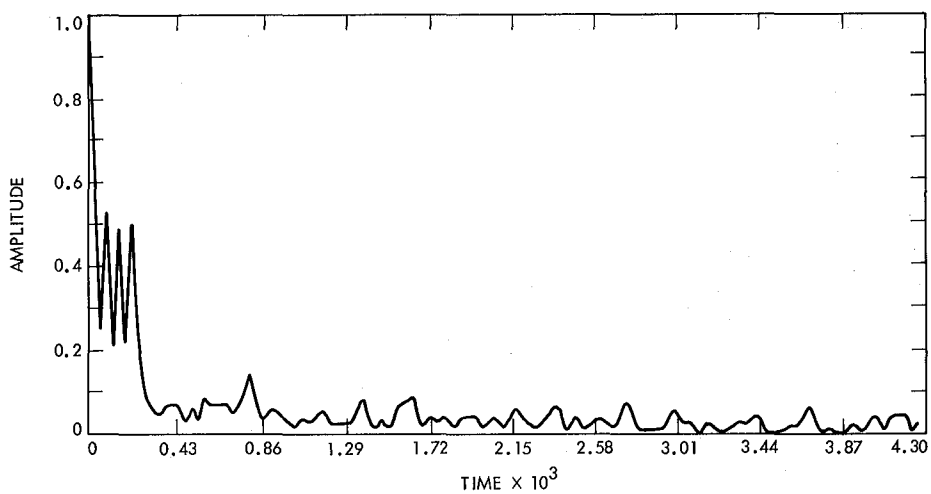


Fig. 8. Adjusted far-field amplitude data. Amplitude normalized to drift curve vs. time.

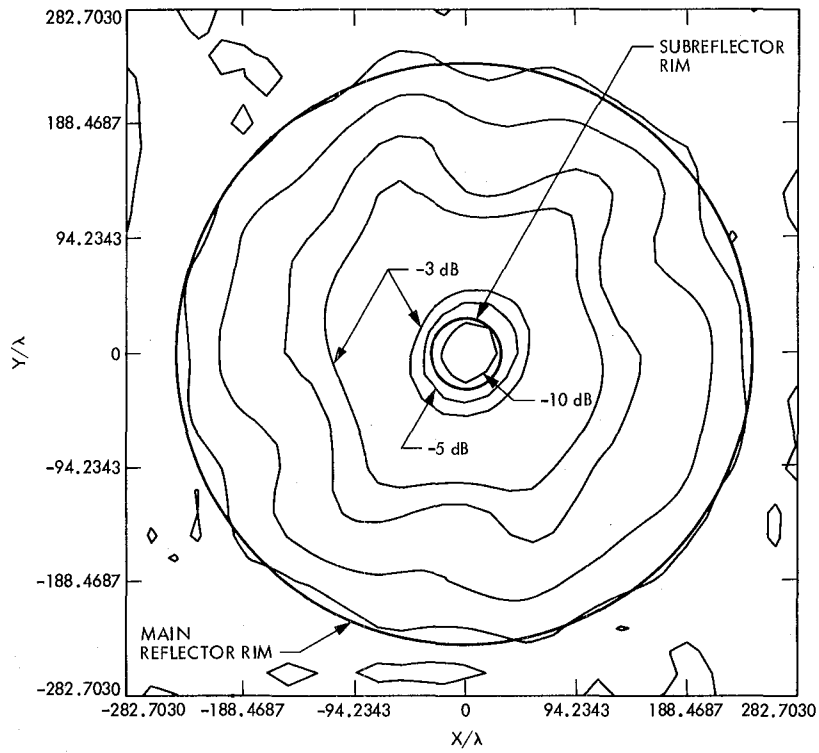


Fig. 9. Contour plot of the reflector aperture amplitude distribution

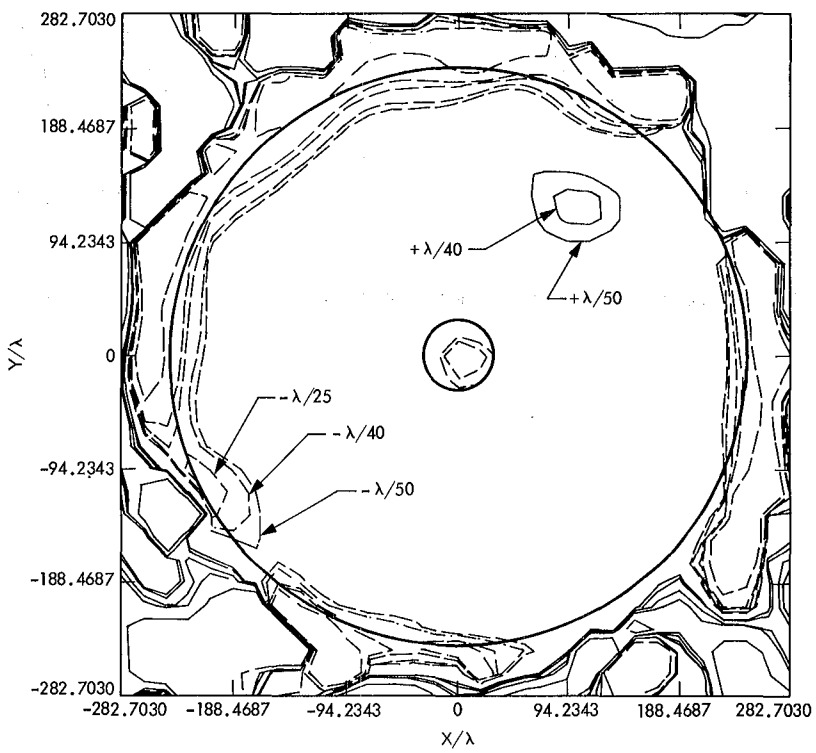


Fig. 10. Contour plot of the surface distortion profile

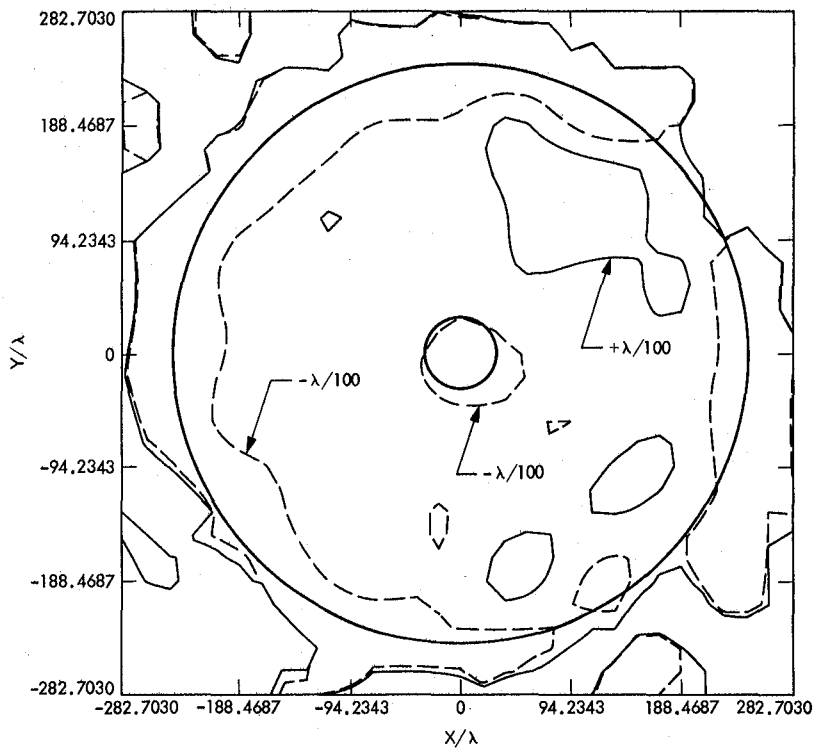


Fig. 11. Contour plot of the surface distortion profile. Contour line is at ±0.01λ.

Loop Gain and Circuit Parameters for Residual Carrier Tracking in the Advanced DSN Block V Receiver

P. H. Young[1]
Radio Frequency and Microwave Subsystems Section

This article presents a circuit level analysis for residual carrier tracking of a deep space network signal by a proposed Block V receiver as conceived by the Receiver Advanced Development Team. The objective is to aid the circuit designer in selection of loop parameters to achieve the required performance objectives.

The topics specifically addressed are (1) loop stability and gain requirements, (2) loop component parametric equations, and (3) phase detector technology.

This paper presents a circuit level analysis for residual carrier tracking of deep space network signal by a proposed Block V receiver, as conceived by an ad hoc design team. The objective is to aid the circuit designer in selection of loop parameters to achieve the required performance objectives.

The topics specifically addressed are as follows:

Loop Stability and Gain Requirements

- Oscillator Stability
- Loop Noise and Signal-to-Noise Ratio
- Doppler and Required Loop Gain
- Need for Coherent AGC and Gain Allowance

Loop Component Parametric Equations

- NCO (Frequency Synthesizer Method)
- Frequency Translation and Multiplication
- Phase Detector
- Analog-to-Digital Converter
- Digital Loop Filter
- Loop Compensation and Filter Frequency Response

Phase Detector Technology

Conclusions and Recommendations

The system analyzed is that of Fig. 1 which is believed to be the latest configuration for Block V.

I. Loop Stability and Gain Requirements

The loop gain requirements for a residual carrier tracking loop can be determined by analyzing the stability of the phase detector input signals. The carrier signal to be controlled is characterized by

$$S(t) = \sqrt{2P}\,[1 + a(t)]\,\sin\,[\omega_o t + \psi(t) + d(t)] \qquad (1)$$

[1]NASA Summer Faculty Fellow, Arizona State University

where $a(t)$ represents amplitude noise. $\psi(t)$ includes phase and frequency noise of the received signal and 2.3 GHz L.O. (see Fig. 1) as well as modulation. In the last term, $d(t)$ represents doppler and other relatively long-term drifts such as circuit aging and temperature effects. The other input to the phase detector—the local reference signal—can be characterized similarly to Eq. (1) and includes noise and drifts of components within the hydrogen maser frequency standard and external multipliers.

A. Oscillator Stability

The problem of oscillator noise has been analyzed by Spilker (Ref. 1) for three sources: frequency flicker noise, white frequency noise, and white phase noise. The phase noise spectral densities were assumed to be K_a/f^3, k_b/f^2, and k_c, respectively, where the proportionality constants are specific to a particular oscillator technology.

For a linear model of the loop, the variance of the phase error is

$$\sigma_\epsilon^2 = \int_0^\infty \phi n(\omega) \mid 1 - H(j\omega) \mid^2 df$$

where $\phi n(\omega)$ is the spectral density of the noise and $1 - H(j\omega)$ gives the loop error function.

Based on the transfer function of a second-order PLL with damping $\zeta = 1/\sqrt{2}$ the residual phase variances

$$\sigma_{\epsilon_i}^2$$

are 8.71 $k_a/b_n{}^2$, 0.179 K_b/B_n and $K_c f_h$, respectively, where B_n is the one-sided loop noise bandwidth and $f_h >> B_n$ is the bandwidth limit imposed by IF filters.[2] Also quantization noise for the Numerically Controlled (digital) Oscillator must be included. The mean square noise for linear quantization is

$$\sigma_q^2 = \frac{q2}{12} \tag{2}$$

where q is the quantum step size. The noise contribution from each mechanism is summed to give a total phase error variance of

$$\sigma_\epsilon^2 = \sigma_a^2 + \sigma_b^2 + \sigma_c^2 + \sigma_q^2 \tag{3}$$

[2]For the DSN receiver this is the post phase-detector low pass filter of 15 MHz bandwidth.

The variance of the phase error is related to frequency variance by

$$\sigma_\epsilon^2 = (2\Pi)^2 T^2 \sigma_f^2 \text{ rad}^2 \tag{4}$$

where T is the measurement time period. The specified frequency stability $\Delta f/f$ will result in a frequency error of

$$\sigma_f^2 = \left[\frac{\Delta f}{f_o}(f_o)\right]^2$$

and, over time T, the variance of the phase error becomes

$$\sigma_\epsilon^2 = (2\Pi \Delta f T)^2 \tag{5}$$

B. Loop Noise and Signal-to-Noise Ratio

Receiver noise within the loop bandwidth B_n will contribute to uncertainty in the loop phase estimate. This can be viewed in terms of a noisy phase detector output voltage or in terms of the frequency uncertainty of the NCO output due to noise modulation.

The receiver noise in the tracking loop can be specified in terms of the carrier-to-noise power ratio, C/N. For a minimum carrier signal defined as that level for which $(C/N)_L$ = 10 dB in the two-sided loop noise bandwidth $2B_{L_o}$, then the worst case C/N at the analog-to-digital converters will be approximately

$$(C/N)_{ADC} = \frac{2B_{L_o}}{2B_{LPF}}(C/N)_L = \frac{0.1 \text{ Hz}}{2(15 \text{ MHz})}(10) \rightarrow -74.8 \text{ dB}$$

where B_{LPF} is the bandwidth of the phase detector output low-pass filter. This value ignores quantizing noise from the proposed 6-bit ADC's (see Eq. [2]) as well as other digital processing noise.

The phase detector input C/N corresponding to the above assumptions and ±250 MHz IF bandwidth would be $(C/N)_{PD}$ = (0.1 Hz) (10)/500 MHz = −87 dB. This IF bandwidth could be limited to, say, 40 MHz since the post detection bandwidth is 15 kHz or less. Then $(C/N)_{PD}$ = −76 dB. In any case, the phase detector will be expected to operate at signal-to-noise ratios substantially below what is conventionally considered to be a threshold value, about −30 dB (Ref. 2). This issue will be explored in a later section on phase detector technology.

C. Doppler and Required Loop Gain

The major determinant of loop gain is based on maintaining a loop phase error of less than 2° (1° goal) during maximum

doppler frequency shifts. Carrier frequency shifts of up to Δf_d = 100 kHz are expected but doppler dynamics are not expected to pose any problems for the tracking loop.

Based on a linearized loop model in which the static phase error θ_e is maintained at less than 0.1 radians (5.7°), the open loop gain (uncompensated loop) at ω = 1 rad/s is given by

$$K_v = \Delta\omega/\theta_e \tag{6}$$

In order to reduce the loop phase error to 1° (0.0175 rad) for a 100 kHz doppler frequency shift, the loop gain must exceed

$$K_v \geqslant 2\Pi\, \Delta f_d/\theta_e(\max)$$

$$= 35.9 \times 10^6 s^{-1}$$

$$= 151 \text{ dB at } \omega = 1 \text{ rad/s}$$

This value, of course, assumes a stand-alone loop with no external intervention.

D. Need for Coherent AGC and Gain Allowance

Equation (15) for phase detector gain (Section II.C) shows that the *PD* gain and therefore the system loop gain and bandwidth are proportional to the residual carrier power. This has a couple of consequences for the system design.

First, a low modulation index is required during the residual carrier tracking mode. For a peak phase deviation of θ_m = 1.57 radians with square-wave modulation, the carrier power is zero (see Eq. [14]). A reasonable upper bound might be $\theta_m \leqslant 1$ rad for which the residual carrier power will be only 5.3 dB less than the total received signal power.

A more important consideration, however, is that while limiters or noncoherent AGC can maintain the noise power into the phase detectors constant for ADC scaling, loop gain and bandwidth will vary with received signal dynamic range. Coherent AGC is required to maintain a constant *carrier* power at the phase detector input.

Since both in-phase and quadrature phase detectors are present in the advanced receiver design it is a simple matter to take the coherent AGC voltage from the inphase (I) detector which acts as a coherent full-wave rectifier (when θ_e = 0). Peak detection of the carrier is then attained by sampling and filtering the I-detector output.

The coherent amplitude detection characteristic of an in-phase phase detector with high input SNR[3] is proportional to

$$\sqrt{2P_c}\, \cos\theta_e \tag{7}$$

Consequently, increasing PLL phase error will cause a decrease in AGC loop gain given by (AGC Voltage Loss),

$$L = 20 \log \cos\theta_e \tag{8}$$

During doppler events the PLL phase error increases. If the dynamics of this process are ignored then we have only to insure that the loop gain is high enough to hold the receiver AGC within specifications.

If the AGC is allowed to vary by 0.5 dB, then the PLL static phase error can increase to $\theta_e = \cos^{-1}(10^{-L/20})$ = 0.336 rad (19.3°). For a worst case doppler of Δf_d = 100 kHz, loop gain (from Eq. [5]) will have to be at least $K_v \geqslant 2\Pi 10^5/0.336 \rightarrow 125$ dB. Of course, the designated loop gain of 151 dB (Section I.C) will insure that AGC variations due to PLL phase error (1° max) will be 19.3 times better than the specification limit.

II. Loop Component Parametric Equations

The simplified residual carrier tracking loop of Fig. 2 delineates loop components and their gain parameters. In this section a parametric expression is determined for each block's contribution to the PLL loop gain. This will allow for the calculation of the overall open loop gain as well as shedding light on the various parameter choices.

A. NCO, Frequency Synthesizer Method

The proposed Numerically Controlled Oscillator is illustrated in Fig. 3. Here, the value of the output frequency (unlocked PLL) or phase (PLL locked) is generated in the accumulator. The accumulator is updated periodically at a rate determined by the stable clock reference frequency f_{ck}. The numerical contents of the accumulator increase in the value for each cycle of the clock. The value of N determines the amount by which the numerical contents increase each clock as illustrated in Fig. 4.

The maximum numerical contents of the accumulator is $M = 2^b$ where b is the number of bits used. Therefore the accumulator will overflow and change output state each time

[3] In order to insure an AGC loop SNR > 10 dB this loop will have to be at least as narrow as B_{L_O} = 0.1 Hz.

M is exceeded. As seen in Fig. 4, the accumulator output frequency will be

$$f_A = \frac{N}{M} f_{ck} = \frac{N}{2^b} f_{ck} \tag{9}$$

which is linearly proportional to the input number N.

The loop gain parameter K_o (and resolution) for the NCO is determined as the frequency deviation Δf_o per change in input number N.

$$\Delta f_o = \frac{N}{2^b} f_{ck} - \frac{N \pm 1}{2^b} f_{ck} = \frac{\pm f_{ck}}{2^b} \tag{10}$$

Hence,

$$K_o \triangleq \frac{\Delta f_o}{N} = \frac{f_{ck}}{2^b} \tag{11}$$

As a typical example, a 32 bit accumulator and 10 MHz clock will give the NCO a sensitivity (and resolution) of $K_o = 10^7/2^{32} = 0.0023$ Hz/numerical value of N.

B. Frequency Translation and Multiplication

The NCO output frequency must be multiplied and translated up to a nominal frequency for the first mixer (2.3 GHz in this case). The translation component of this process will be performed by high-level mixing with an ultrastable reference frequency derived from the maser frequency standard. As such the effect on loop gain and stability will be possibly a multiplication of plus or minus one.

The times-M ($\times M$) multiplication component of this block is needed to increase the deviation sensitivity of the NCO thereby increasing the system open loop gain. The multiplication factor is the gain parameter for this circuit

$$K_m = M \tag{12}$$

M is usually determined after the other loop gain parameters are known since it is typically the most flexible and technologically achievable parameter. The major constraint on M is frequency stability.

As pointed out by Tam and Armstrong in a preliminary report entitled "Description of Block V Receiver Signal Processing Assembly with Technology Assessment" (personal communication), a commercial frequency multiplier with good thermal regulation has demonstrated "a stability of 3×10^{-16}, $\tau = 1000$ s."

C. Phase Detector

The phase detector seems at this point to be the weak link if not the Achilles' heel of the tracking loop. A search of the literature and data sheets (with telephone follow-up) indicates that there is not an off-the-shelf unit available which meets the specifications of 800 MHz ±40 MHz input with an output voltage range of ±2.5 V and 35-40 MHz bandwidth, not to mention output offset voltages of less than "a few tens of microvolts."

I have no doubt that a custom unit is quite feasible. It will be a single-balanced mixer design with an output balance potentiometer and approximately 1 KΩ output impedance. This issue is discussed in some detail later (Section III).

The analog phase detector of block diagram Fig. 2 has as inputs an 800 MHz IF signal including the phase estimate θ from the NCO-XM and an equal frequency reference signal f_{ref} derived from the maser frequency standard. The transfer function of the phase detector is

$$V_d = \sqrt{2P_c} \sin \theta_e \tag{13}$$

where P_c is the residual carrier power and θ_e is the loop phase error. The residual carrier power as a function of modulation index θ_m for a unit power squarewave modulation subcarrier is

$$P_c = P \cos^2 \theta_m \tag{14}$$

where P is the total power.

Since the loop will track the carrier with low phase error, the phase detector gain is just the slope of Eq. (13) at the origin. That is,

$$K_\Phi \triangleq \frac{dV_d}{d\theta_e} = \sqrt{2P_c} \text{ V/rad} \tag{15}$$

A typical value of phase detector sensitivity for high-level inputs is 0.5 rad/V. Unfortunately, most of the input in our application is noise so K_Φ will be low.

D. Analog-to-Digital Converter (ADC)

The ADC samples the phase detector output $V_d(t)$ and quantizes it into a numerical estimate $V_d(n)$. $V_d(n)$ is more commonly written as $X(n)$, X_n or X_k.

The output of the phase detector is essentially all noise with SNR = -75 dB (Section I.C) at the operating threshold defined as +10 dB SNR in $2B_{L_o}$ = 0.1 Hz. The $3\sigma_n$ noise peaks are scaled to ±2.5 V for input to the ADC and held constant by bandpass limiting or noncoherent AGC. If the number of bits used for linear quantizing in the ADC is a, then the ADC gain constant is

$$K_{\text{ADC}} \triangleq \frac{\Delta n}{\Delta V_d} = \frac{2^a}{6\sigma_n} \tag{16}$$

which for a = 6 bits and a full sampler input range of 5V pk-pk will be $2^6/5$ V = 12.8 numerical values per volt.

E. Digital Processing and Loop Filter

The phase detector output samples are taken at a very high rate (about 30M samples/s) compared to the expected tracking dynamics (kHz range). These noisy samples are accumulated (perhaps an Accumulate and Dump operation) and averaged to form a reduced rate sequence of numeric values which approximate the loop phase/frequency error. The static gain of this digital filter can be scaled to unity if desired so that it would have no effect on the loop gain computation.

Figure 5 is a block diagram of the second order loop filter implemented digitally. G_1 and G_2 are filter constants and the z-transform parameter z^{-1} indicates a 1-bit delay. For a third order loop an additional block is added (dashed) implementing $G_3/(1 - z^{-1})$.

For the purposes of computing the system parameter K_v ("velocity constant"), the static gain of the loop filter is G_1 and for the overall digital processing gain is K_F.

The loop filter output for updating the NCO phase/frequency is $z(n)$. As indicated in the system block diagram of Fig. 2, a numerical value N_o is set which establishes the center frequency of the NCO. This numerical value is incremented by the digitally processed numerical estimate of the phase and frequency error from the summing junction, $y(n)$. $y(n)$ is the sum of the Residual Carrier Error (RCE) and the Suppressed Carrier Error (SCE) where either may be zero depending on the data transmission mode.

More work needs to be done in the digital processing area to better define the parameters and their effects on loop gain and dynamics.

F. Loop Compensation and Loop-Filter Frequency Response

Tracking loops for receivers are usually compensated for damping factor ζ between 0.5 and 1.5 where $\zeta = 1/\sqrt{2}$ is most common. Lead-lag compensation allows for the independent setting of loop gain, bandwidth and damping.

For the advanced receiver design the digital loop filter will provide the loop compensation. Figure 6 illustrates the open loop frequency response where the simple pole at ω_p (= $1/\tau_1$) and zero at ω_z (= $1/\tau_2$) are set by the frequency response of the digital loop filter. For $\zeta = 1/\sqrt{2}$ and a desired one-sided loop bandwidth of B rads/s, the zero should be approximately one octave below B. That is,

$$\omega_z = B/2 \tag{17}$$

If the 40 dB/decade line ($1/s^2$) is projected from ω_z to the K_v = 1 axis, the intercept is at $B/\sqrt{2}$. Hence, the $\zeta = 1/\sqrt{2}$ compensated open-loop response between ω_p and ω_z is on a slope given by

$$40 \log \frac{B}{\sqrt{2}\,\omega} \tag{18}$$

The pole ω_p must be placed at that frequency for which the $1/s^2$ and K_v/s curves intersect. In log notation for Bode plotting, the K_v/s curve is given by

$$20 \log \frac{K_v}{\omega} \tag{19}$$

The solution for ω_p therefore is $B^2/2\omega_p^2 = K_v/\omega_p$, or

$$\omega_p = \frac{B^2}{2K_v} \tag{20}$$

The second order loop filter frequency response is equivalent to a lead-lag network with a pole and zero given approximately by Eqs. (17) and (20) respectively.

For the tracking loop with $\zeta = 1/\sqrt{2}$, B_{L_o} = 0.1 Hz, and K_v(dB) = 151 dB ($K_v = 35.9 \times 10^6$), the results are $\omega_p = 1.38 \times 10^{-9}$ rad/s (0.22 nHz) and ω_z = 0.314 rad/s (0.05 Hz). For ζ = 1.5, ω_z is decreased by a factor of 1.65 and ω_p is decreased by 1.28.

The exact result relating damping factor, loop gain and compensation time-constants τ_1 and τ_2 is from Ref. 3, $\zeta = (1 + \tau_2/r\tau_1)\sqrt{r}/2 \approx \sqrt{r}/2$ where $r = AK\tau_2^2/\tau_1$. The approximation $\zeta \approx \sqrt{r}/2$ can be rewritten in my notation as $\zeta \approx \sqrt{K_v\,\omega_p}/2\omega_z$ which for $\zeta = 1/\sqrt{2}$ and $\omega_z = B/2$ yields Eq. (20). Failure to use the exact relationship of Ref. 3 for the high gain tracking loop will result in an error for the damping

factor of less than 9 parts per billion. The correction factors applied to ω_p and ω_z for $\zeta_x \neq 1/\sqrt{2}$ are computed from

$$x = (2\zeta_x^2)^{1/3} \tag{21}$$

where

$$\omega_p(\zeta_x) = \sqrt{x}\ \omega_p \text{ and } \omega_z(\zeta_x) = x\omega_z \tag{22}$$

III. Phase Detector Technology

The concerns of phase detector DC offset are explored in this section.

Commercially available phase detectors produced by mixer manufacturers come in two circuit configurations, i.e., the double-balanced mixer (DBM) of Fig. 7(a) and the single-balanced mixer (SBM) of Fig. 7(b). In addition, for completeness, Fig. 7(c) illustrates a second form of the SBM.

As analyzed in Ref. 4, circuit C can be dismissed immediately because of poor dc balance, low gain (about 9.9 dB less than circuit B), and limited output range because the transfer characteristic has flat tops for $V_d > V_D$ due to the discharge of the output capacitor C. Circuit B is a more conventional version of the SBM phase detector for which Eqs. (12) and (14) give the transfer function and gain assuming that the reference input $V_{ref} \gg \sqrt{2P_c}$. The analysis of Ref. 4 indicates best linearity and dc balance when the reference and input signals are equal. However, the circuit sensitivity will suffer by $\sqrt{2}$ as the transfer function becomes

$$V_d = \sqrt{P_c}\left[\sqrt{1+\sin\theta_e} - \sqrt{1-\sin\theta_e}\right] \tag{23}$$

which is a nearly triangular version of the sinusoidal characteristic of Eq. (12). Also, noise was not considered in the analysis although a reference oscillator or the VCO of a narrow bandwidth PLL will be expected to be relatively noiseless compared to the signal input. Consequently one would not expect a thresholding effect as exhibited in discriminators. The nearly triangular characteristic when operating with equal level inputs is advantageous for demodulator applications *and* where the nonlinearity of the sin θ_e detector becomes a problem in high-level input applications. The block V receiver qualifies on both counts.

The Varil PD 100 and PD 101 DBMs are redesigned version of the standard DBM. The redesign takes advantage of the sensible notion that developing a reasonably high output voltage requires using more than a 50 Ω load resistance. The PD 100-101 is designed for an output load of 500 Ω. As might be expected the dc offset is worse with the higher impedance. Indeed the equation from Ref. 5 (p. 83) must be modified to include $R_L \neq 50\ \Omega$. Also this equation, derived in Ref. 5, is already in error by $\sqrt{2}$ because in Eq. (19) (p. 83), $V_{IF} = \sqrt{50\ P_{IF}}$ gives the rms voltage whereas the $2V_{IF}/\Pi$ relationship for the full-wave rectifying due to poor isolation (IS) requires the *peak* V_{IF}. The modified and corrected expression is

$$V_{\text{offset}} = 6.4\ \sqrt{R_L/50}\ 10^{\alpha}$$

or

$$= 0.9\ \sqrt{R_L}\ 10^{\alpha} \tag{24}$$

where

$$\alpha = (\text{LO} - \text{IS} - 30)/20$$

and LO power is in dBm with isolation in dB.

As stated in Ref. 5, DBM's often have better isolation than SBM's and they always are superior with respect to third-order (and higher) intermodulation products which produce dc offset. However, SBM's are simpler; it's easier to match diodes and to maintain the match[4]. Also the dc offset voltage due to noise can easily be zeroed. Such a circuit is illustrated in Fig. 8. One such unit (PD 100-101) is guaranteed only to 500 MHz with a 1.5 MHz baseband width having ±2 V output range into 1 kΩ (+17 dBm RF input). While these specs don't match our application, it should be noted that they are guaranteed over a temperature range of -54°C to 100°C, and custom designs with reduced temperature requirements may achieve our specifications.

The double-balanced mixer approach is available from another manufacturer, Mini-Circuits. This unit is designed specifically for phase detector applications and isolation is greater than 50 dB. The dc offset is 200 μV typical (1 mV, max). The available unit is rated only to 100 MHz with a ±1 V output range into 500 Ω (+7 dBm input) but the baseband width is 50 MHz. DC balance zeroing is not practical with four-diode DBM's.

The SBM and DBM units available as phase detectors designed for relatively high impedance loads have absolute maximum input power ratings of +20 and +17 dBm, respec-

[4]This is especially important in high-level detectors because they may have more than one diode/branch (leg) or a diode and series resistor in order to reduce dissipation or breakdown effects.

tively. Indeed the high level mixers available from Avantek and Watkins-Johnson (Relcom Division) are rated to +27 dBm of LO input power. None of these devices can be driven with one Watt (+30 dBm).

IV. Conclusions and Recommendations

An analysis of the requirements for residual carrier tracking for the advanced DSN receiver indicates a minimum loop gain of 151 dB be used. This is based on maintaining a maximum loop phase error of 1 deg without external intervention.

The contribution to system loop gain for each loop component has been assessed and a parametric equation determined. This will aid the receiver designer in making the tradeoffs necessary to achieve the required loop gain.

Coherent AGC will be required in order to maintain loop gain and bandwidth over a large received signal dynamic range.

Concerns over phase detector thresholding effects due to large negative SNR's and dc voltage offset have been addressed. The findings suggest a custom design incorporating the single-balanced mixer configuration of Fig. 8. The suggested chopped reference solution can be used if necessary to further reduce the dc offsets. Indeed, since this is a technique similar to the chopped amplifier solution to dc stability, it suggests the use of an ac-coupled amplifier at the chopped phase detector output to obviate the question of obtaining ±2.5 V peak on noise.

As it is, attaining the required loop gain appears to pose a substantial challenge as the following calculation will attest:

$$K_v = K_\Phi K_{ADC} K_F K_o M \tag{25}$$

If the representative loop component gain values calculated in the text are used then K_v = 151 dB = 35.9×10^6 and the multiplication factor in the frequency translation section becomes $M \geqslant 35.9 \times 10^6/(\sqrt{2P_c} \times 12.8 \times K_F \times 0.0023 \times 2\Pi) = 137 \times 10^6/K_F \sqrt{P_c}$. Because of the negative SNR's, the carrier signal into the phase detector will be $\sqrt{2P_c} \ll 2.5\ V_{pk}$ and therefore the digital processing will be required to provide a considerable processing gain.

On the other hand the 151 dB of loop gain is based on requiring the loop to reduce, *without external intervention*, the loop phase error to 1 deg during a 100 kHz doppler frequency shift. External methods for reducing loop stress will be very important.

More work is needed to be done to better define the digital processing area of the tracking loop for better definition of the parameters which affect the loop gain and dynamics and some to overcome the lack of published research on microwave phase detectors.

References

1. Spilker, J. J., *Digital Communications by Satellite*, Chapter 12, Prentice Hall, 1977.
2. Gardner, F. M., *Phase Lock Techniques*, Wiley, 1966.
3. Tausworthe, R., "Theory and Practical Design of Phase-Locked Receivers," *JPL Technical Report No. 32-819*, Feb. 1966, p. 31.
4. Ohm, G. and Alberty, M. "Microwave Phase Detectors for PSK Demodulators," *IEEE Transactions on MTT*, Vol. MTT-29, No. 7, July 1981, p. 724-731.
5. Kurtz, S. R., "Specifying Mixers as Phase Detectors," *Microwaves*, Vol. 17, No. 1, pp.80-89, Jan. 1978.

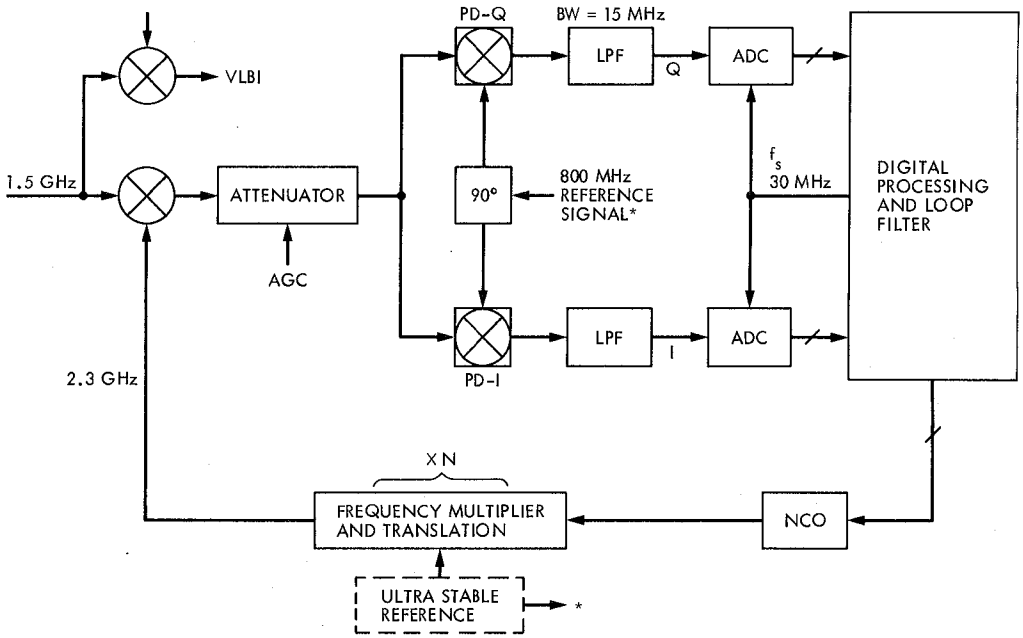


Fig. 1. Basic block diagram for the carrier tracking loop

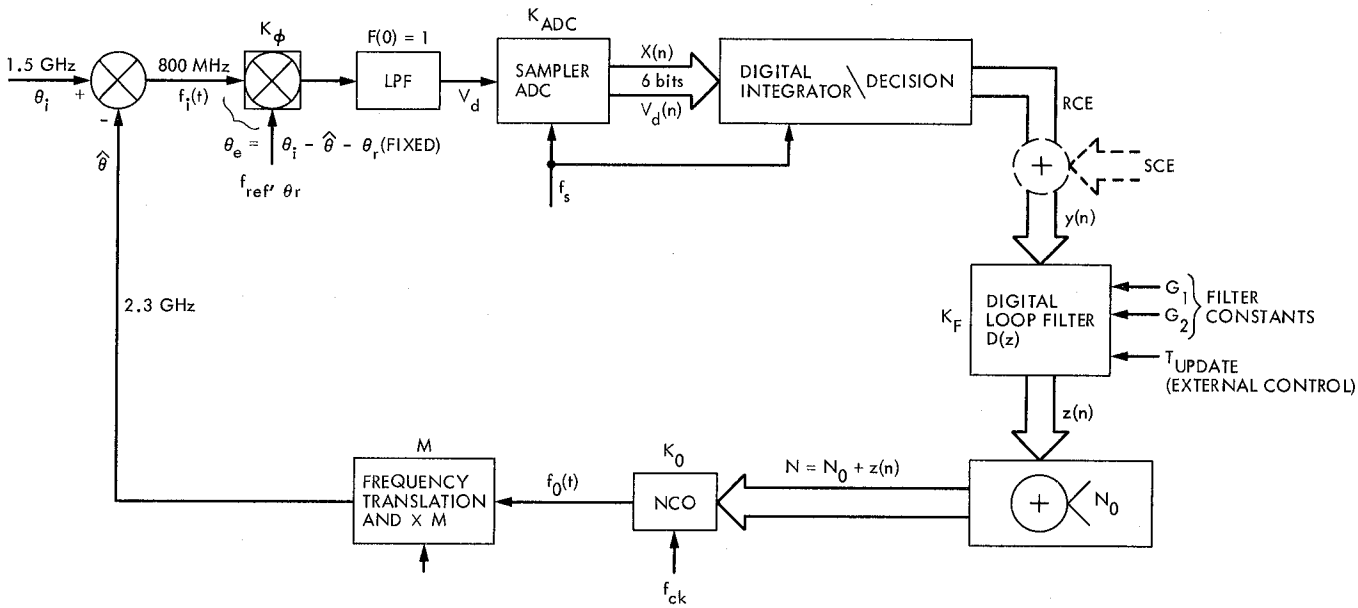


Fig. 2. Simplified residual carrier tracking loop

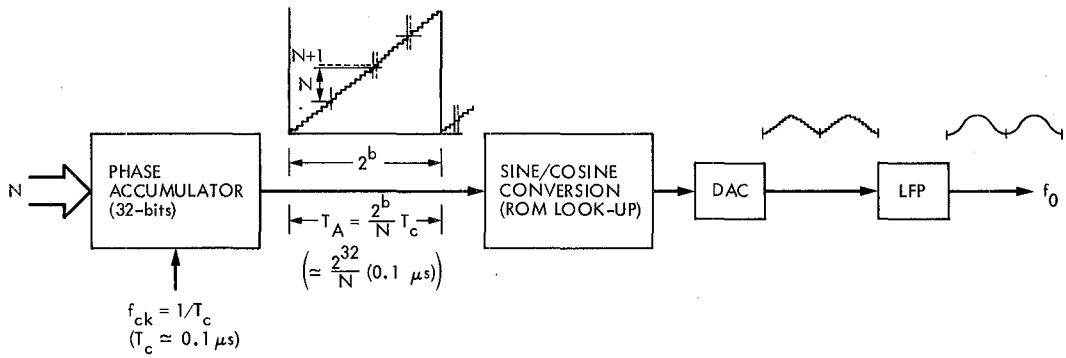


Fig. 3. Numerically controlled oscillator (frequency synthesizer)

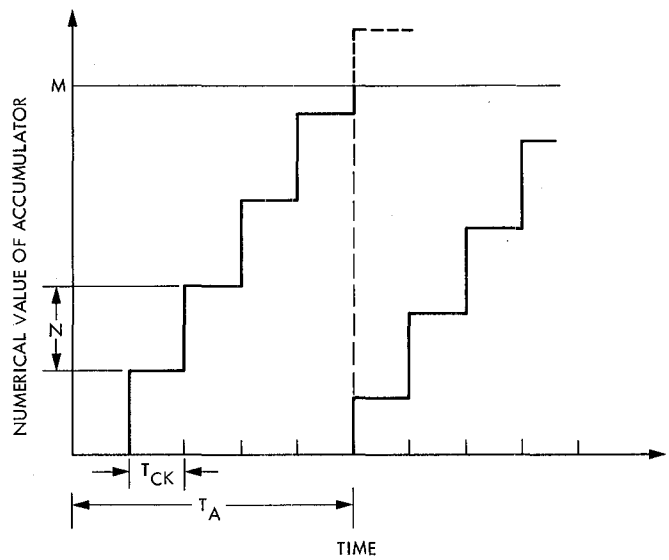


Fig. 4. Numerical value of accumulator as a function of time

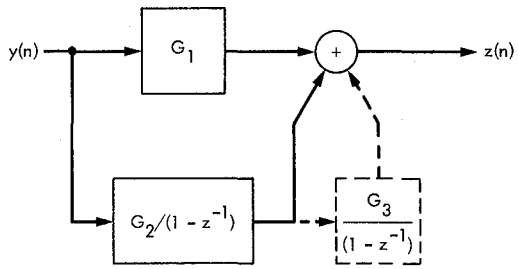


Fig. 5. Digital loop filter

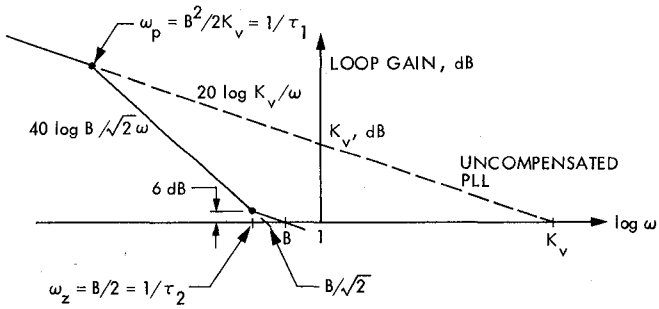


Fig. 6. Open loop gain versus frequency

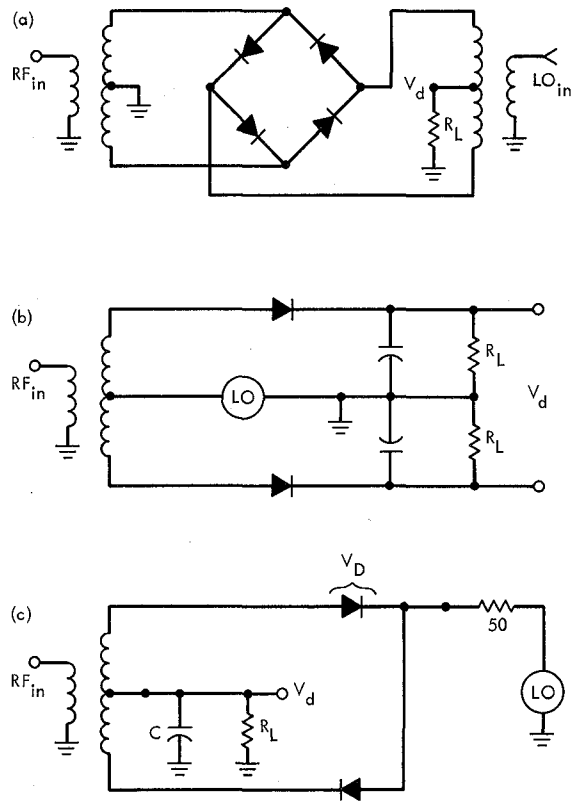


Fig. 7. Typical phase detector circuits

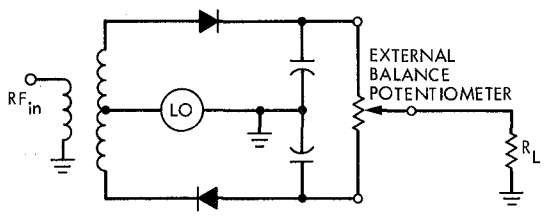


Fig. 8. Improved single-balanced phase detector

X-Band Uplink Ground Systems Development

C. E. Johns
Radio Frequency and Microwave Subsystems Section

The development of the X-band exciter and Doppler extractor equipment for the X-band uplink has been completed. Stability measurements have been made on the exciter and Doppler reference signals and the results are presented.

I. Introduction

The development of a stable X-band RF subsystem was prompted by requirements for increased accuracies of spacecraft navigation and charged particle calibration. The design goal was to develop an X-band uplink with a fractional frequency ($\Delta f/f$) stability of 5×10^{-15} over a 1000-s integration period.

To accomplish this goal, a stable X-band exciter and Doppler extractor system was developed (Ref. 1) that would provide the required stability.

The fabrication of the prototype exciter and Doppler extractor has been completed and tests were performed at JPL to verify the stabilities of the exciter and the S- and X-band Doppler reference signals. The results of the tests are presented.

II. Exciter Functions

A. Exciter Stability

Exciter stability measurements were made at JPL using the Maser Test Facility. A simplified diagram of the test set-up is shown in Fig. 1. Briefly, the exciter signal was set to 7.2 GHz + 1 Hz and phase compared with a stable 7.2-GHz signal generated from a hydrogen maser. The 1-Hz difference signal, at the phase comparator output, was fed to the Allan variance computer.

In addition, maser-derived 10-MHz and 100-MHz references were supplied to the exciter synthesizer and the exciter, respectively.

The result of the exciter stability measurement is shown in Fig 2. The stability at the 1000-s integration period is seen to be about 2.5×10^{-15}, which agrees closely with prior computations (Ref. 1 and Fig. 3).

B. Doppler Reference Signals

The exciter also generates coherent X- and S-band Doppler reference signals. The frequency of these signals, relative to the exciter frequency, fx, are:

$$\text{S-band Doppler} = 240/749\, fx$$

$$\text{X-band Doppler} = 880/749\, fx$$

To generate the Doppler references, a $fx/80$ signal (Ref. 1 and Fig. 4) is multiplied to $131/749\, fx$ and then summed with fx to yield the X-band Doppler. In a second path, the $fx/80$ signal is multiplied to $509/749\, fx$ and subtracted from the fx signal to yield the $240/749\, fx$ Doppler reference.

The stability of both Doppler references were also measured and the results are shown in Figs. 3 and 4.

III. Doppler Extractor

Included with X-band exciter hardware for DSS 13 is a new Doppler extractor to permit Doppler extraction using the receiver's second local oscillator (LO). The extraction scheme is shown in Fig. 5. As shown, the Doppler first IF is generated by mixing the stable 8.1-GHz first LO frequency with the Doppler reference signal generated within the exciter. The Doppler first IF does not contain Doppler information, but after it is mixed with the receiver's second LO, the resulting 1050-MHz IF signal contains Doppler information. This signal is subsequently down-converted to 50 MHz and then again down-converted to 1 MHz and 5 MHz biased Doppler signals.

This method of Doppler extraction reduces receiver phase instabilities by a large factor and was considered a prerequisite for achieving the 5×10^{-15} stability goal.

IV. Receiver Configuration

To utilize the new Doppler extractors at DSS 13, the present Block II receivers will be modified as shown in Fig. 6. The X- and S-band receivers are identical in configuration and only one is shown for simplicity.

Functionally, the X-band receiver is similar to the X-band Doppler extractor to the extent that from the X-band inputs to the 50-MHz outputs, the IF signals in each are identical. In the same respect, the S-band receiver is similar to the S-band Doppler extractor.

The receiver 50-MHz IF is fed into the standard Block III receiver input. The Block III LO was modified to allow locking to the 350-MHz IF.

V. Conclusions

The initial tests made on the X-band exciter and Doppler reference signals have verified the accuracy of the original stability estimate of 2.5×10^{-15} for 1000-second integration periods.

Receiver and Doppler hardware instabilities will degrade the overall system performance. Estimates have been made that a total system stability of 5×10^{-15} will be attained, as measured at the biased Doppler output. Future testing will be performed at DSS 13 to evaluate the complete system and verify the estimates.

All RF hardware has been delivered to DSS 13 and is awaiting final interface cabling to complete the X-band uplink installation.

References

1. Hartop, R., Johns, C., and Kolbly, R., "X-band Uplink Ground Systems Development," *The Deep Space Progress Report 42-56, January and February 1980*, pp. 48-58. Jet Propulsion Laboratory, Pasadena, California.

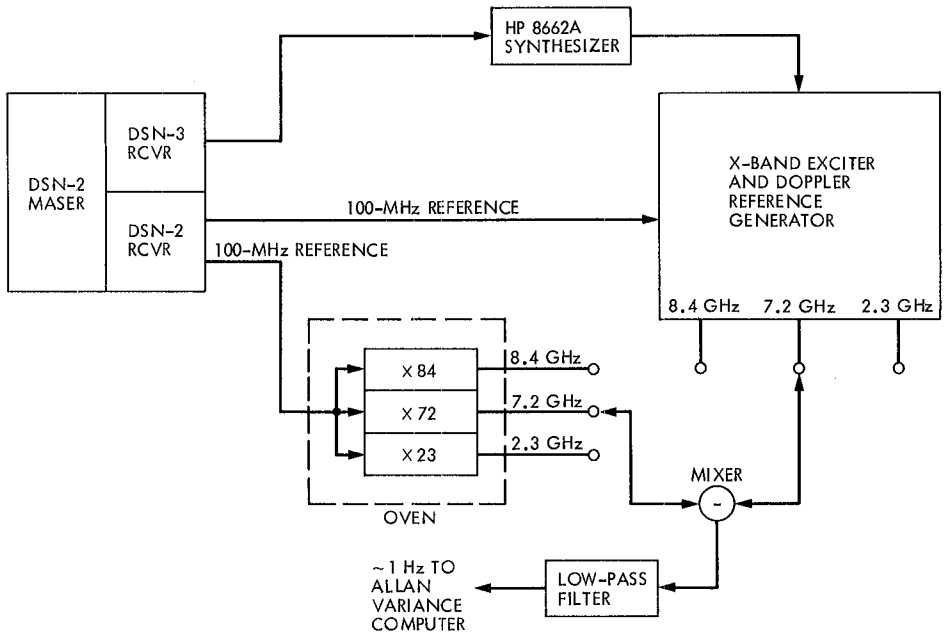


Fig. 1. Simplified diagram of stability test setup

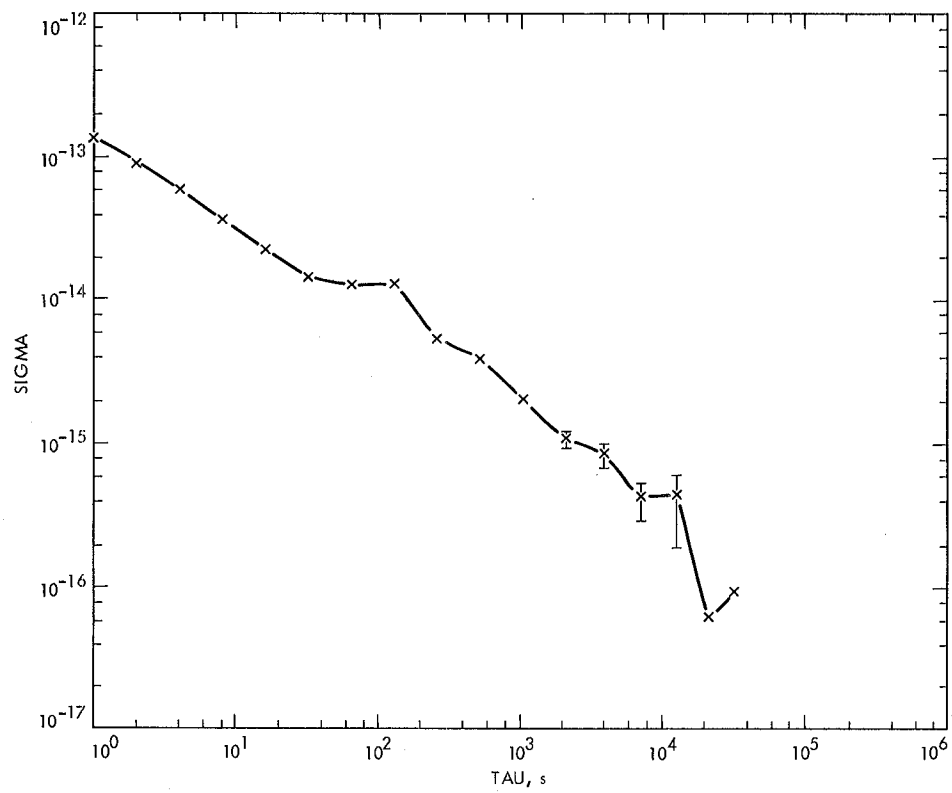


Fig. 2. 7.2-GHz exciter output stability

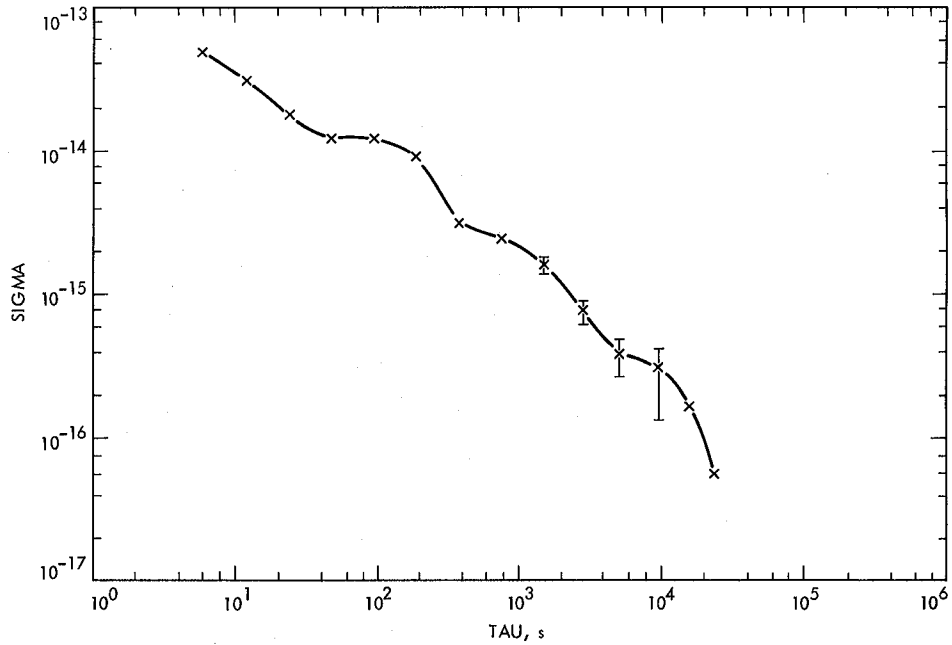


Fig. 3. 8.4-GHz Doppler reference stability

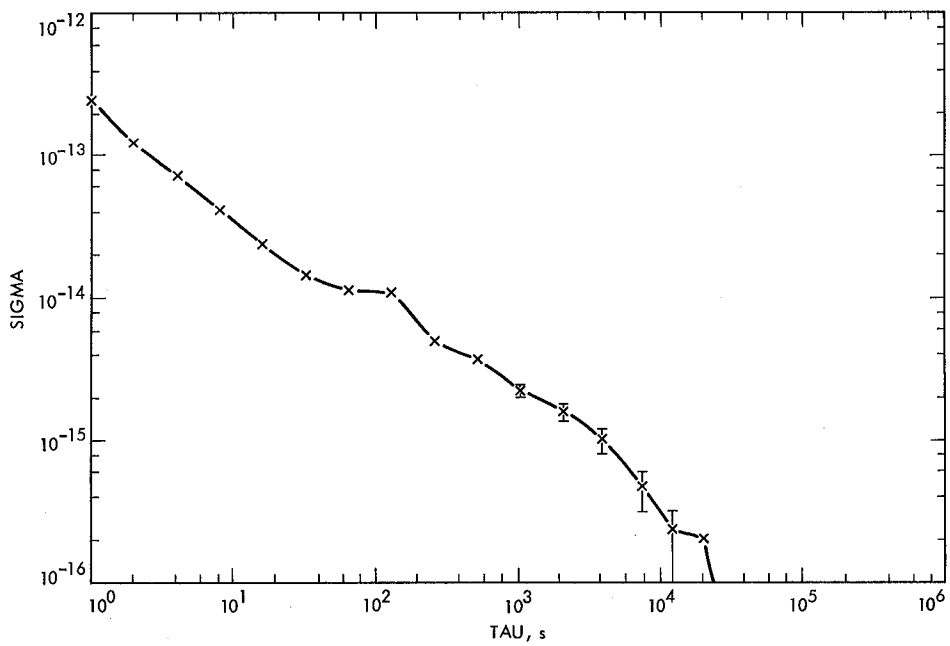


Fig. 4. 2.3-GHz Doppler reference stability

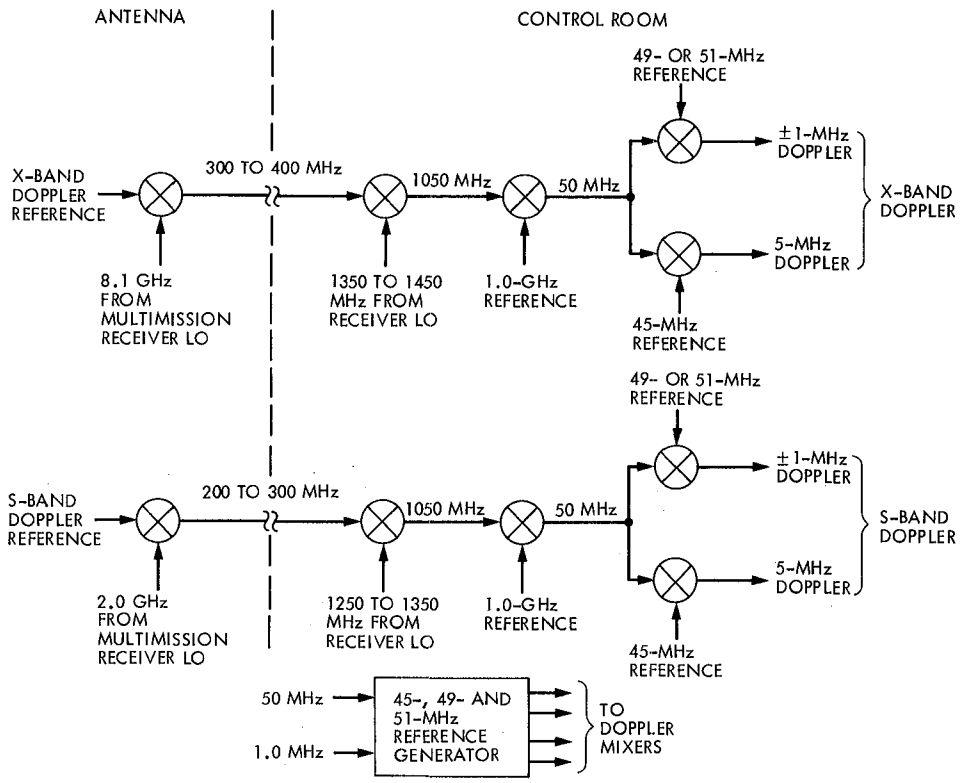


Fig. 5. X- and S-band Doppler extractor scheme

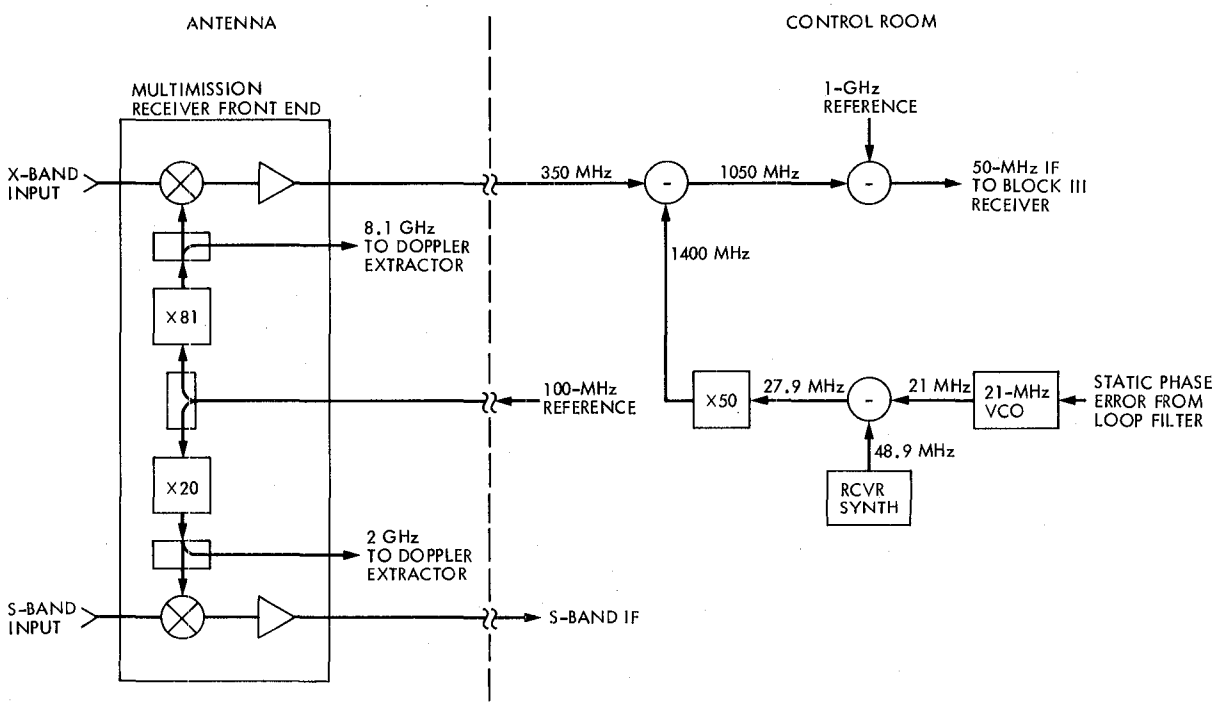


Fig. 6. Simplified diagram of the X-band receiver

Doppler System Phase Transfer Functions for a System With an X-Band Uplink and X-Band and S-Band Downlinks

M. A. Koerner
Telecommunications Systems Section

A new DSN RF system is being developed that can transmit at X-band and receive at both X-band and S-band. End-to-end tests are planned to measure the phase stability of this system. This article derives equations for the phase transfer functions between the error sources in the system and the X-band and S-band doppler extractor outputs. This analysis considers both test modes, using a test translator, and actual spacecraft tracking where the spacecraft is at some distance from the DSN station. The results indicate that the proposed end-to-end tests will not accurately reflect the error that will occur during actual spacecraft tracking.

I. Introduction

The next step in the evolution of radio links for planetary spacecraft is the use of an X-band frequency (nominally 7162.3125 MHz) instead of an S-band frequency (nominally 2113.3125 MHz) for the uplink. Experimental X-band receivers will be flown on Galileo and Venus Radar Mapper. A new DSN RF system is being developed that can transmit at X-band and receive at both X-band and S-band.

One justification for an X-band uplink is increased doppler measurement accuracy, primarily for radio-science experiments to measure gravitational waves. The two largest sources of doppler measurement error in the present system, which uses an S-band uplink, are those caused by charged particles, in both the Earth's ionosphere and the interplanetary plasma, and receiver thermal noise. A new method of extracting frequency estimates from doppler phase data would make thermal noise effects negligible as long as no receiver cycle slips occur. Use of an X-band uplink should reduce the doppler measurement error caused by the Earth's ionosphere and the interplanetary plasma by at least an order of magnitude. Thus, use of an X-band uplink yields the potential for substantial improvement in doppler system accuracy.

A critical question is the phase stability of the ground radio equipment. Considerable care has been taken in the design of this equipment, and end-to-end tests, using test translators instead of a transponder, will be made to verify the phase stability of the DSN RF system hardware. However, test results with a test translator may differ significantly from those that would be observed with a transponder at a significant distance from the DSN station.

The analysis presented here determines the phase transfer function between each of a number of phase error sources in

the DSN RF system equipment and the X-band and S-band doppler extractor outputs. Both test configurations, using test translators, and the actual tracking configuration with a transponder and finite two-way time delay are considered. Comparing the phase transfer functions for the test configurations with those for the actual tracking configuration will yield considerable qualitative insight into which phase error sources will be accurately measured by the planned end-to-end measurements and which will not be accurately measured by these tests. These phase transfer functions can also be used in a future analysis to quantitatively calculate the contribution of each phase error source, given an Allen variance curve or phase noise power spectrum for the error source, to the doppler system inaccuracy.

A number of phase error sources are considered in this analysis. Θ_A is the phase of the hydrogen maser's 100 MHz output. Since both the X-band and S-band doppler extractor outputs analyzed in this article are 1 MHz biased doppler, the $0.01\Theta_A$ terms in the equations for the doppler extractor outputs are the 1.0 MHz bias. All other terms in Θ_A represent error. Other sources in the transmitting path are the phase error ϕ_{CC}, introduced by the temperature compensated cable between the hydrogen maser and the antenna-mounted equipment; the phase error ϕ_{UC}, introduced by the uncompensated cable between the exciter frequency synthesizer in the control room and the antenna-mounted equipment; the phase error ϕ_{ES}, introduced by the exciter frequency synthesizer; the phase error ϕ_{EM}, contributed by the exciter modulator; and the phase error ϕ_{PA}, contributed by the RF power amplifier. X-band receiver phase errors are the phase error ϕ_{XV}, contributed by the X-band receiver VCO; and the phase error ϕ_{XS}, contributed by the X-band receiver frequency synthesizer. S-band receiver phase errors are the phase error ϕ_{SV}, contributed by the S-band receiver VCO; and the phase error ϕ_{SS}, contributed by the S-band receiver frequency synthesizer.

A very simplified block diagram of the DSN RF system is shown in Fig. 1, which also shows the spacecraft transponder. The analytical models for the blocks shown in Fig. 1 are discussed in Section II. The results of combining the analytical models discussed in Section II are shown in Sections III and IV. Sections III and IV show the transfer function between each error source and the X-band and S-band doppler extractor outputs. In each case, the transfer functions for the test configurations are compared with those for the actual tracking configuration. Section V presents the report's conclusions.

II. System Elements

This section presents the analytical model and transfer functions for the major elements of the (X/X,X/S) system. The numbers printed on the figures associated with this section are the frequencies in MHz at various points of the (X/X,X/S) system, given that the uplink frequency is 7162.3125 MHz, the X-band downlink frequency is 8415 MHz, and the S-band downlink frequency is 2295 MHz. In addition, the symbol Θ with various subscripts is used to denote the total signal phase at various points in the system, and the symbol ϕ with appropriate subscripts is used for the phase noise or phase instabilities in the system.

The symbols M_E, M_X, and M_S are introduced in this section to denote the exciter, X-band RF receiver, and S-band RF receiver frequency synthesizer multiplication factors. These multiplication factors are adjusted to change the transmitter, X-band receiver and S-band receiver frequencies. The Appendix derives the relationships between M_E, M_X, and M_S and the transmitter and receiver frequencies and the interrelationship between M_E, M_S, and M_X

A. Exciter RF Frequency Generator

The analytical model for the exciter RF frequency generator is shown in Fig. 2. Θ_A is the phase of the 100 MHz hydrogen maser frequency standard output. The maser and the exciter frequency synthesizer, which performs the "$\times M_E$" frequency multiplication, are located in the DSN station control room. The phase error ϕ_{ES} rerpesents the phase error contribution of the exciter frequency synthesizer.

The rest of the equipment modeled in Fig. 2 is antenna-mounted. The 41.39453125 MHz (nominal) output of the frequency synthesizer is transmitted to the antenna via an uncompensated cable. $M_E\,\phi_{UC}$ represents the phase instability of this cable, where ϕ_{UC} represents the phase instability normalized to 100 MHz. The 100 MHz maser output is sent to the antenna via a compensated cable, which introduces a phase error ϕ_{CC}. On the antenna the "$\times$ 16" multiple of the uncompensated cable output is added (in phase) to the "$\times$ 65" multiple of the compensated cable output to yield the RF signal phase Θ_D. The outputs, Θ_B and Θ_C, of the uncompensated and compensated cables are also used by the doppler extractors.

From Fig. 2,

$$\Theta_B = M_E\,(\phi_{UC} + \Theta_A) + \phi_{ES} \tag{1}$$

$$\Theta_C = \phi_{CC} + \Theta_A \tag{2}$$

and

$$\Theta_D = (65 + 16\,M_E)\,\Theta_A + 16\,M_E\,\phi_{UC} + 65\,\phi_{CC} + 16\,\phi_{ES} \tag{3}$$

B. Transmitter

In the transmitter analytical model, ϕ_{EM} will be used to denote the phase error introduced by the exciter-modulator and ϕ_{PA} to denote that introduced by the RF power amplifier. The exciter-modulator/power amplifier element incorporates a phase compensation loop. The feedback point for this loop may be either the exciter-modulator output or the RF power amplifier output. Thus, two analytical models are required for the exciter-modulator/power amplifier element. Figure 3 shows the analytical model for feedback from the exciter output. Figure 4 shows the analytical model for feedback from the power amplifier output. Other than the feedback point for the phase compensation loop, the two analytical models are identical.

If feedback is taken from the exciter output, the exciter output will be

$$\Theta_E = \Theta_D + [1 - H(s)]\ \phi_{EM} \tag{4}$$

and the power amplifier output will be

$$\Theta_F = \Theta_D + [1 - H(s)]\ \phi_{EM} + \phi_{PA} \tag{5}$$

where

$$H(s) = \left(\frac{k}{k+1}\right) \frac{1}{1 + \frac{\tau}{k+1} s} \tag{6}$$

$$1 - H(s) = \left(\frac{1}{k+1}\right) \frac{1 + \tau s}{1 + \frac{\tau}{k+1} s} \tag{7}$$

and

$$k = k_D G_A k_\phi \tag{8}$$

is the gain of the phase compensation loop.

If feedback is taken from the power amplifier output, the exciter output will be

$$\Theta_E = \Theta_D + [1 - H(s)]\ \phi_{EM} + H(s)\ \phi_{PA} \tag{9}$$

and the power amplifier output will be

$$\Theta_F = \Theta_D + [1 - H(s)]\ (\phi_{EM} + \phi_{PA}) \tag{10}$$

Figure 5 shows sketches of $|H(s)|$ and $|1 - H(s)|$ as a function of frequency.

$$f_L = \frac{1}{(2\pi\tau)} \tag{11}$$

and

$$f_H = (k+1) f_L \tag{12}$$

Since $k_D = 0.9 \times 10^{-3}$ V/deg, $G_A = 1220$, and $k = 47$ deg/V, k is equal to 51.606. τ may be equal to either 4 s or 0.528×10^{-2} s. For $\tau = 4$ s, $f_L = 0.040$ Hz and $f_H = 2.09$ Hz. For $\tau = 0.528 \times 10^{-2}$ s, $f_L = 30.142$ Hz and $f_H = 1585.70$ Hz.

Which phase compensation loop mode is used will be determined by the spectral characteristics of the signal to be phase-modulated onto the uplink carrier. Since the uplink phase modulator is inside the transmitter phase compensation loop, uplink command or ranging modulation will be filtered by the $1 - H(s)$ response of the loop. If the phase modulation has significant components at frequencies below 1585.70 Hz, the narrowband ($\tau = 4$ s) mode must be used. For example, the narrowband mode must be used for systems using a Voyager-type command system. For systems whose uplink phase modulation has no significant spectral components below 1585.70 Hz, the wideband mode can be used. This results in better transmitter phase noise compensation and, consequently, in better doppler tracking accuracy. Systems that use the NASA standard command system or that operate in a mode where uplink command modulation is not currently required can use the wideband mode.

C. Test Translators

Two test translators are provided in the exciter/receiver system. One generates an X-band downlink signal (8415 MHz nominal) from the X-band uplink signal (7162.3125 MHz nominal). The other generates an S-band downlink signal (2295 MHz nominal) from the X-band uplink signal. The appropriate references for the frequency translation are provided by the X-band and S-band doppler extractors. Analytical models for the two test translators are shown in Figs. 6 and 7. Using the notation shown in these figures, the X-band test translator output is

$$\Theta_I = \Theta_H + \Theta_G \tag{13}$$

where Θ_H is the X-band test translator input and Θ_G is the (nominally 1252.6875 MHz) reference from the X-band doppler extractor, and the S-band test translator output is

$$\Theta_L = \Theta_K - \Theta_J \tag{14}$$

where Θ_K is the S-band test translator input and Θ_J is the (nominally 4867.3125 MHz) reference signal from the S-band doppler extractor.

D. X-Band Receiver Analytical Model

Figure 8 shows the analytical model for the X-band receiver. The first down-converter and the "X 81" frequency multiplier are antenna mounted. The remainder of the equipment is located in the control room. After the first down-conversion to 315 MHz (nominal), the "X 50" multiple of the difference between the 48.3 MHz (nominal) output of the X-band receiver synthesizer and the 21 MHz VCO output is used to translate the received signal to 1050 MHz. Two additional stages of down-conversion, to 50 MHz and 10 MHz, precede the phase detector, which operates at 10 MHz. The reference signals for these two down-converters and the phase detector are derived from the 100 MHz hydrogen maser frequency standard. The standard also drives the receiver frequency synthesizer.

Three sources of phase noise are shown in Fig. 8 in addition to the phase noise on the hydrogen maser frequency reference. As noted previously ϕ_{CC} is the phase instability in the temperature-compensated cable used to carry the 100 MHz reference signal up to the antenna-mounted equipment. ϕ_{XS} is the phase noise contributed by the X-band receiver frequency synthesizer. ϕ_{XV} is the phase noise contributed by the X-band receiver VCO.

Analyzing Fig. 8, if we define

$$H_X(s) = \frac{1 + \tau_2 s}{1 + \left(\tau_2 + \frac{1}{k_L}\right)s + \frac{\tau_1}{k_L}s^2} \tag{15}$$

where

$$k_L = 50\,\alpha\,k_D k_F k_V, \tag{16}$$

the phase of the X-band receiver output is

$$\Theta_N = H_X(s)\,(\Theta_M - 81\,\phi_{CC}) + [1 - H_X(s)]\,50\,\phi_{XV} - [1 - H(s)]50\,\phi_{XS} - [50\,M_X + (70.5 - 50\,M_X)\,H_X(s)]\,\Theta_A \tag{17}$$

where Θ_M is the phase of the X-band receiver input.

E. S-Band Receiver Analytical Model

Figure 9 shows the analytical model for the S-band receiver. The first down-converter and the "X 20" frequency multiplier are antenna-mounted. The remainder of the equipment is located in the control room.

The block diagram of the S-band receiver is very nearly identical to that of the X-band receiver. The initial down-conversion uses a "X 20" multiple of the 100 MHz hydrogen maser reference signal. Thus, for a 2295 MHz received signal, the initial down-conversion is to 295 MHz. This IF frequency requires the S-band receiver frequency synthesizer output to be at 47.9 MHz. Otherwise, the block diagrams of the X-band and S-band receivers are identical.

Analyzing Fig. 9, if we define

$$H_S(s) = \frac{1 + \tau_2 s}{1 + \left(\tau_2 + \frac{1}{k_L}\right)s + \frac{\tau_1}{k_L}s^2} \tag{18}$$

where

$$k_L = 50\,\alpha\,k_D k_F k_V, \tag{19}$$

$$\Theta_P = H_S(s)\,(\Theta_O - 20\,\phi_{CC}) + [1 - H_S(s)]\,50\,\phi_{SV} - [1 - H_S(s)]\,50\,\phi_{SS} - [50\,M_S + (9.5 - 50\,M_S)\,H_S(s)]\,\Theta_A \tag{20}$$

Note that the values of τ_1, τ_2, k_L, α, k_D, k_F, and k_V for the S-band receiver may not be the same as those for the X-band receiver.

F. X-Band Doppler Extractor

Figure 10 shows the analytical model for the X-band doppler extractor. Input phases are Θ_B, Θ_C, and Θ_D from the exciter RF frequency generator, the hydrogen maser frequency standard phase Θ_A, and the phase Θ_N of the X-band receiver output. Note that the phase of the "X 10480/749" frequency multiplier output is the phase Θ_G of the reference signal used by the X-band test translator.

Analyzing Fig. 10, the phase of the reference signal used by the X-band test translator is

$$\Theta_G = \left(\frac{2096}{749}\right)\Theta_B + \left(\frac{8515}{749}\right)\Theta_C \tag{21}$$

and the phase of the X-band doppler extractor output is

$$\Theta_Q = \Theta_N - 10.49\,\Theta_A - \left(16 + \frac{2096}{749}\right)\Theta_B + \left(16 - \frac{8515}{749}\right)\Theta_C \tag{22}$$

G. S-Band Doppler Extractor

Figure 11 shows the analytical model for the S-band doppler extractor. Input phases are Θ_B and Θ_C from the exciter RF frequency generator, the hydrogen maser frequency standard phase Θ_A, and the phase Θ_P of the S-band receiver output. Note that the phase of the "X 40720/749" frequency multipler output is the phase Θ_J of the reference signal used by the S-band test translator.

In the analysis of Fig. 11, the phase of the reference signal used by the S-band test translator is

$$\Theta_J = \left(\frac{8144}{749}\right)\Theta_B + \left(\frac{33085}{749}\right)\Theta_C \tag{23}$$

and the phase of the S-band doppler extractor output is

$$\Theta_R = \Theta_P - 10.49\,\Theta_A - \left(16 - \frac{8144}{749}\right)\Theta_B - \left(45 - \frac{33085}{749}\right)\Theta_C \tag{24}$$

III. X/X Transfer Functions

A. Testing

If the system is in the test mode, the X-band receiver input is equal to the X-band test translator output. Thus,

$$\Theta_M = \Theta_I = \Theta_H + \Theta_G \tag{25}$$

Switching in the transmitter permits the test translator input to be either the RF power amplifier output ($\Theta_H = \Theta_F$) or the exciter output ($\Theta_H = \Theta_E$). Futhermore, additional switching permits the feedback point for the transmitter phase-compensation loop to be taken from either the RF power amplifier output or the exciter output. For phase-compensation loop feedback from the power amplifier output, Θ_E and Θ_F are given by Eqs. (9) and (10). For feedback from the exciter output, Θ_E and Θ_F are given by Eqs. (4) and (5).

Combining the appropriate relationship between the X-band test translator input Θ_H and either Θ_E or Θ_F and the transfer functions derived in Section 2 of this report yields

$$\Theta_Q = 0.01\,\Theta_A + \phi_X \tag{26}$$

where Θ_Q is the X-band doppler extractor output, $0.01\,\Theta_A$ is the 1 MHz bias, and the X-band doppler phase error is

$$\begin{aligned}\phi_X = {} & \left[-(65 + 16\,M_E)\left(\frac{880}{749}\right) + (70.5 - 50\,M_X)\right] \\ & \times [1 - H_X(s)]\,\Theta_A - 16\left(\frac{880}{749}\right)M_E\,[1 - H_X(s)]\,\phi_{UC} \\ & + \left(81 - 65 \cdot \frac{880}{749}\right)[1 - H_X(s)]\,\phi_{CC} \\ & - 16\left(\frac{880}{749}\right)[1 - H_X(s)]\,\phi_{ES} \\ & + [1 - H_X(s)]\,50\,(\phi_{XV} - \phi_{XS}) \\ & + H_X(s)\,\phi_T\end{aligned} \tag{27}$$

The net transmitter phase error is

$$\left.\begin{aligned}\phi_T &= [1 - H(s)]\,(\phi_{EM} + \phi_{PA}), && O = PA, FB = PA \\ &= [1 - H(s)]\,\phi_{EM}, && O = E, FB = E \\ &= [1 - H(s)]\,\phi_{EM} + \phi_{PA} && O = PA, FB = E \\ &= [1 - H(s)]\,\phi_{EM} + H(s)\,\phi_{PA}, && O = E, FB = PA\end{aligned}\right\} \tag{28}$$

In Eq. (28), $O = PA$ denotes output from the power amplifier, $O = E$ denotes output from the exciter, $FB = PA$ denotes feedback from the power amplifier output, and $FB = E$ denotes feedback from the exciter output.

B. Actual Spacecraft Tracking

During actual spacecraft tracking, the X-band receiver input will be

$$\Theta_M = \left(\frac{880}{749}\right)H_T(s)\exp(-\tau s)\,\Theta_F \tag{29}$$

where $H_T(s)$ is the spacecraft receiver transfer function, τ is the two-way group delay and Θ_F is the RF power amplifier output. Note that error sources in the spacecraft and the ground receiving system thermal noise have been neglected.

Combining Eq. (29) and the transfer functions derived in Section II yields, for this case,

$$\phi_X = \Big\{ -(65 + 16\,M_E)\left(\frac{880}{749}\right)[1 - H_T(s)\,H_X(s)\exp(-\tau s)] + (70.5 - 50\,M_X)[1 - H_X(s)] \Big\}\,\Theta_A - 16\left(\frac{880}{749}\right) M_E\,[1 - H_T(s)\,H_X(s)\exp(-\tau s)]\,\phi_{UC} + \Big\{ 81\,[1 - H_X(s)] - 65\left(\frac{880}{749}\right) \times [1 - H_T(s)\,H_X(s)\exp(-\tau s)] \Big\}\,\phi_{CC} - 16\left(\frac{880}{749}\right)[1 - H_T(s)\,H_X(s)\exp(-\tau s)]\,\phi_{ES} + [1 - H_X(s)]\,50\,(\phi_{XV} - \phi_{XS}) + \left(\frac{880}{749}\right) H_T(s)\,H_X(s)\exp(-\tau s)\,\phi_T \qquad (30)$$

where

$$\phi_T = [1 - H(s)]\,(\phi_{EM} + \phi_{PA}), \quad FB = PA$$
$$= [1 - H(s)]\,\phi_{EM} + \phi_{PA}, \quad FB = E \qquad (31)$$

where $FB = PA$ indicates the feedback is taken from the power amplifier output and $FB = E$ indicates the feedback is taken from the exciter output. Note that for the configurations possible during actual spacecraft tracking, Eqs. (28) and (31) are the same.

In either the test or tracking configurations, ϕ_X is the total X-band doppler phase error. Compare Eq. (27) for ϕ_X during testing with Eq. (30) for ϕ_X during actual spacecraft tracking. Note that the error contribution of some terms will be much greater in the tracking configuration than in a test configuration. One difference is in the transfer functions for Θ_A, ϕ_{UC}, ϕ_{CC}, and ϕ_{ES}. In a test configuration the appropriate transfer function is, from Eq. (27), $1 - H_X(s)$. In the tracking configuration, the appropriate transfer function is from Eq. (30), either $1 - H_X(s)$, $1 - H_T(s)\,H_X(s)\exp(-\tau s)$, or some weighted sum of these two transfer functions.

Assume $H_T(s) = 1$ and $H_X(s) = 1$ over the frequency band occupied by the error sources. During testing the error contributions from Θ_A, ϕ_{UE}, ϕ_{CC}, and ϕ_E would be zero. During actual spacecraft tracking, however,

$$1 - H_T(i\omega)\,H_X(i\omega)\exp(-i\omega\tau) = 1 - \exp(-i\omega\tau) = 2i\sin(\omega\tau/2)\exp(-i\omega\tau/2) \qquad (32)$$

For large τ (large spacecraft distance), this transfer function has significant low frequency response. Thus, the doppler phase errors caused by Θ_A, ϕ_{UC}, ϕ_{CC}, and ϕ_{ES} may be much greater during actual spacecraft tracking than during testing. However, the doppler phase errors caused by the X-band receiver phase noise, ϕ_{XV} and ϕ_{XS}, will be the same during tests as during actual spacecraft tracking. Thus, doppler system tests will accurately measure the X-band doppler system phase errors resulting from ϕ_{XV} and ϕ_{XS} but not those resulting from Θ_A, ϕ_{UC}, ϕ_{CC}, and ϕ_{ES}.

For the same transmitter mode of operation, the transmitter contribution to the X-band doppler phase noise should be somewhat greater during actual tracking than during testing. Comparison of Eqs. (27) and (30) shows that the transmitter contribution to X-band doppler phase noise is $H_X(s)\phi_T$ during testing and $(880/749)\,H_T(s)\,H_X(s)\exp(-\tau s)\,\phi_T$ during actual tracking. Thus, if one assumes $H_T(s) = 1$ over the frequency band of interest, the Allen variance of the transmitter contribution to X-band doppler phase noise will be a factor of $(880/749)^2$ greater during actual spacecraft tracking than during testing with the same transmitter mode.

IV. X/S Transfer Functions

A. Testing

If the system is in the test mode, the S-band receiver input is equal to S-band test translator output. Thus

$$\Theta_R = \Theta_L = \Theta_K - \Theta_J \qquad (33)$$

Thus, following the same approach used for the X-band doppler phase, the S-band doppler phase will be

$$\Theta_R = 0.01\,\Theta_A + \phi_S \qquad (34)$$

where

$$\phi_S = \Big[-(65 + 16\,M_E)\left(\frac{240}{749}\right) + (9.5 - 50\,M_S)\Big]\,[1 - H_S(s)]\,\Theta_A - 16\left(\frac{240}{749}\right) M_E\,[1 - H_S(s)]\,\phi_{UC} - \left(65\cdot\frac{240}{749} - 20\right)[1 - H_S(s)]\,\phi_{CC}$$

$$- 16\left(\frac{240}{749}\right)[1 - H_S(s)]\ \phi_{ES}$$

$$+ [1 - H_S(s)]\ 50\ (\phi_{SV} - \phi_{SS})$$

$$+ H_S(s)\ \phi_T \qquad (35)$$

is the S-band doppler phase error, and ϕ_T is as specified in Eq. (28).

B. Actual Spacecraft Tracking

During actual spacecraft tracking, the S-band receiver input will be

$$\Theta_O = \left(\frac{240}{749}\right) H_T(s) \exp(-\tau s)\ \Theta_F \qquad (36)$$

where $H_T(s)$ is the spacecraft receiver transfer function, τ is the two-way group delay and Θ_F is the X-band RF power amplifier output. Note that error sources in the spacecraft and the ground receiving system thermal noise have been neglected.

Combining Eq. (36) and the transfer functions derived in Section II yields, for this case,

$$\phi_S = \left\{-(65 + 16\,M_E)\left(\frac{240}{749}\right)[1 - H_T(s)\,H_S(s) \exp(-\tau s)]\right.$$

$$\left. + (9.5 - 50\,M_S)\,[1 - H_S(s)]\right\}\Theta_A$$

$$- 16\,M_E\left(\frac{240}{749}\right)[1 - H_T(s)\,H_S(s) \exp(-\tau s)]\ \phi_{UC}$$

$$- \left\{65\left(\frac{240}{749}\right)[1 - H_T(s)\ H_S(s) \exp(-\tau s)]\right.$$

$$\left. - 20\,[1 - H_S(s)]\right\}\phi_{CC}$$

$$- 16\left(\frac{240}{749}\right)[1 - H_T(s)\,H_S(s) \exp(-\tau s)]\ \phi_{ES}$$

$$+ [1 - H_S(s)]\ 50\ (\phi_{SV} - \phi_{SS})$$

$$+ \left(\frac{240}{749}\right) H_T(s)\,H_S(s) \exp(-\tau s)\ \phi_T \qquad (37)$$

where the transmitter phase error ϕ_T is that specified in Eq. (31). Comparing Eqs. (35) and (37), one notes that, as seen previously in the X/X analysis, the effects of Θ_A, ϕ_{UC}, ϕ_{CC}, and ϕ_{ES} will be significantly greater during actual spacecraft tracking than during testing. The effect of ϕ_{SV} and ϕ_{SS} will be the same during tests and actual tracking. Finally, assuming $H_T(s) = 1$ over the frequency band of interest, the Allen variance of the transmitter contribution to S-band doppler phase noise will be a factor of $(240/749)^2$ less during actual spacecraft tracking than during tests with the same transmitter mode.

V. Conclusion

Phase transfer functions between each phase error source and the X-band and S-band doppler extractor outputs for both the system configurations used during testing and those used during actual spacecraft tracking have been determined here. These results show that the contribution of the exciter frequency synthesizer phase error ϕ_{ES}, the hydrogen maser phase Θ_A, the compensated cable phase error ϕ_{CC} and the uncompensated cable phase error ϕ_{UC} to the X-band and S-band doppler phase errors will be very much less in test configurations than in actual spacecraft tracking. The primary reason for this difference is the two-way time delay in the tracking configuration. The contributions of the exciter-modulator phase error ϕ_{EM} and the RF power amplifier phase error ϕ_{PA} in the test configurations will be slightly less at X-band and significantly greater at S-band than during actual spacecraft tracking. The reason for this difference is that the transponder frequency "gains" are 880/749 for the X-band downlink and 240/749 for the S-band downlink, whereas the frequency "gains" of both the X-band and the S-band test translators, for the transmitter phase noises ϕ_{EM} and ϕ_{PA}, are unity. Finally, the contributions of the VCO and frequency synthesizer phase errors in both the X-band and S-band receivers will be the same during testing as during actual spacecraft tracking.

Given power spectra for the phase error sources, the transfer functions presented in this article can be used to calculate quantitative values instead of the qualitative observations presented above. The DSN station stability program will supply the measurements of the individual error sources. A subsequent article will compare the results obtained by combining the individual measurements, using the test mode phase transfer functions presented in this article, with the results from the end-to-end tests. This comparison is expected to aid in the understanding of the end-to-end test data and to either verify or suggest modifications to the preceding analysis.

Appendix

Exciter, X-Band Receiver, and S-Band Receiver Frequency Synthesizer Multiplication Factors

The exciter, X-band receiver, and S-band receiver frequency synthesizer multiplication factors (M_E, M_X, and M_S) will be adjusted to obtain the desired transmitter and receiver frequencies.

I. Exciter

The frequency of the transmitted RF signal (in MHz) will be

$$f_T = 6500 + 1600\,M_E \tag{A-1}$$

Thus,

$$M_E = \frac{(f_T - 6500)}{1600} \tag{A-2}$$

For f_T = 7162.3125 MHz, M_E = 0.4139453125.

II. X-Band Receiver

If f_{RX} is the center frequency (in MHz) of the X-band receiver,

$$50\,(100\,M_X - 21) - (f_{RX} - 8100) = 1050 \tag{A-3}$$

or

$$M_X = \frac{(f_{RX} - 6000)}{5000} \tag{A-4}$$

Since

$$f_{RX} = \left(\frac{880}{749}\right) f_T\,, \tag{A-5}$$

$$M_X = \frac{(1408\,M_E + 1226)}{3745} \tag{A-6}$$

For f_{RX} = (880/749) 7162.3125 MHz = 8415 MHz, M_X = 0.483.

III. S-Band Receiver

If f_{RS} is the center frequency (in MHz) of the S-band receiver,

$$50\,(100\,M_S - 21) - (f_{RS} - 2000) = 1050 \tag{A-7}$$

or

$$M_S = \frac{(f_{RS} + 100)}{5000} \tag{A-8}$$

Since

$$f_{RS} = \left(\frac{240}{749}\right) f_T \tag{A-9}$$

$$M_S = \frac{(3840\,M_E + 16349)}{37450} \tag{A-10}$$

For f_{RS} = (240/749) 7126.3125 MHz = 2295 MHz, M_S = 0.479.

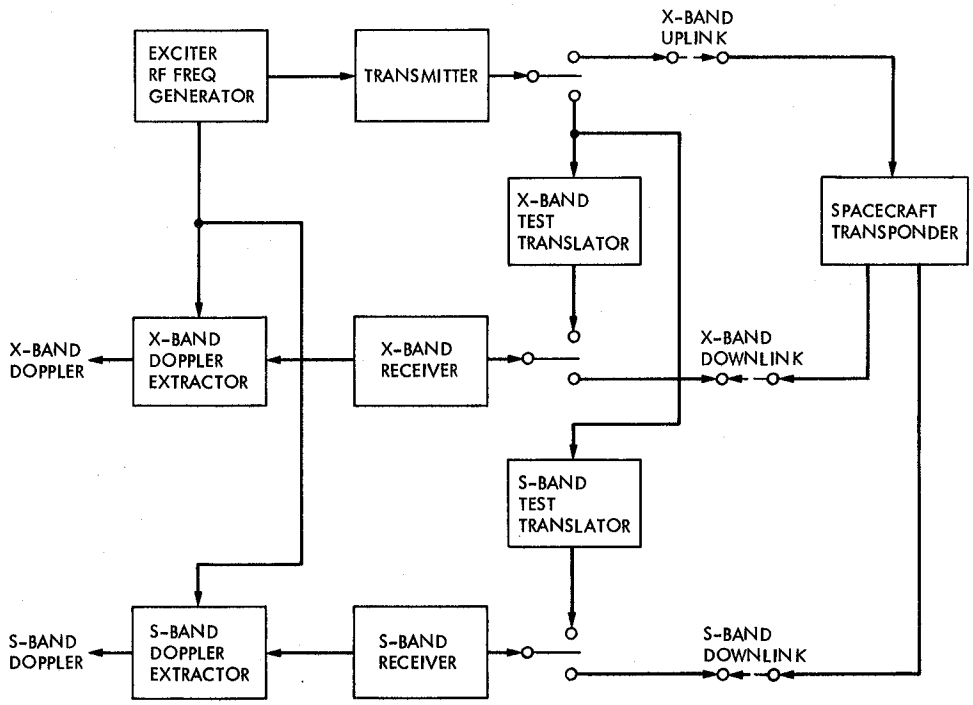


Fig. 1. (X/X, X/S) system functional block diagram

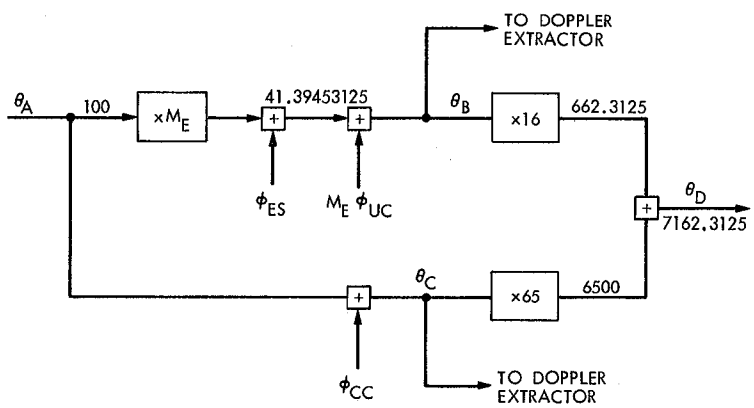


Fig. 2. Exciter RF frequency generator analytical model

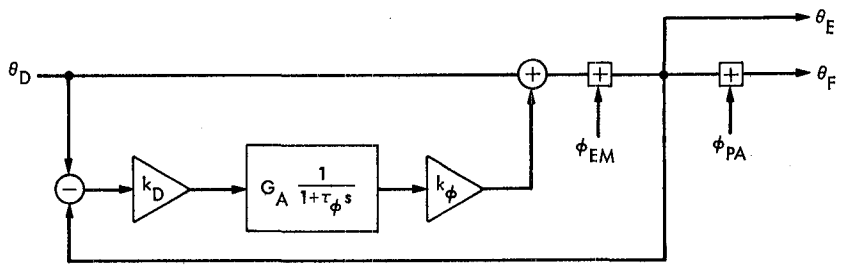


Fig. 3. Transmitter analytical model for feedback from the exciter output

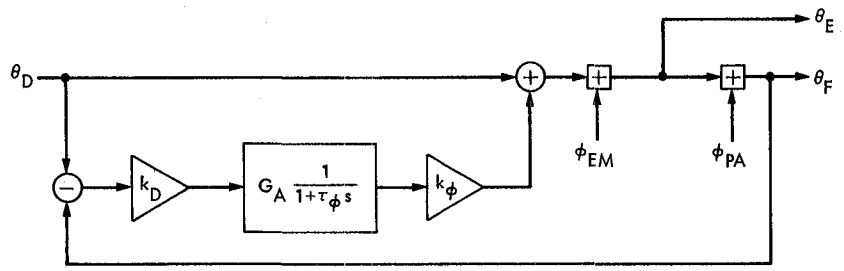


Fig. 4. Transmitter analytical model for feedback from the power amplifier output

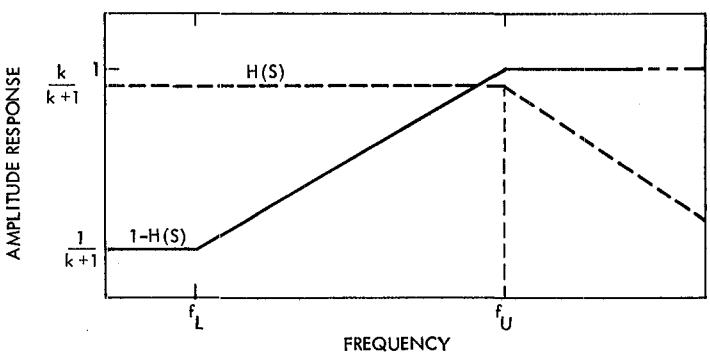


Fig. 5. Transmitter phase compensation loop amplitude responses

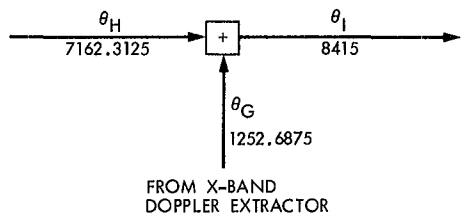


Fig. 6. X-band test translator analytical model

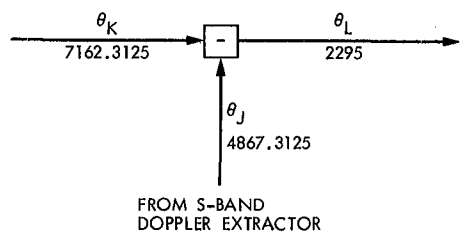


Fig. 7. S-band test translator analytical model

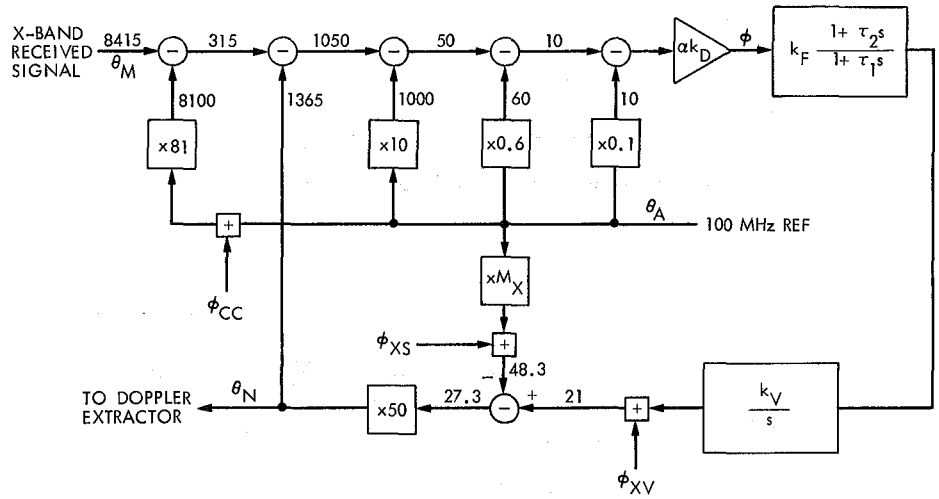


Fig. 8. X-band receiver analytical model

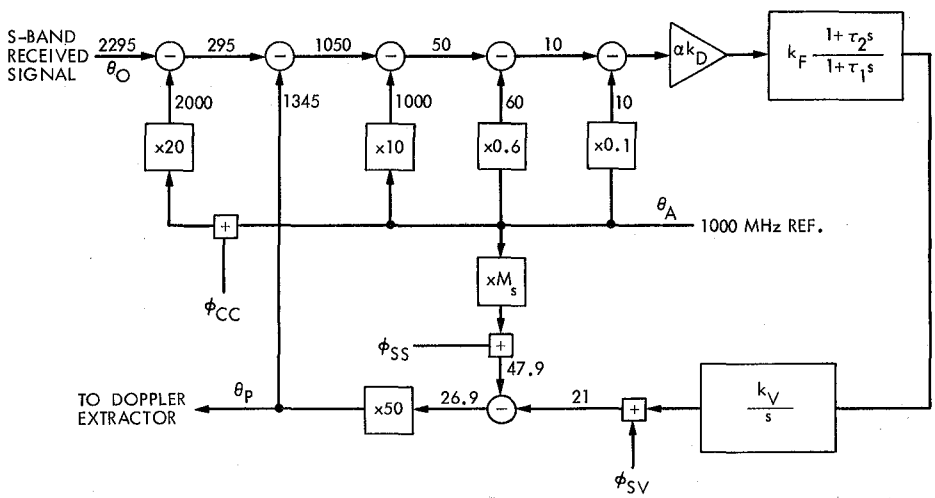


Fig. 9. S-band receiver analytical model

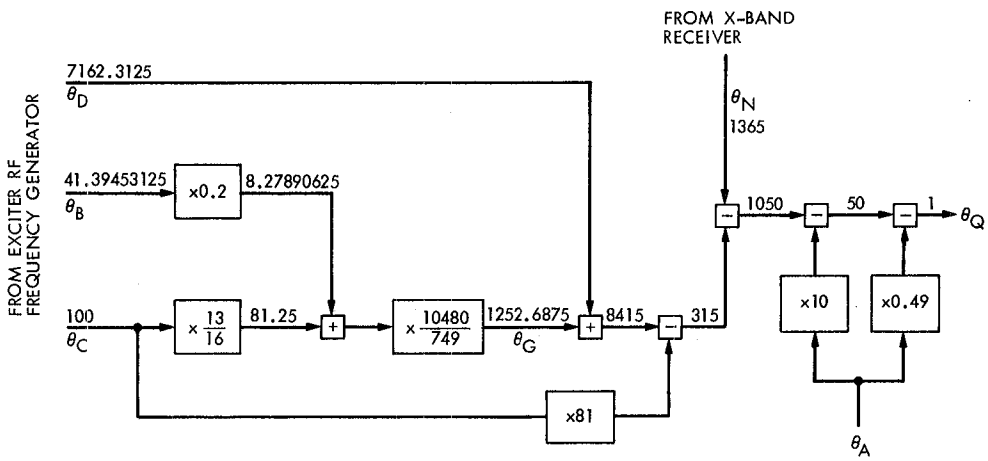


Fig. 10. X-band doppler extractor analytical model

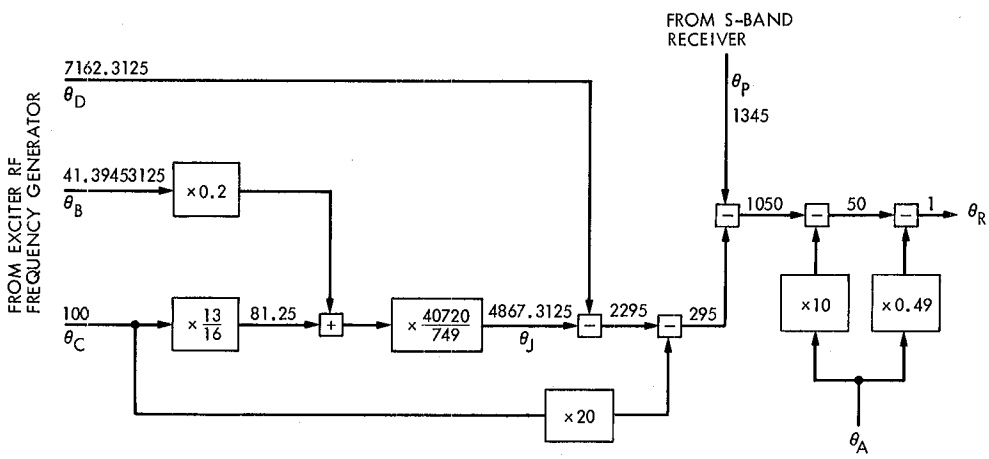


Fig. 11. S-band doppler extractor analytical model

Transfer Function Bounds for Partial-Unit-Memory Convolutional Codes Based on Reduced State Diagram

P. J. Lee
Communication Systems Research Section

The performance of a coding system consisting of a convolutional encoder and a Viterbi decoder can be analytically found by the well-known transfer function bounding technique. For the partial-unit-memory byte-oriented convolutional encoder with m_o binary memory cells and k_o ($> m_o$) inputs, a state diagram of 2^{k_o} states has been used for the transfer function bound. In this article, it is shown that a reduced state diagram of $(2^{m_o} + 1)$ states can be used for easy evaluation of transfer function bounds for partial-unit-memory codes.

I. Introduction

A class of convolutional codes called unit-memory byte-oriented convolutional codes (UM codes) was introduced by Lee (Ref. 1) and Lauer (Ref. 2). The encoder structure of an $(m_o, k_o/n_o)$ UM code is shown in Fig. 1, where m_o is the number of binary memory cells, k_o is the number of inputs, and n_o is the number of outputs. The code rate r of this UM code is k_o/n_o. UM codes with $m_o < k_o$ are called partial-UM codes (PUM codes) (Ref. 2). For consistency in terminology, we will call UM codes with $m_o = k_o$, full-UM codes (FUM codes).

In Ref. 1, it was shown that FUM codes are superior to conventional convolutional codes in the sense of having larger free distances for given pairs of m_o and r. In Ref. 2, PUM codes are shown to be even better than FUM codes in the same sense. Also, as inner codes for the concatenated coding systems with Reed-Solomon (RS) outer codes, FUM codes were shown to be better than conventional convolutional codes, due to their byte-oriented natures (Refs. 1 and 3). For example, it was shown (Ref. 3) that, with a 6-bit RS outer code and convolutional inner codes with $m_o = 6$ and $r = 1/3$, the use of FUM code can save 0.3 dB in required signal-to-noise ratio over the use of conventional convolutional code for a given performance. For the same applications, we expect that PUM codes will be more useful since we can employ larger symbol size RS codes with inner codes of small complexities (e.g., 8-bit RS code and $(5, 8/n_o)$ PUM code).

The performance of a coding system employing a convolutional encoder and a Viterbi decoder can be analytically evaluated by the transfer function bound (TFB) based on the corresponding state diagram (SD) for that code (Refs. 4 and 5). However, for an $(m_o, k_o/n_o)$ PUM code, the TFB should be evaluated from a SD of 2^{k_o} states, and hence the inversion of a $(2^{k_o} - 1) \times (2^{k_o} - 1)$ matrix is required. The purpose of this

report is to present a reduced state diagram of $(2^{m_o} + 1)$ states to reduce the required computational burden for the evaluation of TFB.

II. Preliminaries

This section briefly reviews the encoder structure and the state diagram (SD) for the evaluation of transfer function bound (TFB) for FUM codes. Readers who are not familiar with TFB are referred to Ref. 4 or Ref. 5.

The encoder structure of an $(m_o, k_o/n_o)$ UM code is shown in Fig. 1. The $(k_o + m_o)$-input to n_o-output connection box including n_o mod-2 adders is often represented by an $n_o \times (k_o + m_o)$ binary matrix G, called a code generator matrix. The nth bit in the tth output vector, y_n^t, $n = 1,2,\ldots, n_o$, $t = 1,2,\ldots$, is then

$$y_n^t = \sum_{k=1}^{k_o} G(n,k) \cdot x_k^t \oplus \sum_{k=1}^{m_o} G(n, k_o + k) \cdot x_k^{t-1} \quad (1)$$

where $\sum$ and $\oplus$ represent the mod-2 summations and $x_k^t \in (0,1)$, for $k = 1,2,\ldots, k_o$, and $t = 1,2,\ldots$ ($x_k^o = 0$, $k = 1,2,\ldots, m_o$, by convention).

First, consider FUM codes with $m_o = k_o$. We define the "present state" at time t, S^t, as

$$S^t = \left(x_1^{t-1}, x_2^{t-1}, \ldots, x_{k_o}^{t-1}\right)$$

The "next state" at time t is then S^{t+1}. Notice that the number of possible states is 2^{k_o} $(= M_1)$, regardless of the number of outputs n_o and the time t. Define the state space, $\mathcal{S}$, as

$$\mathcal{S} = \{s_0, s_1, \ldots, s_{M_1 - 1}\}$$

where

$$s_i = (x_1, x_2, \ldots, x_{k_o})$$

with

$$i = \sum_{k=1}^{k_o} x_k \cdot 2^{k_o - k}$$

for any time t.

The SD for TFB of this $(k_o, k_o/n_o)$ FUM code consists of M_1 nodes, and M_1^2 directed branches with associated branch metrics. Node i represents the state s_i, and the directed branch from node i to node j represents the transition from state s_i to state s_j. The branch metric on the directed branch from node i to node j, $m(i,j)$, is given by

$$m(i,j) = D^{H^O(i,j)} Z^{H^I(i,j)} \quad (2)$$

where the values of H^O's depend on the code generator G and the type of channel to be used, and the values of H^I's depend on the type of performance measure of interest. Here, for the type of channel we will consider only the binary-input symmetric-output (BISO) channel, and for the type of performance measure we will consider the bit error rate (BER) and the $M_1 (= 2^{k_o})$-ary symbol error rate (SER). For the BISO channel, the H^O's in Eq. (2) are then

$$H^O(i,j) = \sum_{n=1}^{n_o} y_n(i,j)$$

where $y_n(i,j)$ is the y_n^t in Eq. (1) with $(x_1^{t-1}, x_2^{t-1}, \ldots, x_{k_o}^{t-1}) = s_i$ and $(x_1^t, x_2^t, \ldots, x_{k_o}^t) = s_j$. That is, $H^O(i,j)$ is the binary Hamming weight of the corresponding output vector. The H^I's in Eq. (2) when using the BER criterion, H_B^I's, are then

$$H_B^I(i,j) = \frac{\text{(number of 1's in the next state } s_j)}{k_o}$$

i.e., $H_B^I(i,j)$ is the normalized binary Hamming weight of the corresponding input vector. Likewise, the H^I's in Eq. (2) for M_1-ary SER, H_S^I's, are

$$H_S^I(i,j) = 0, \text{ if } j = 0$$

$$= 1, \text{ otherwise}$$

or $H_S^I(i,j)$ is the M_1-ary Hamming weight of the corresponding input symbol.

From the SD described, we find the transfer function $T(D,Z)$ as (Refs. 4 and 5)

$$T(D,Z) = B \cdot (I - A)^{-1} \cdot C$$

where I is the $(M_1 - 1) \times (M_1 - 1)$ unit matrix, and the $(M_1 - 1) \times (M_1 - 1)$ matrix A, the $(M_1 - 1)$-dimensional row vector B, and the $(M_1 - 1)$-dimensional column vector C are obtained from the SD for $i = 1,2,\ldots, M_1 - 1$ and $j = 1,2,\ldots,$

$M_1 - 1$, by $A(i, j) = m(i, j)$, $B(j) = m(0, j)$, and $C(i) = m(i, 0)$. Then with

$$\frac{\partial}{\partial Z} T(D, Z) = \frac{\partial B}{\partial Z} \cdot (I - A)^{-1} \cdot C + B \cdot (I - A)^{-1} \cdot \frac{\partial A}{\partial Z} \cdot (I - A)^{-1} \cdot C$$

the TFB for BER and SER can be found as

$$\text{BER} \leqslant c_o \cdot \frac{\partial}{\partial Z} T_B(D, Z) \Big|_{D=D_o,\ Z=1}$$

and

$$\text{SER} \leqslant c_o \cdot \frac{\partial}{\partial Z} T_S(D, Z) \Big|_{D=D_o,\ Z=1}$$

where D_o is the union Bhattacharyya distance of the coding channel (everything inside the encoder-decoder pair) and the constant c_o depends on the type of coding channel and the code. For example, when we use the binary antipodal signaling over an additive white Gaussian noise channel with no channel output quantization, D_o and c_o are given by (Ref. 5)

$$D_o = \exp\left(-\frac{E_s}{N_o}\right)$$

and

$$c_o = Q\left(2 \cdot d_f \cdot \frac{E_s}{N_o}\right) \cdot \exp\left(d_f \cdot \frac{E_s}{N_o}\right)$$

where N_o is the one sided noise spectral density, E_s is the received signal energy per channel bit, d_f is the free distance of the code, and

$$Q(w) = \int_w^\infty \exp\left(-\frac{z^2}{2}\right) \cdot \frac{dz}{\sqrt{2\pi}}$$

An illustrating example is shown in Fig. 2. Figure 2(a) shows the encoder structure of a (2, 2/3) FUM code including the code connections. The code generator matrix G is then

$$G = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

Its SD when using the BER criterion is shown in Fig. 2(b). From this, we have

$$B = \begin{bmatrix} D\ Z^{0.5} & D^3\ Z^{0.5} & D^2\ Z \end{bmatrix}$$

$$C = \begin{bmatrix} D^2 \\ D^2 \\ D^2 \end{bmatrix}, \quad A = \begin{bmatrix} D\ Z^{0.5} & D\ Z^{0.5} & D^2\ Z \\ D^3\ Z^{0.5} & D\ Z^{0.5} & Z \\ D\ Z^{0.5} & D\ Z^{0.5} & D^2\ Z \end{bmatrix}$$

And it is easy to show

$$\left.\frac{\partial T_B}{\partial Z}\right|_{Z=1} = \frac{0.5\,D^3 + 2\,D^4 + 1.5\,D^5 - 3\,D^6 - D^7 + 4\,D^8}{(1 - 3\,D + D^3 + D^4)^2}$$

If the SER is the performance measure of interest, then the corresponding SD is exactly the same as Fig. 2(b) but with $Z^{0.5}$ replaced by Z. This gives

$$\left.\frac{\partial T_S}{\partial Z}\right|_{Z=1} = \frac{D^3 + 2\,D^4 + 3\,D^5 - 5\,D^6 - 2\,D^7 + 5\,D^8 - D^{10}}{(1 - 3\,D + D^3 + D^4)^2}$$

III. Reduced State Diagram for TFB of PUM Codes

The transfer function bounding technique described in the previous section is valid when only one transition exists from a given state to another state. In other words, each encoder input requires at least one memory cell. Hence, the performance of an $(m_o, k_o/n_o)$ PUM code can be found by introducing $(k_o - m_o)$ dummy binary memory cells and by treating all of that as a $(k_o, k_o/n_o)$ FUM code. For an illustrating example, a (1, 2/3) PUM code with one dummy memory cell is shown in Fig. 3(a). The corresponding SD of 4 states is shown in Fig. 3(b). From this, we have

$$B = \begin{bmatrix} D^2\ Z^{0.5} & D^2\ Z^{0.5} & D^2\ Z \end{bmatrix}$$

$$C = \begin{bmatrix} 1 \\ D \\ D \end{bmatrix}, \quad A = \begin{bmatrix} D^2\ Z^{0.5} & D^2\ Z^{0.5} & D^2\ Z \\ D^3\ Z^{0.5} & D\ Z^{0.5} & D\ Z \\ D^3\ Z^{0.5} & D\ Z^{0.5} & D\ Z \end{bmatrix}$$

and

$$\left.\frac{\partial T_B}{\partial Z}\right|_{Z=1} = \frac{0.5D^2 - 0.5D^3 + 2D^4 + 2D^5 - 4D^6 + 2D^8}{(1 - 2D - D^2 + 2D^3 - 2D^5)^2} \tag{3}$$

and

$$\left.\frac{\partial T_S}{\partial Z}\right|_{Z=1} = \frac{D^2 - 2D^3 + 4D^4 + 4D^5 - 8D^6 + 4D^8}{(1 - 2D - D^2 + 2D^3 - 2D^5)^2} \tag{4}$$

Notice in Fig. 3(b) that $m(0, j) = m(1, j)$ and $m(2, j) = m(3, j)$ for all j. In general, for any $(m_o, k_o/n_o)$ PUM code, since there are no connections from the last $(k_o - m_o)$ dummy binary cells, $m(i_a, j) = m(i_b, j)$ for any j, if the first m_o elements in s_{i_a} are identical to those in s_{i_b}. This observation is the key for reducing the number of necessary states in the SD for the TFB.

Let $M_2 = 2^{m_o} (\leqslant M_1/2)$ and $M_3 = M_1/M_2 = 2^{k_o - m_o}$ $(\geqslant 2)$. Since the number of actually working memory cells is m_o, one may naturally consider that the number of necessary states is M_2. That is, one may define the "present state" of the PUM code at time t, $\bar{S}^t$, as

$$\bar{S}^t = (x_1^{t-1}, x_2^{t-1}, \ldots, x_{m_o}^{t-1})$$

regardless of the values of $x_{m_o+1}^{t-1}, \ldots, x_{k_o}^{t-1}$, and the corresponding state space as;

$$\bar{\mathcal{S}} = \{\bar{s}_0, \bar{s}_1, \bar{s}_2, \ldots, \bar{s}_{M_2-1}\} \tag{5}$$

where

$$\bar{s}_i = (x_1, x_2, \ldots, x_{m_o})$$

with

$$i = \sum_{k=1}^{m_o} x_k \cdot 2^{m_o - k} \tag{6}$$

In other words, one may define $\bar{s}_i$ instead of defining $s_{i \cdot M_3}$, $s_{i \cdot M_3 + 1}, \ldots, s_{i \cdot M_3 + M_3 - 1}$ without distinction. The SD on this state space $\bar{\mathcal{S}}$ for any purpose has M_2 nodes and $M_3 \times M_2^2$ directed branches, where node i represents the state s_i and the M_3 directed branches from node i to node j represent the transitions from state s_i to state s_j. In Ref. 2, this state space $\bar{\mathcal{S}}$ with M_2 states was used for finding free distances of PUM codes.

However, for TFB, we can not use this $\bar{\mathcal{S}}$ directly. We may use $\bar{s}_i$'s for $i = 1,2, \ldots, M_2 - 1$, but we have to be careful about $\bar{s}_o$. That is, we have to distinguish s_i's, $i = 1,2, \ldots, M_3 - 1$ from s_o, since nonzero inputs can cause transitions with nonzero outputs from s_i, $i = 0,1, \ldots, M_3 - 1$, to s_j, $j = 1,2, \ldots, M_3 - 1$, although the contents of actually working memories are all 0's. Hence, if we represent s_o by $\bar{s}_o$, then we require another state, say $\bar{s}_{0'}$, in order to represent s_i's, $i = 1,2, \ldots, M_3 - 1$. As a result, we require the state space $\tilde{\mathcal{S}}$ of $(M_2 + 1)$ states as depicted by

$$\tilde{\mathcal{S}} = \{s_o, s_{0'}, s_1, s_2, \ldots, s_{M_2-1}\}$$

where $\bar{s}_i$'s are defined as in (5) and (6) for $i = 1,2, \ldots, M_2 - 1$, and

$$\bar{s}_0 = (0,0, \ldots, 0) \text{ with } (x_{m_o+1}, \ldots, x_{k_o}) = (0, \ldots, 0)$$

$$\bar{s}_{0'} = (0,0, \ldots, 0) \text{ with } (x_{m_o+1}, \ldots, x_{k_o}) \neq (0, \ldots, 0)$$

Now, we can define the reduced SD (RSD) for the TFB of an $(m_o, k_o/n_o)$ PUM code, which consists of $(M_2 + 1)$ nodes and $(M_2 + 1)^2$ directed branches with branch metrics defined for $i = 0,1,2, \ldots, M_2 - 1$, as

$$\bar{m}(i, 0) = m(i \cdot M_3, 0)$$

$$\bar{m}(i, 0') = \sum_{\nu=1}^{M_3-1} m(i \cdot M_3, \nu)$$

$$\bar{m}(i, j) = \sum_{\nu=0}^{M_3-1} m(i \cdot M_3, j \cdot M_3 + \nu), \quad j = 1,2, \ldots, M_2 - 1$$

and

$$\bar{m}(0', j) = \bar{m}(0, j), \quad j = 0,0',1,2, \ldots, M_2 - 1$$

And the corresponding transfer function $\bar{T}(D, Z)$ is then

$$\bar{T}(D, Z) = \bar{B} \cdot (\bar{I} - \bar{A})^{-1} \cdot \bar{C}$$

where $\bar{I}$ is the $M_2 \times M_2$ unit matrix, and for $i = 0',1,2, \ldots, M_2 - 1$, and $j = 0',1,2, \ldots, M_2 - 1$, $\bar{A}(i, j) = \bar{m}(i, j)$, $\bar{B}(j) = \bar{m}(0, j)$, and $C(i) = \bar{m}(i, 0)$.

For the (1, 2/3) PUM code in Fig. 3(a), the RSD for the TFB using the BER criterion is shown in Fig. 3(c). From this RSD we see that

$$B = \begin{bmatrix} D^2 Z^{0.5} & D^2 Z^{0.5} + D^2 Z \end{bmatrix}$$

$$C = \begin{bmatrix} 1 \\ D \end{bmatrix}, \quad A = \begin{bmatrix} D^2 Z^{0.5} & D^2 Z^{0.5} + D^2 Z \\ D^3 Z^{0.5} & D\ Z^{0.5} + D\ Z \end{bmatrix}$$

And the resulting expressions for $(\partial \overline{T}_B/\partial Z)|_{Z=1}$ and $(\partial \overline{T}_s/\partial Z)|_{Z=1}$ are exactly the same as the right-hand sides of Eqs. (3) and (4).

IV. Conclusions and Discussion

For $(m_o, k_o/n_o)$ PUM codes, the reduced SD's of $(2^{m_o} + 1)$ states for TFB's were presented. The TFB's for BER and SER criteria with BISO channel were described. TFB's with other types of channels and/or for other performance criteria may also be of interest. For example, one may want to use $M(= 2^{n_o})$-ary orthogonal input, symmetric output channel. For another example, one may want to use mean square error for the performance criterion for k_o-bit representations of certain quantities. For these cases with PUM codes, the corresponding TFB's may also be found based on the reduced SD of $(2^{m_o} + 1)$ states.

References

1. Lee, L. N., "Short, Unit-Memory, Byte-Oriented, Binary-Convolutional Codes Having Maximal Free Distance," *IEEE Trans. Info. Theory*, Vol. IT-22, pp. 349-352, May 1976.
2. Lauer, G. S., "Some Optimal Partial-Unit-Memory Codes," *Technical Report No. EE-767*, Univ. of Notre Dame, Notre Dame, Indiana, July 1976. With the same title, its shortened version was published in *IEEE Trans. Info, Theory*, Vol. IT-25, pp. 240-242, March 1979.
3. Lee, L. N., *Concatenated Coding Systems Employing Unit-Memory Convolutional Codes and Symbol-Oriented Optimal Decoding Algorithms*, Ph.D. Thesis, Dept. Elec. Eng., Univ. Notre Dame, Notre Dame, Indiana, May 1976.
4. Omura, J. K., and M. K. Simon, "Modulation/Demodulation Techniques for Satellite Communications, Part IV: Appendix," *JPL Publication 81-73*, Jet Propulsion Laboratory, Pasadena, California, Nov. 1981.
5. Viterbi, A. J., and J. K. Omura, *Principles of Digital Communications and Coding*, McGraw-Hill, New York, 1979.

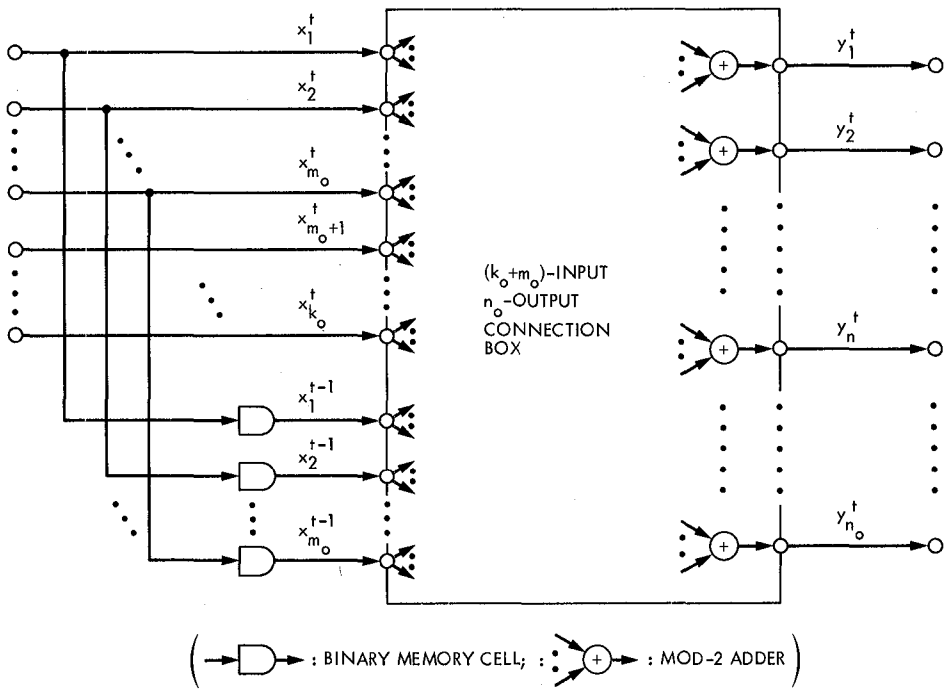


Fig. 1. Encoder structure of (m_o, k_o/n_o) UM code

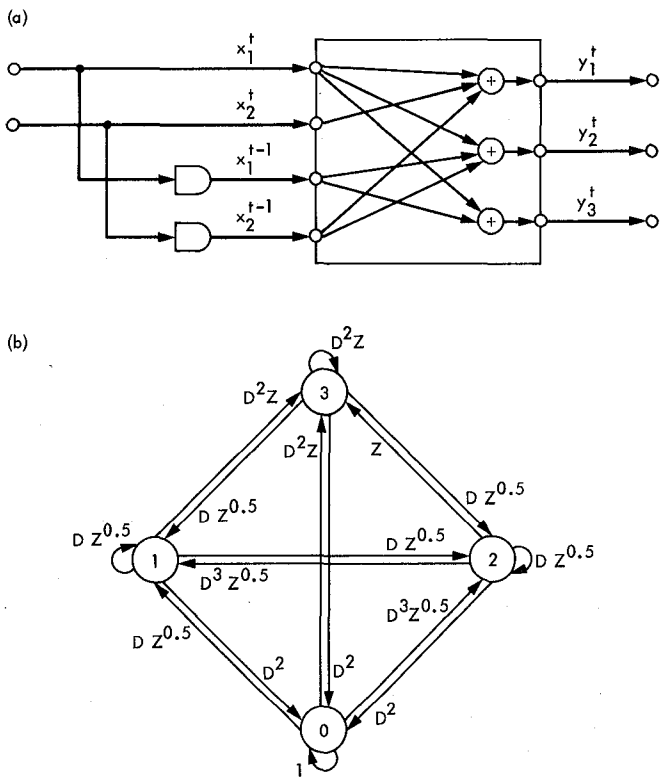


Fig. 2. Example of a (2,2/3) FUM code: (a) encoder structure of a (2,2/3) FUM code; (b) SD of the above code for BER with BISO channel

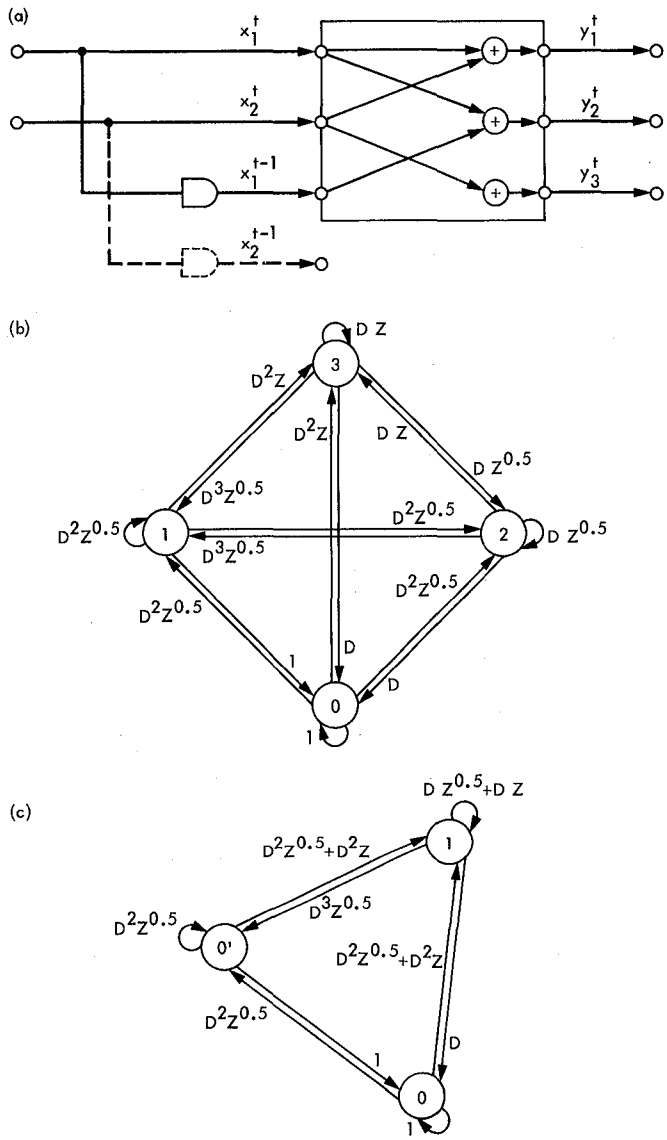


Fig. 3. Example of a (1,2/3) PUM code: (a) encoder structure of a (1,2/3) PUM code; (b) usual SD of the above code for BER with BISO channel; (c) reduced SD of the above code for BER with BISO channel

Simulations for Full Unit-Memory and Partial Unit-Memory Convolutional Codes With Real-Time Minimal-Byte-Error Probability Decoding Algorithm

Q. D. Vo
Communications Systems Research Section

This report describes a program which was written to simulate Real-Time Minimal-Byte-Error Probability (RTMBEP) decoding of full unit-memory (FUM) convolutional codes on a 3-bit quantized AWGN channel. This program was used to compute the symbol-error probability of FUM codes and to determine the signal-to-noise (SNR) required to achieve a bit error rate (BER) of 10^{-6} for corresponding concatenated systems. A (6, 6/30) FUM code, 6-bit Reed-Solomon code combination was found to achieve the required BER at a SNR of 1.886 dB. The RTMBEP algorithm was then modified for decoding partial unit-memory (PUM) convolutional codes. A simulation program was also written to simulate the symbol-error probability of these codes.

I. Introduction

To achieve reliable communications over a very noisy channel with relatively small coding complexity, concatenated coding systems using a convolutional code as the inner code and a Reed-Solomon code as the outer code are usually suggested (Fig. 1). In our baseline system, the inner code is a (7, 1/2) convolutional code with Viterbi decoding (Ref. 1) while the outer code is an 8-bit Reed-Solomon code. The overall system achieves a bit error rate (BER) of 10^{-6} with a signal-to-noise ratio (SNR) of 2.53 dB. Our goal is to obtain a 2-dB improvement in SNR (i.e., to find an inner code with a corresponding decoding algorithm to achieve a BER of 10^{-6} at a SNR of 0.53 dB). It was found by Lin-nan Lee (Refs. 2 and 3) that the use of byte-oriented full unit-memory (FUM) convolutional codes in conjunction with real-time minimal-byte-error probability (RTMBEP) decoding algorithm provides an improvement of about 0.3 dB compared to regular bit-oriented convolutional codes with same state complexity. In this report, we simulated the RTMBEP decoding of FUM codes on a 3-bit quantized additive white Gaussian noise (AWGN) channel (as shown by the dashed box in Fig. 1). This simulation program is used as a tool to compute symbol-error probability for FUM codes. We then modified the RTMBEP algorithm for decoding a subclass of FUM codes called partial-unit-memory (PUM) codes (Ref. 4). A simulation program was also written for this modified algorithm to simulate symbol-error probability of PUM codes.

II. Simulation of FUM Codes With RTMBEP Decoding Algorithm

A general (*lo, ko/no*) unit-memory convolutional encoder is shown in Fig. 2. Here, we have *ko* bits of input to be encoded, *lo* bits of delayed input, and *no* bits of encoded output. For $lo = ko$ we have FUM convolutional codes. For $lo < ko$ we have PUM convolutional codes.

A (*ko, ko/no*) FUM code may be written as

$$\mathbf{b}_t = \mathbf{a}_t G_o + \mathbf{a}_{t-1} G_1; \quad t = 1, 2, \ldots$$

where $\mathbf{a}_t$ is the k_o-bit byte of information to be encoded at time t, $\mathbf{b}_t$ is the corresponding encoded no-bit byte, and G_o and G_1 are $ko \times no$ encoding matrices (by convention, $\mathbf{a}_o = \mathbf{0}$). The sequence $\{\mathbf{b}_t\}$ is sent over a 3-bit quantized AWGN channel and the corresponding sequence $\mathbf{r}_t$ is received. For convenience, we denote $\mathbf{a}_{[t,t^1]}$ to be $[\mathbf{a}_t, \mathbf{a}_{t+1}, \ldots, \mathbf{a}_{t^1}]$ and similarly for $\mathbf{b}_{[t,\,t^1]}, \mathbf{r}_{[t,\,t^1]}$.

The RTMBEP decoding rule (Ref. 3) chooses its estimate $\hat{\mathbf{a}}_t$ to be the value of $\mathbf{a}_t$ which maximizes $P(\mathbf{a}_t|\mathbf{r}_{[1,t+\Delta]})$ where $\mathbf{r}_{[1,\,t+\Delta]}$ is the observed sequence with delay Δ. This algorithm is based on the facts that the channel is memoryless and that the code has unit memory to develop a recursive method for computing $P(\mathbf{a}_t|\mathbf{r}_{[1,\,t+\Delta]})$

$$P(\mathbf{a}_t|\mathbf{r}_{[1,t+\Delta]}) = \frac{P(\mathbf{a}_t, \mathbf{r}_{[1,t+\Delta]})}{\sum_{\mathbf{a}_t} P(\mathbf{a}_t, \mathbf{r}_{[1,t+\Delta]})}$$

$$P(\mathbf{a}_t, \mathbf{r}_{[1,t+\Delta]}) = P(\mathbf{a}_t, \mathbf{r}_{[1,t]})\, P(\mathbf{r}_{[t+1,t+\Delta]}|\mathbf{a}_t)$$

$$P(\mathbf{a}_t, \mathbf{r}_{[1,t]}) = \sum_{\mathbf{a}_{t-1}} P(\mathbf{a}_{t-1}, \mathbf{r}_{[1,t-1]})$$

$$\times P(\mathbf{a}_t, \mathbf{r}_t|\mathbf{a}_{t-1}, \mathbf{r}_{[1,t-1]})$$

$$= 2^{-ko} \sum_{\mathbf{a}_{t-1}} P(\mathbf{a}_{t-1}, \mathbf{r}_{[1,t-1]})\, P(\mathbf{r}_t|\mathbf{b}_t)$$

$$P(\mathbf{r}_{[t+i,t+\Delta]}|\mathbf{a}_{t+i-1}) = \sum_{\mathbf{a}_{t+1}} P(\mathbf{r}_{[t+i+1,t+\Delta]}|\mathbf{a}_{t+i})$$

$$\times P(\mathbf{a}_{t+i}, \mathbf{r}_{t+i}|\mathbf{a}_{t+i-1})$$

$$= 2^{-ko} \sum_{\mathbf{a}_{t+i}} P(\mathbf{r}_{[t+i+1,t+\Delta]}|\mathbf{a}_{t+i})$$

$$\times P(\mathbf{r}_{t+i}|\mathbf{b}_{t+i})$$

The last recursion is initialized with $i = \Delta$

$$P(\mathbf{r}_{t+\Delta}|\mathbf{a}_{t+\Delta-1}) = \sum_{\mathbf{a}_{t+\Delta}} P(\mathbf{r}_{t+\Delta}, \mathbf{a}_{t+\Delta}|\mathbf{a}_{t+\Delta-1})$$

$$= 2^{-ko} \sum_{\mathbf{a}_{t+\Delta}} P(\mathbf{r}_{t+\Delta}|\mathbf{b}_{t+\Delta})$$

and computed backward with $i = \Delta - 1, \ldots, 1$. Here, we assume all information sequences are equally likely (i.e., $P(\mathbf{a}_t) = 2^{-ko}$). We modified the algorithm slightly so that it is computer compatible. Let

$$f(\mathbf{a}_t) = P(\mathbf{a}_t, \mathbf{r}_{[1,t]})$$

$$h(\mathbf{a}_{t+i-1}) = P(\mathbf{r}_{[t+i,\,t+\Delta]}|\mathbf{a}_{t+i-1})$$

At time $t = 0$ (initialization) the whole system is set up as follows:

(1) Set up coder matrix which gives $\mathbf{b}_t$ for each $\mathbf{a}_t$ and $\mathbf{a}_{t-1}$

(2) Simulate $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_\Delta$ on a 3-bit quantized AWGN channel and compute probability matrices $P(\mathbf{r}_i|\mathbf{b}_i)$

(3) Set $f(\mathbf{a}_o = \mathbf{0}) = 1$ and $f(\mathbf{a}_o \neq \mathbf{0}) = 0$

For each time t (main loop; $t = 1, 2, \ldots$) the following steps are taken:

Step 0: Simulate $\mathbf{r}_{t+\Delta}$ and compute $P(\mathbf{r}_{t+\Delta}|\mathbf{b}_{t+\Delta})$ matrix

Step 1: Compute

$$f(\mathbf{a}_t) = \sum_{\mathbf{a}_{t-1}} f(\mathbf{a}_{t-1})\, P(\mathbf{r}_t|\mathbf{b}_t)$$

and normalize

$$f(\mathbf{a}_t) = f(\mathbf{a}_t)/f(\mathbf{0})$$

Note that the normalization is needed since, for large t, $f(\mathbf{a}_t)$ tends to go to zero.

Step 2: Set

$$h(\mathbf{a}_{t+\Delta-1}) = \sum_{\mathbf{a}_{t+\Delta}} P(\mathbf{r}_{t+\Delta} \mid b_{t+\Delta})$$

and normalize

$$h(\mathbf{a}_{t+\Delta-1}) = \frac{h(\mathbf{a}_{t+\Delta-1})}{h(\mathbf{0})}$$

Step 3: For $i = \Delta - 1$ to $i = 1$, compute

$$h(\mathbf{a}_{t+i-1}) = \sum_{\mathbf{a}_{t+i}} h(\mathbf{a}_{t+i}) P(\mathbf{r}_{t+i} \mid \mathbf{b}_{t+i})$$

and normalize

$$h(\mathbf{a}_{t+i-1}) = \frac{h(\mathbf{a}_{t+i-1})}{h(\mathbf{0})}$$

Step 4: Choose the estimate $\hat{\mathbf{a}}_t$ such that

$$f(\hat{\mathbf{a}}_t)\, h(\hat{\mathbf{a}}_t) \geq f(\mathbf{a}_t)\, h(\mathbf{a}_t) \quad ; \quad \text{for all } \mathbf{a}_t\text{'s}$$

and compute the probability

$$P(\mathbf{a}_t = \hat{\mathbf{a}}_t \mid \mathbf{r}_{[1,t+\Delta]}) = \frac{f(\hat{\mathbf{a}}_t)\, h(\hat{\mathbf{a}}_t)}{\sum_{\mathbf{a}_t} f(\mathbf{a}_t)\, h(\mathbf{a}_t)}$$

Note that this probability could be used as a reliability indicator for soft decision (including erasure) decoding of the convolutional code. Since large Δ requires more computation, it is desirable to keep Δ as small as possible. It is shown by Lee (Ref. 3) that $\Delta = 8$ gives virtually the same performance as $\Delta = \infty$. We checked the simulation with several small FUM convolutional codes and ran it for the maximal d_{free} (6, 6/18), (6, 6/24) codes found by Lin-nan Lee (Ref. 2) and a (6, 6/30) code found by Pil J. Lee. These codes are given in Table 1 in hexadecimal format. Simulation results are given in Table 2 for different levels of signal-to-noise ratio, E_b'/No (E_b' is the input bit energy at the inner encoder and No is the one-sided noise power spectral density.) The symbol (byte) error probability P_s, which is calculated based on 8000-byte decoding simulation, is plotted versus E_b'/No in Fig. 3 for the three FUM codes. These FUM codes are concatenated with various 6-bit rate k/n ($n = 2^6 - 1 = 63$) Reed-Solomon codes. Given that the interleaving is perfect, the overall bit-error probability can be computed as

$$P_b = \frac{n+1}{2n} \sum_{i=t+1}^{n} \frac{i}{n} \binom{n}{i} p^i (1-p)^{n-i}$$

where t is the error correcting capability ($t = n - k/2$) and p is the symbol-error probability at the inner decoder output. The required p to achieve a BER of 10^{-6} is shown in Table 3 for several 6-bit Reed-Solomon codes. From Fig. 2, the requirement in p specifies a requirement in E_b'/No and E_b/No ($E_b/No = nE_b'/kNo$) as also shown in Table 3. In Fig. 4, we plot the signal-to-noise ratio (E_b/No) required to achieve a BER of 10^{-6} versus the Reed-Solomon code rate for the three FUM codes. An improvement of 0.64 dB is obtained with the (6, 6/30) code compared to the baseline (7, 1/2) convolutional code, Reed-Solomon code combination. To achieve a 2 dB improvement in required E_b/No, FUM codes with higher state space complexity will be required. Unfortunately, it takes a long time to search for and to simulate such codes. For example, it takes 16 hours to simulate one point for a (7, 7/28) FUM code on a fast computer. Therefore, we consider next a subclass of FUM codes called partial-unit-memory (PUM) codes (Ref. 4) which may provide similar performance with less complexity.

III. Simulation of PUM Codes With Modified RTMBEP Decoding Algorithm

Let $\mathbf{a}_t = \tilde{\mathbf{a}}_t : \hat{\mathbf{a}}_t$ (the colon denotes concatenation of $k_o - l_o$-bit byte $\tilde{\mathbf{a}}_t$ with l_o-bit byte $\hat{\mathbf{a}}_t$) be the k_o-bit byte of information to be encoded at time t and $\mathbf{b}_t$ be the corresponding encoded n_o-bit byte. A $(l_o, k_o/n_o)$ PUM code may be written as (see Fig. 2)

$$\mathbf{b}_t = \mathbf{a}_t G_o + \hat{\mathbf{a}}_{t-1} G_1; \quad t = 1, 2, 3, \ldots$$

where G_0 and G_1 are encoding matrices with dimensions $k_o \times n_o$ and $l_o \times n_o$, respectively. The state complexity is defined to be l_o, which is the number of delay cells required to realize the encoder. Let $\tilde{G}_0$ be the first $k_o - l_o$ rows of G_0 and $\hat{G}_0$ be the remaining l_o rows, then

$$\tilde{\mathbf{b}}_t = \tilde{\mathbf{a}}_t G_o + \hat{\mathbf{a}}_t \hat{G}_o + \hat{\mathbf{a}}_{t-1} G_1$$

We want to find the estimate $\mathbf{a}_t^o$ which maximizes $P(\mathbf{a}_t \mid \mathbf{r}_{[1,\, t+\Delta]})$ where $\mathbf{r}_{[1,\, t+\Delta]}$ is the received sequence with delay Δ. As in Section II, we have

$$P(\mathbf{a}_t \mid \mathbf{r}_{[1,t+\Delta]}) = \frac{P(\mathbf{a}_t, r_{[1,t+\Delta]})}{\sum_{\alpha} P(\alpha, r_{[1,t+\Delta]})}$$

$$P(\mathbf{a}_t, \mathbf{r}_{[1,t+\Delta]}) = P(\mathbf{a}_t, \mathbf{r}_{[1,t]}) P(\mathbf{r}_{[t+1,t+\Delta]} \mid \mathbf{a}_t)$$

$$P(\mathbf{a}_t, \mathbf{r}_{[1,t]}) = 2^{-k_o} \sum_{\mathbf{a}_{t-1}} P(\mathbf{a}_{t-1}, \mathbf{r}_{[1,t-1]}) P(\mathbf{r}_t \mid \mathbf{b}_t)$$

However, taking into account that $\mathbf{b}_t$ just depends on $\hat{\mathbf{a}}_{t-1}$, we can write

$$P(\mathbf{a}_t, \mathbf{r}_{[1,t]}) = 2^{-k_o} \sum_{\hat{\mathbf{a}}_{t-1}} P(\mathbf{r}_t \mid \mathbf{b}_t) \sum_{\tilde{\mathbf{a}}_{t-1}} P(\mathbf{a}_{t-1}, \mathbf{r}_{[1,t-1]})$$

$$= 2^{-k_o} \sum_{\hat{\mathbf{a}}_{t-1}} P(\mathbf{r}_t \mid \mathbf{b}_t) P(\hat{\mathbf{a}}_{t-1}, \mathbf{r}_{1,t-1]})$$

where

$$P(\hat{\mathbf{a}}_{t-1}, \mathbf{r}_{[1,t-1]}) = \sum_{\tilde{\mathbf{a}}_{t-1}} P(\mathbf{a}_{t-1}, \mathbf{r}_{[1,t-1]})$$

could be computed outside of the $\mathbf{a}_t$ loop. Since the dimension of $\hat{\mathbf{a}}_{t-1}$ (i.e., l_o) is less than the dimension of $\mathbf{a}_{t-1}$ (i.e., k_o), this results in a saving of computation time. For example, with $k_o = 6$, $l_o = 6$ (regular FUM codes) it requires 4096 multiplications compared to 2048 multiplications for the case $k_o = 6$, $l_o = 5$ or 1024 multiplications for the case $k_o = 6$, $l_o = 4$.

On the other hand,

$$P(\mathbf{r}_{[t+1,t+\Delta]} \mid \mathbf{a}_t) = P(\mathbf{r}_{[t+1,t+\Delta]} \mid \tilde{\mathbf{a}}_t, \hat{\mathbf{a}}_t)$$

$$= P(\mathbf{r}_{[t+1,t+\Delta]} \mid \hat{\mathbf{a}}_t)$$

since $\mathbf{r}_{[t+1,t+\Delta]}$ is independent of $\tilde{\mathbf{a}}_t$. A backward recursion is developed to compute $P(\mathbf{r}_{[t+1,t+\Delta]} \mid \hat{\mathbf{a}}_t)$:

$$P(\mathbf{r}_{[t+i,t+\Delta]} \mid \hat{\mathbf{a}}_{t+i-1}) = \sum_{\mathbf{a}_{t+i}} P(\mathbf{a}_{t+i}, \mathbf{r}_{[t+i,t+\Delta]} \mid \hat{\mathbf{a}}_{t+i-1})$$

$$= \sum_{\mathbf{a}_{t+i}} P(\mathbf{r}_{[t+i+1,t+\Delta]} \mid \hat{\mathbf{a}}_{t+i}) \times P(\mathbf{a}_{t+i}, \mathbf{r}_{t+i} \mid \hat{\mathbf{a}}_{t+i-1})$$

$$= \sum_{\mathbf{a}_{t+i}} P(\mathbf{r}_{[t+i-1,t+\Delta]} \mid \hat{\mathbf{a}}_{t+i}) \times P(\mathbf{r}_{t+i} \mid \mathbf{a}_{t+i}, \hat{\mathbf{a}}_{t+i-1}) P(\mathbf{a}_{t+i})$$

$$= 2^{-k_o} \sum_{\mathbf{a}_{t+i}} P(\mathbf{r}_{[t+i-1,t+\Delta]} \mid \hat{\mathbf{a}}_{t+i}) \times P(\mathbf{r}_{t+i} \mid \mathbf{b}_{t+i})$$

$$= 2^{-k_o} \sum_{\hat{\mathbf{a}}_{t+i}} P(\mathbf{r}_{[t+i-1,t+\Delta]} \mid \hat{\mathbf{a}}_{t+i}) \times \sum_{\tilde{\mathbf{a}}_{t+1}} P(\mathbf{r}_{t+1} \mid \mathbf{b}_{t+1})$$

This recursion is initialized with $i = \Delta$:

$$P(\mathbf{r}_{t+\Delta} \mid \hat{\mathbf{a}}_{t+\Delta-1}) = 2^{-k_o} \sum_{\mathbf{a}_{t+\Delta}} P(\mathbf{r}_{t+\Delta} \mid \mathbf{b}_{t+\Delta})$$

For $k_o = 6$, $l_o = 5$ this recusion requires 1048 multiplications compared to 4096 multiplications of the regular FUM codes ($k_o = 6$, $l_o = 6$). Let

$$f(\mathbf{a}_t) = P(\mathbf{a}_t, \mathbf{r}_{[1,t]})$$

$$h(\hat{\mathbf{a}}_{t+i-1}) = P(\mathbf{r}_{[t+i,t+\Delta]} \mid \hat{\mathbf{a}}_{t+i-1})$$

Then the modified algorithm for PUM codes is summarized as follows: At time $t = 0$, the system is initalized as follows:

(1) Set up coder matrix which gives $\mathbf{b}_t$ for each $\mathbf{a}_t$ and $\hat{\mathbf{a}}_{t-1}$

(2) Simulate $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_\Delta$ on a 3-bit quantized AWGN channel and compute probability matrices $P(\mathbf{r}_i|\mathbf{b}_i)$

(3) Set $f(\mathbf{a}_o = \mathbf{0}) = 1$ and $f(\mathbf{a}_o \neq \mathbf{0}) = 0$

At time t ($t = 1, 2, \ldots$) the following steps are taken:

Step 0: Simulate $\mathbf{r}_{t+\Delta}$ and compute $P(\mathbf{r}_{t+\Delta} \mid \mathbf{b}_{t+\Delta})$ matrix.

Step 1: Compute

$$f(\hat{\mathbf{a}}_{t-1}) = \sum_{\tilde{\mathbf{a}}_{t-1}} f(\mathbf{a}_{t-1})$$

For each $\mathbf{a}_t$, compute

$$f(\mathbf{a}_t) = \sum_{\hat{\mathbf{a}}_{t-1}} f(\hat{\mathbf{a}}_{t-1}) P(\mathbf{r}_t \mid \mathbf{b}_t)$$

$$f(\mathbf{a}_t) = f(\mathbf{a}_t)/f(\mathbf{0})$$

Step 2: Set

$$h(\hat{\mathbf{a}}_{t+\Delta-1}) = \sum_{\mathbf{a}_{t+\Delta}} P(\mathbf{r}_{t+\Delta} \mid \mathbf{b}_{t+\Delta})$$

$$h(\hat{\mathbf{a}}_{t+\Delta-1}) = \frac{h(\hat{\mathbf{a}}_{t+\Delta-1})}{h(\mathbf{0})}$$

Step 3: For $i = \Delta - 1$ to $i = 1$, compute for each $\hat{\mathbf{a}}_{t+i-1}$

$$h(\hat{\mathbf{a}}_{t+i-1}) = \sum_{\hat{\mathbf{a}}_{t+i}} h(\hat{\mathbf{a}}_{t+i}) \sum_{\tilde{\mathbf{a}}_{t+i}} P(\mathbf{r}_{t+i} \mid \mathbf{b}_{t+i})$$

$$h(\hat{\mathbf{a}}_{t+i-1}) = \frac{h(\hat{\mathbf{a}}_{t+i-1})}{h(\mathbf{0})}$$

Step 4: Choose the estimate $\mathbf{a}_t^o$ such that (recall that $h(\mathbf{a}_t) = h(\hat{\mathbf{a}}_t)$)

$$f(\mathbf{a}_t^o)\, h(\mathbf{a}_t^o) \geq f(\mathbf{a}_t)\, h(\mathbf{a}_t)$$

for all $\mathbf{a}_t$'s and compute as a reliability indicator

$$P(\mathbf{a}_t = \mathbf{a}_t^o \mid \mathbf{r}_{[1,t+\Delta]}) = \frac{f(\mathbf{a}_t^o)\, h(\mathbf{a}_t^o)}{\sum_{\mathbf{a}_t} f(\mathbf{a}_t)\, h(\mathbf{a}_t)}$$

Again, this reliability indicator could be used for soft decision (including erasure) decoding of the convolutional code.

We checked the simulation with some small PUM codes given in Table 4. The simulated symbol-error probability for each code is also given in Table 4. Good PUM codes with higher state complexity are needed to achieve the required performance.

IV. Summary

A program is written to simulate the RTMBEP decoding of FUM codes on a 3-bit quantized AWGN channel. This program is used to compute the symbol-error probability for various FUM codes. From this, the SNR required to achieve a BER of 10^{-6} can be determined for a concatenated system. A 6, 6/30) FUM code is found to achieve the required BER at a SNR of 1.886 dB. Finally the RTMBEP algorithm is modified for decoding a subclass of FUM codes called PUM codes. A simulation program is also written to simulate symbol-error probability of PUM codes.

References

1. Viterbi, A. J., and Omura, J. K., *Principles of Digital Communication and Coding*, McGraw-Hill, N.Y., 1979.
2. Lee, Lin-nan, "Short Unit-Memory Byte-Oriented Binary Convolutional Codes Having Maximal Free Distance," *IEEE Transactions on Information Theory*, Vol. IT-22, pp. 349-352, May 1976.
3. Lee, Lin-nan, *Concatenated Coding Systems Employing Unit-Memory Convolutional Codes and Symbol-Oriented Optimal Decoding Algorithms*, Ph.D. Dissertation, Department of Electrical Engineering, University of Notre Dame, IN., May 1976.
4. Lauer, G. S., "Some Optimal Partial Unit Memory Codes," *IEEE Transactions on Information Theory*, Vol. IT-25, pp. 240-243, March 1979.

Table 1. Some UM convolutional codes

Rate	no	ko	G_0	G_1	d_{Free}
1/3	18	6	38D30	031CB	16
			1C698	06396	
			0E34C	0C72C	
			07986	18E19	
			234C3	30C72	
			31A61	218E5	
1/4	24	6	83EEB0	3C52D3	24
			41F758	78A5A6	
			22FB8C	F14B0D	
			1375C6	E6865A	
			0BBAC3	CD1CB4	
			07DD61	9E2969	
1/5	30	6	20FBAC1C	0F14B4C1	29
			107DD60E	1E296982	
			08BEE307	3C52C344	
			04DD71A3	39A19688	
			02EEB0F1	33472D10	
			01F75878	278A5A60	

Table 2. Symbol-error probability for codes in Table 1

E'_b/No (dB)	0.50	0.75	1.00	1.25	1.50
(6,6/18) P_S		0.05925	0.0355	0.02125	0.01275
(6,6/24) P_S	0.060875	0.03775	0.02275	0.01225	0.00675
(6,6/30) P_S		0.024625	0.0125	0.007875	0.005

Table 3. Signal-to-noise ratio requirement to achieve a BER of 10^{-6}

k	45	47	49	51	53	55	57
k/n	0.7143	0.7460	0.7778	0.8095	0.8413	0.8730	0.9048
(dB)	-1.461	-1.272	-1.091	-0.918	-0.751	-0.590	-0.435
P	0.02875	0.02294	0.01764	0.01290	0.00879	0.00538	0.00277
(6,6/18)							
E'_b/No (dB)	1.104	1.212	1.342	1.494	1.681	1.920	
E_b/No (dB)	2.565	2.484	2.433	2.412	2.432	2.510	
(6,6/24)							
E'_b/No (dB)	0.887	0.995	1.109	1.233	1.388	1.587	
E_b/No (dB)	2.348	2.267	2.200	2.151	2.139	2.177	
(6,6/30)							
E'_b/No (dB)		0.758	0.846	0.968	1.169	1.392	
E_b/No (dB)		2.030	1.937	1.886	1.920	1.982	

Table 4. Simulated symbol-error probability for some PUM codes

Codes	G_0	G_1	E'_b/No (dB) and P_S				
(1,2/4)	F	6	2 dB		4 dB		6 dB
	3		0.04036		0.00665		0.00039
(3,4/8)	FF	17	2.5 dB	3.0 dB	3.5 dB	4.0 dB	4.5 dB
	OF	2D	0.00946	0.00376	0.00125	0.00034	0.00005
	33	8B					
	55						

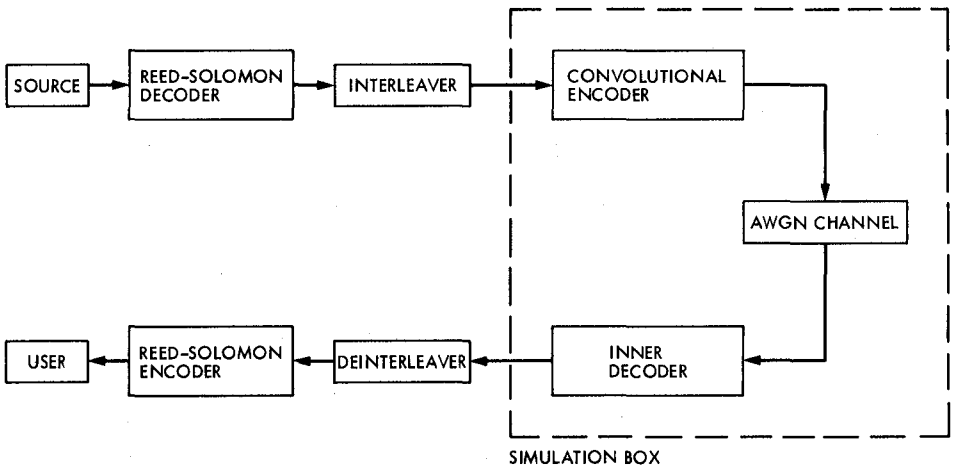


Fig. 1. A concatenated coding system

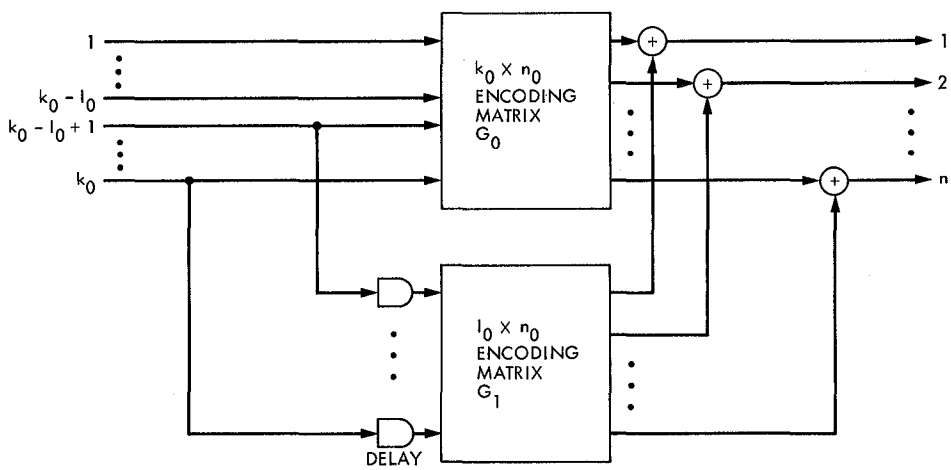


Fig. 2. A general (l_o, *ko/no*) unit-memory convolutional encoder

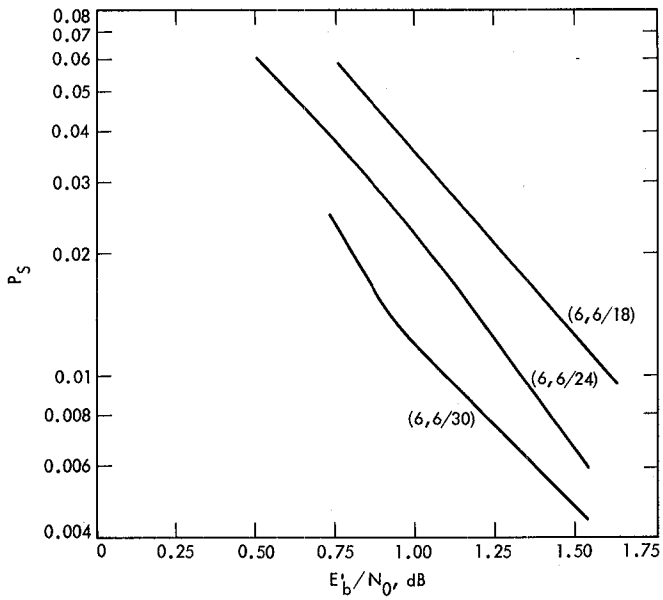


Fig. 3. Symbol error probability for three FUM codes

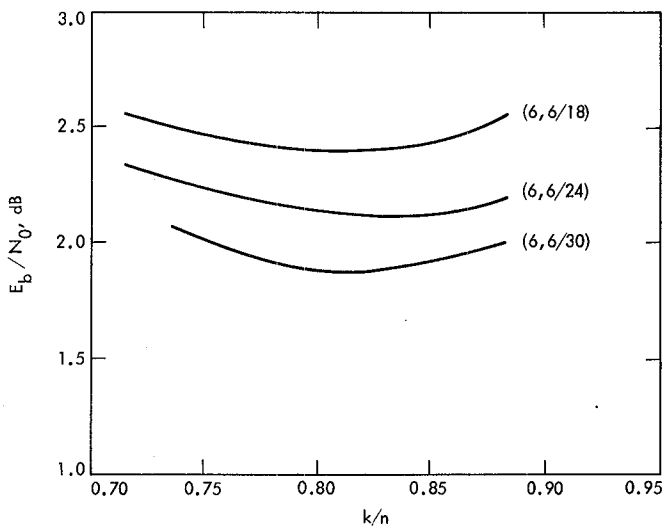


Fig. 4. Required E_b/No to achieve a BER of 10^{-6}

Signal-to-Noise Ratio and Combiner Weight Estimation for Symbol Stream Combining

Q. D. Vo
Communications Systems Research Section

A method is presented for signal-to-noise ratio (SNR) and symbol stream combiner weight estimation. The SNR estimator employs absolute value moments as in an earlier method. The main contribution of this work is that a new algorithm is derived for the combiner weight estimator to remove the large bias at low SNRs. The new algorithm is simulated to combine two independent symbol streams at various SNRs. As an example, in combining two symbol streams at SNRs of −1 dB and −7 dB, combiner weight estimates using 1000 samples for the −1 dB stream and 10,000 samples for the −7 dB stream achieve an output SNR of −0.039 dB, which is just 0.012 dB below the theoretical limit achievable with perfect knowledge of the SNRs.

I. Introduction

Signals received from interplanetary telecommunications are usually very weak. An improvement in signal-to-noise ratio (SNR) can be obtained by properly combining the outputs from several receiving stations. The optimal combining strategy depends on the SNR levels of the received signals. In this report, we consider a technique of SNR and combiner weight estimation for possible use in symbol stream combining. This is as opposed to baseband combining wherein combining is done prior to subcarrier demodulation and symbol synchronization. The SNR estimators, both biased and unbiased, are considered by Layland (Ref. 1). We follow the same procedure to define the biased and unbiased combiner weight estimators. The SNR loss of the combined output stream is analyzed and compared for both biased and unbiased combiner weight estimates. The combining process has negligible loss using either method for two symbol streams with the same SNRs. However, with the biased method, the SNR loss grows significantly as the difference in the two SNRs gets larger due to large bias introduced by the use of absolute moment as an approximation to the mean at low SNRs. For the unbiased method, simulation results show that the bias is practically removed with a sufficiently large number of samples. For example, with 1000 samples for the −1 dB stream and 10,000 samples for the −7 dB stream, the obtained output SNR is just 0.012 dB less than the maximum achievable limit (−0.027 dB) which assumes perfect knowledge of the input SNRs.

II. Background

As shown in Fig. 1, a spacecraft signal is received simultaneously at two tracking stations. This signal is demodulated and integrated to produce two sequences of symbols, $\{x_k\}$ and $\{y_k\}$. Due to differences in receiving systems, the two symbol streams have different levels of SNR. The symbol stream combiner is designed to maximize the overall SNR. Let

$$E(x_k) = \pm m_1, \quad \text{var}(x_k) = \sigma_1^2$$

and

$$E(y_k) = \pm m_2, \quad \text{var}(y_k) = \sigma_2^2$$

Then

$$\text{SNR}_{x_k} = \frac{m_1^2}{2\sigma_1^2}, \quad \text{SNR}_{y_k} = \frac{m_2^2}{2\sigma_2^2}$$

(The factor of two in the denominator is used so that the definition of SNR is the same as symbol energy to noise spectral density.) The combining strategy is

$$z_k = \alpha_1 x_k + \alpha_2 y_k$$

where α_1 and α_2 are chosen to maximize

$$\text{SNR}_{z_k} = \frac{(\alpha_1 m_1 + \alpha_2 m_2)^2}{2\alpha_1^2 \sigma_1^2 + 2\alpha_2^2 \sigma_2^2} \leqslant \text{SNR}_{x_k} + \text{SNR}_{y_k}$$

Taking derivatives with respect to α_1, α_2 and setting them to zero, it is found that the output SNR is maximized when

$$\frac{\alpha_1}{\alpha_2} = \frac{m_1}{\sigma_1^2}\frac{\sigma_2^2}{m_2} = \frac{\sqrt{\text{SNR}_{x_k}}}{\sqrt{\text{SNR}_{y_k}}} \times \frac{\sigma_2}{\sigma_1}$$

Since m_k and σ_k^2 (k = 1,2) are not known we must have some way to estimate them. The conventional method which deals with absolute moment is considered first, and then improvement by bias removal is investigated.

III. Conventional SNR Estimator

Let the input signal $x(t)$ be either $+V$ or $-V$ in the intervals t_k to t_{k+1}, $k = 1,2,\ldots$. This signal is corrupted by additive white Gaussian noise $n(t)$ having zero mean and two-sided spectral density $N_0/2$. Then received signal is integrated over the symbol time to produce the symbol stream $\{x_k\}$,

$$x_k = \pm VT + n_k$$

where n_k is a zero-mean Gaussian random variable with variance $N_0T/2$. Thus,

$$m = E(x_k) = \pm VT, \quad \sigma^2 = \text{Var}(x_k) = \frac{N_0 T}{2}$$

and

$$R = \text{SNR}_{x_k} = \frac{m^2}{2\sigma^2} = \frac{V^2 T}{N_0}$$

The conventional SNR estimator uses the sample mean and variance in conjunction with absolute moment to estimate m and σ^2

$$\widetilde{m} = \frac{1}{N}\sum_{k=1}^{N} |x_k|$$

$$\widetilde{\sigma}^2 = \frac{1}{N-1}\sum_{k=1}^{N}(|x_k| - \widetilde{m})^2 = \frac{N}{N-1}\left[\frac{1}{N}\sum_{k=1}^{N} x_k^2 - \widetilde{m}^2\right]$$

So the estimate $\widetilde{R}$ for the SNR is

$$\widetilde{R} = \frac{\left(\frac{1}{N}\sum_{k=1}^{N}|x_k|\right)^2}{\frac{2}{N-1}\sum_{k=1}^{N}\left(|x_k| - \frac{1}{N}\sum_{i=1}^{N}|x_i|\right)^2}$$

The performance of this estimator was analyzed by Layland (Ref. 1), and then by Lesh (Ref. 2). The mean of the estimator is found to be dependent heavily on the input noise but relatively insensitive to changes in the sample size. For large enough $N(N>500)$ we have

$$E(\widetilde{R}) \simeq \frac{[E(|x_k|)]^2}{2\,\text{Var}(|x_k|)} = f(R)$$

where

$$E(|x_k|) = \sqrt{N_0 T}\left[\frac{\exp^{-R}}{\sqrt{\pi}} + \sqrt{R}\,\text{erf}(\sqrt{R})\right]$$

$$\text{Var}(|x_k|) = E(x_k^2) - E(|x_k|)^2$$

$$E(x_k^2) = N_0 T(R + \tfrac{1}{2})$$

and

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp^{-t^2} dt$$

The estimator mean is plotted in Fig. 2 over the interested range of input SNRs. We see that there is a significant bias at low SNRs due to the nonlinearity of the absolute moment. This problem can be solved by defining another estimator (Ref. 1) $\hat{R} = f^{-1}(\tilde{R})$; then for large N we have

$$E(\hat{R}) \simeq R$$

and

$$\text{Var}(\hat{R}) \simeq \text{Var}(\tilde{R}) \cdot \left[\left. \frac{df(x)}{dx} \right|_{x=R} \right]^{-2}$$

In real time implementation, the inverse of the function f could be evaluated using table lookup.

The variance of $\tilde{R}$ is found to be strongly dependent on the sample size and relatively insensitive with respect to noise. A computer program is written to evaluate this variance based on equations provided by Lesh (Ref. 2). Figure 3 shows the dependence of the standard deviation on the sample size for $R = -1$ dB and $R = -7$ dB. To be consistent with the previous paper (Ref. 2), the standard deviation in dB is defined as

$$\sigma_{\text{dB}}(\tilde{R}) = 10 \log_{10} \left(1 + \frac{\sigma(\tilde{R})}{R} \right)$$

In Fig. 4 we plot the number of symbols (i.e., sample size) required to achieve a standard deviation of 0.1 dB versus input SNRs.

IV. Implementation for the Symbol Stream Combiner

The combining strategy is

$$z_k = \alpha_1 x_k + \alpha_2 y_k$$

Before we consider the algorithms for estimating α's, we would like to see how sensitive the output SNR is with respect to these coefficients. Since the optimum weights result in an SNR equal to the sum of the input SNRs, the SNR loss is

$$\text{SNR loss (dB)} = 10 \log_{10}(R_1 + R_2) - 10 \log_{10} R$$

where

$$R_1 = \text{SNR}_{x_k}$$

$$R_2 = \text{SNR}_{y_k}$$

$$R = \text{SNR}_{z_k}$$

Then

$$E(z_k) = \alpha_1 E(x_k) + \alpha_2 E(y_k) = \alpha_1 m_1 + \alpha_2 m_2$$

$$\text{Var}(z_k) = \alpha_1^2 \text{Var}(x_k) + \alpha_2^2 \text{Var}(y_k) = \alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2$$

So the output SNR will be

$$R = \text{SNR}_{z_k} = \frac{(\alpha_1 m_1 + \alpha_2 m_2)^2}{2\left[\alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2\right]}$$

Figure 5 shows the SNR loss as a function of R_2 and α_2 for $R_1 = -1$ dB and $\alpha_1 = \sqrt{2R_1}$ (we consider a normalized case where $\sigma_1^2 = \sigma_2^2 = 1$). We see that highly accurate estimates of α's (hence of input SNRs) are unnecessary since the SNR loss does not depend very critically on these coefficients around their optimal values. (For example with a change of 0.1 from the optimal value of α_2, the SNR loss increases by less than 0.02 dB). However, if large biases are present, the SNR loss grows very fast and becomes very sensitive to changes in α's. So we need to derive an algorithm with negligible bias. We first consider the biased combining method and then try to remove the bias with appropriate adjustment.

A. Biased Combining Algorithm

The biased estimates of α's are just the sample means divided by the sample variances:

$$\tilde{\alpha}_1 = \frac{\frac{1}{N}\sum_{i=1}^{N} |x_i|}{\frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} |x_i|\right)^2\right]}$$

$$\tilde{\alpha}_2 = \frac{\frac{1}{N}\sum_{i=1}^{N} |y_i|}{\frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N} y_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} |y_i|\right)^2\right]}$$

Since the optimal strategy just depends on the ratio between α_1 and α_2, for simple implementation we could define

$$\tilde{\alpha}_1 = \frac{\sum_{i=1}^{N} |x_i|}{\sum_{i=1}^{N} x_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} |x_i|\right)^2}$$

$$\tilde{\alpha}_2 = \frac{\sum_{i=1}^{N} |y_i|}{\sum_{i=1}^{N} y_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} |y_i|\right)^2}$$

The combined output stream is

$$z_k = \tilde{\alpha}_1 x_k + \tilde{\alpha}_2 y_k$$

The SNR loss can be evaluated by computing the output SNR:

$$E(z_k) = E(\tilde{\alpha}_1) m_1 + E(\tilde{\alpha}_2) m_2$$

$$\begin{aligned}\mathrm{Var}(z_k) &= E(z_k^2) - E(z_k)^2 \\ &= E\left[(\tilde{\alpha}_1 x + \tilde{\alpha}_2 y)^2\right] - \left(E(\tilde{\alpha}_1) m_1 + E(\tilde{\alpha}_2) m_2\right)^2 \\ &= E(\tilde{\alpha}_1^2) E(x^2) + E(\tilde{\alpha}_2^2) E(y^2) \\ &\quad - \left(E(\tilde{\alpha}_1) m_1\right)^2 - \left(E(\tilde{\alpha}_2) m_2\right)^2\end{aligned}$$

From Appendix A, we can find

$$C_1 = E(\tilde{\alpha}_1) m_1, \quad C_2 = E(\tilde{\alpha}_2) m_2$$

$$D_1 = E(\tilde{\alpha}_1^2) E(x^2), \quad D_2 = E(\tilde{\alpha}_2^2) E(y^2)$$

Then

$$R = \mathrm{SNR}_{z_k} = \frac{[E(z_k)]^2}{2\,\mathrm{VAR}(z_k)} = \frac{(C_1 + C_2)^2}{2[D_1 + D_2 - C_1^2 - C_2^2]}$$

which is a function of R_1, R_2 and N_1, N_2 where N_1, N_2 denote the number of samples of the sequences $\{x_k\}$ and $\{y_k\}$ respectively.

Figure 6 shows the SNR loss as a function of R_1 and N_1 for $R_2 = -7$ dB and for $N_2 = 1000$ and 10,000. Figure 7 plots the SNR loss versus R_2 and N_2 for $R_1 = -1$ dB and for $N_1 = 1000$ and 10,000. We see that the combining loss is negligible for two symbol streams with same SNRs. The SNR loss grows significantly as the difference in the two SNRs gets larger. This is due to the large bias at low SNRs introduced by the use of absolute moment as an approximation to the mean.

B. Unbiased Combining Algorithm

Since we will often be dealing with low and unequal SNRs (from -12 dB to -1 dB), we must find some way to remove the combiner bias. For convenience let us define

$$a_1 = \sum_{i=1}^{N_1} |x_i|$$

$$s_1 = \sum_{i=1}^{N_1} x_i^2$$

Let $\tilde{\alpha}_1$ be the biased estimate of α_1 and $\tilde{R}_1$ be the biased estimate of R_1, then

$$\tilde{\alpha}_1 = (N_1 - 1)\frac{a_1}{N_1 s_1 - a_1^2}$$

$$\tilde{R}_1 = \frac{\tilde{\alpha}_1 a_1}{2N_1}$$

From Appendix A, for large N we have

$$E[\tilde{\alpha}_1] = \frac{\mu}{\sigma_v^2}(1 + B) \simeq \frac{\mu}{\sigma_v^2}$$

$$= \frac{m_1}{2\sigma_1^2} \frac{\frac{\exp -R_1}{\sqrt{\pi R_1}} + \mathrm{erf}(\sqrt{R_1})}{\left\{R_1 + \frac{1}{2} - \left[\frac{\exp -R_1}{\sqrt{\pi}} + \sqrt{R_1}\,\mathrm{erf}(\sqrt{R_1})\right]^2\right\}}$$

But $\alpha_1 = m_1/\sigma_1^2$, so

$$\alpha_1 = \frac{2\left[R_1 + \frac{1}{2} - \left(\frac{\exp -R_1}{\sqrt{\pi}} + \sqrt{R_1}\ \mathrm{erf}\ (\sqrt{R_1})\right)^2\right]}{\frac{\exp -R_1}{\sqrt{\pi R_1}} + \mathrm{erf}\ (\sqrt{R_1})} E(\tilde{\alpha}_1)$$

$$= g(R_1)\, E(\tilde{\alpha}_1) \qquad (1)$$

Since we don't know R_1, we also have to estimate it. From Section III,

$$E(\tilde{R}_1) = f(R_1)$$

$$= \frac{\left[\frac{\exp -R_1}{\sqrt{\pi}} + \sqrt{R_1}\ \mathrm{erf}\ \left(\sqrt{R_1}\right)\right]^2}{2\left[R_1 + \frac{1}{2} - \left(\frac{\exp -R_1}{\sqrt{\pi}} + \sqrt{R_1}\ \mathrm{erf}\ (\sqrt{R_1})\right)^2\right]}$$

Layland (Ref. 1) defines the asymtotically unbiased estimate of R_1 to be

$$\hat{R}_1 = f^{-1}(\tilde{R}_1)$$

Using $\hat{R}_1$ in place of R_1 in Eq. (1), we define the asymtotically unbiased estimate of α_1 to be

$$\hat{\alpha}_1 = g(\hat{R}_1)\,\tilde{\alpha}_1 = gof^{-1}(\tilde{R}_1)\,\tilde{\alpha}_1 = h(\tilde{R}_1)\,\tilde{\alpha}_1$$

The function $h(\tilde{R})$ is shown in Fig. 8 for $\tilde{R}$ from 0.8796 to 1.3893 (for R from -13 dB to 0 dB). Using second-order polynominal interpolation, this function can be approximated with an error less than 0.002 by

for $0.8796 \leqslant \tilde{R} \leqslant 0.9000$

$$h(\tilde{R}) = -121.49\,\tilde{R}^2 + 221.5\,\tilde{R} - 100.6772$$

for $0.9000 < \tilde{R} \leqslant 0.9600$

$$h(\tilde{R}) = -14.6\,\tilde{R}^2 + 29.5263\,\tilde{R} - 14.4791$$

for $0.9600 < \tilde{R} \leqslant 1.0500$

$$h(\tilde{R}) = -3.5596\,\tilde{R}^2 + 8.4951\,\tilde{R} - 4.4639$$

for $1.0500 < \tilde{R} \leqslant 1.3893$

$$h(\tilde{R}) = -1.0004\,\tilde{R}^2 + 3.0811\,\tilde{R} - 1.5990$$

Note that, for $\tilde{R} \leqslant 0.9000$, the polynominal coefficients are quite large compared to the function values. So the function values are *very* sensitive to small changes in coefficients, and one must be careful in using this approximation in finite arithmetic applications.

Since it is extremely complicated to analyze the performance of this estimator, we simulate the algorithm (see Appendix B) to find the SNR loss for different R_1, R_2 and N_1, N_2. Figure 9 shows the SNR loss as a function of R_1 and N_1 for R_2 = -7 dB and $\log_{10} N_2$ = 3,4. Figure 10 plots the SNR loss versus R_2 and N_2 for R_1 = -1 dB and $\log_{10} N_1$ = 3.4. We see that as N_1, N_2 get larger, the SNR loss goes to zero. In Fig. 9, the R_1 = -1 dB curves cross because with a very small number of samples (N_1 = 100) the simulation results are not very reliable, especially for low SNR sequences. The 100-sample block may fall into either a very low noise section or a very noisy one. The same thing happens for the R_2 = -3 dB curves in Fig. 10. The combining loss goes down to approximately 0.012 dB with N_1 = 1000 for the -1 dB stream and N_2 = 10,000 for the -7 dB stream. In general larger numbers of samples are required for lower SNR sequences.

V. Conclusion

The conventional SNR and combiner weight estimator has large bias at low SNRs. As a result, the output SNR loss grows significantly as the difference in SNR of the two symbol streams gets larger. When the algorithm is modified to remove the bias, the bias is essentially removed with a sufficiently large number of samples. Sample sizes of 1000 for the -1 dB stream and 10,000 for the -7 dB stream are sufficient to achieve an output SNR which is just 0.012 dB less than the maximum achievable limit.

References

1. Layland, J. W., "On S/N Estimation," *Space Programs Summary 37-48*, Vol. III, pp. 209-212, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 31, 1967.

2. Lesh, J. R., "Accuracy of the Signal to Noise Ratio Estimator," *Technical Report 32-1526*, Vol. X, pp. 217-235, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1972.

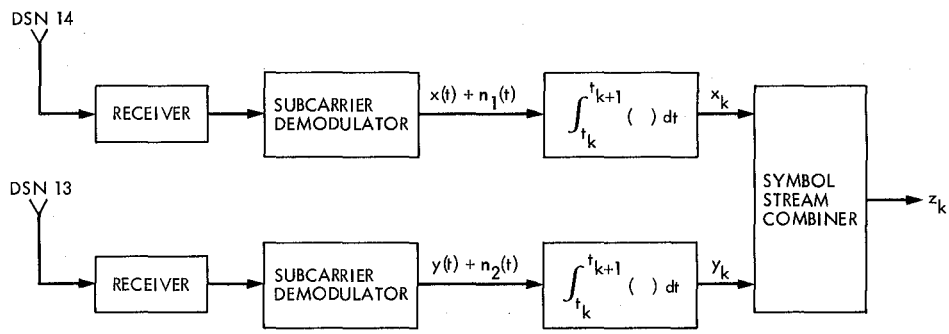


Fig. 1. Model of the symbol stream combining process

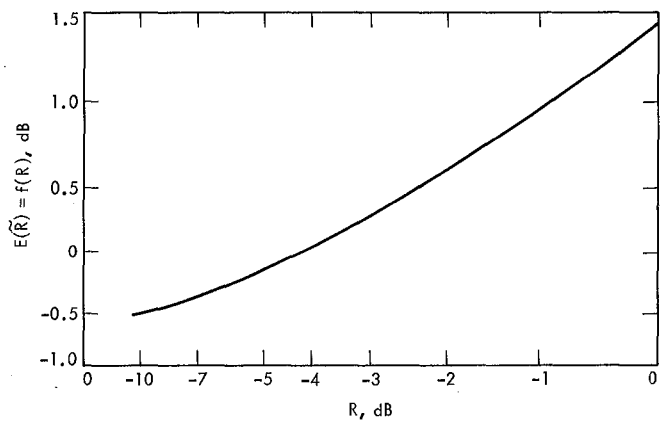


Fig. 2. Biased SNR estimator mean vs actual SNR

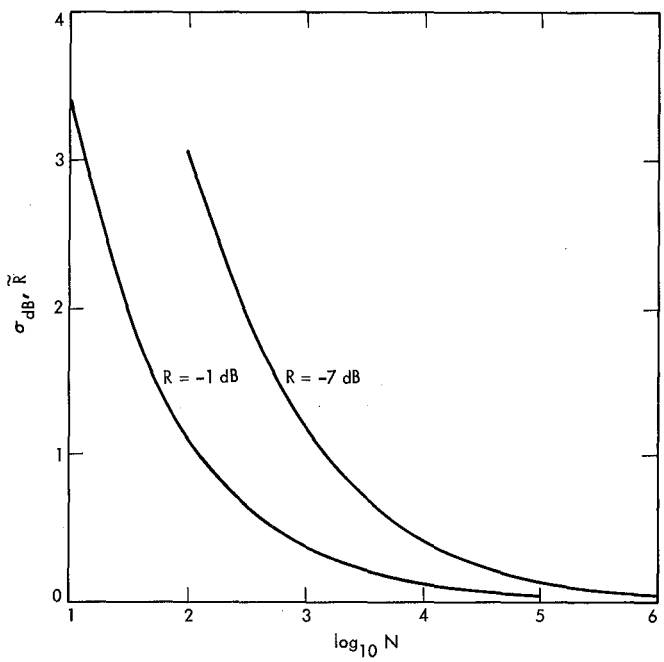


Fig. 3. Dependence of standard deviation on sample size

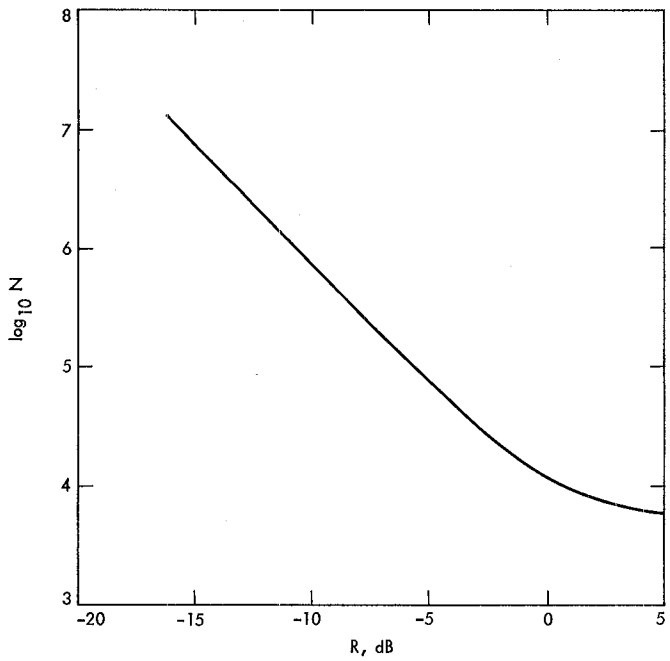


Fig. 4. Number of symbols required to achieve σdB($\hat{R}$) = 0.1 dB

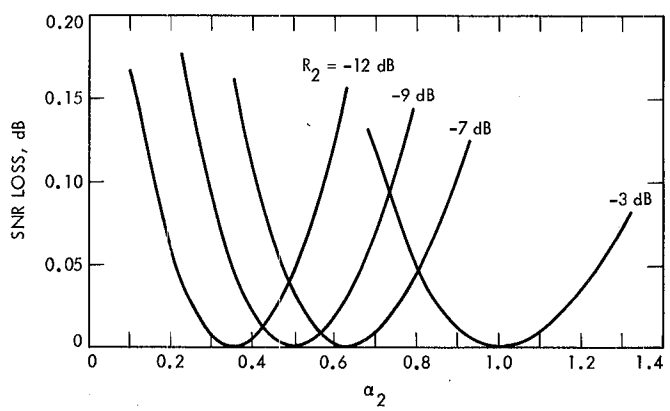


Fig. 5. SNR loss vs R_2 and α_2 for $R_1 = -1$ dB

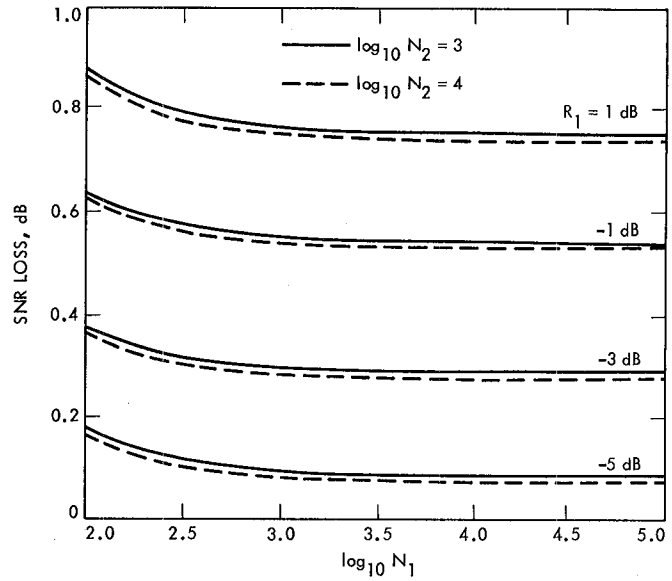


Fig. 6. SNR loss vs R_1 and N_1 for $R_2 = -7$ dB (conventional method, without bias removed)

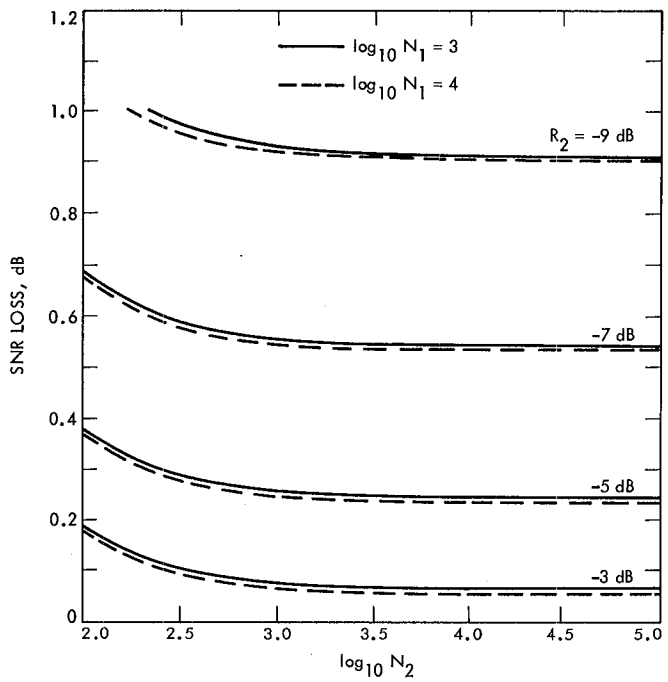


Fig. 7. SNR loss vs R_2 and N_2 for $R_1 = -1$ dB (conventional method, without bias removed)

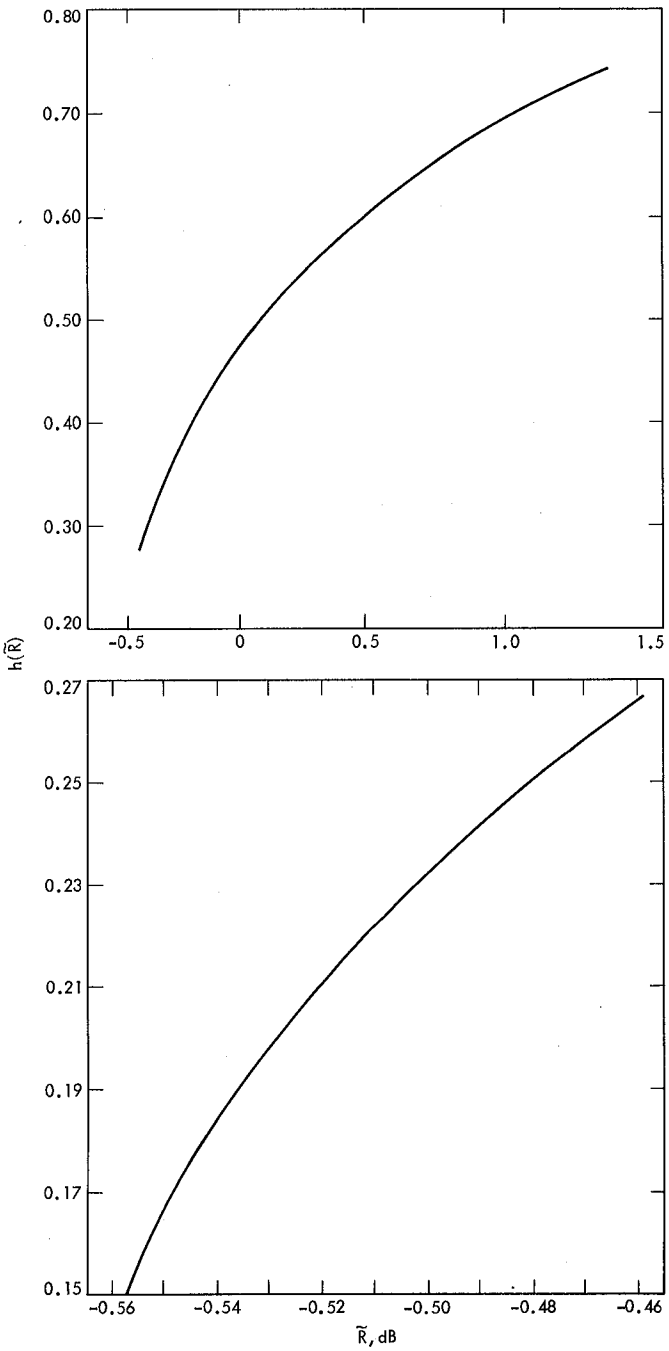


Fig. 8. Bias-removed adjustment function h(R̃)

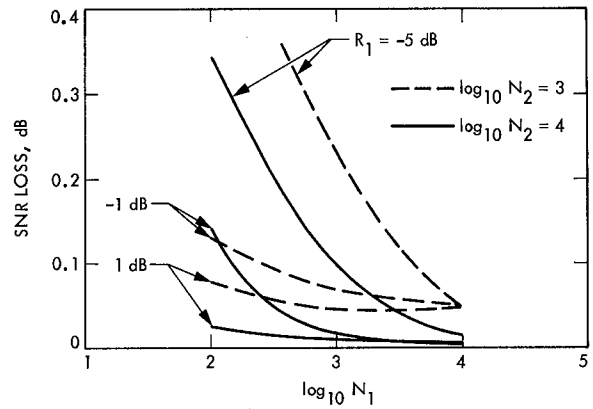


Fig. 9. Simulation of SNR loss vs R_1 and N_1 for $R_2 = -7$ dB (conventional method, with bias removed)

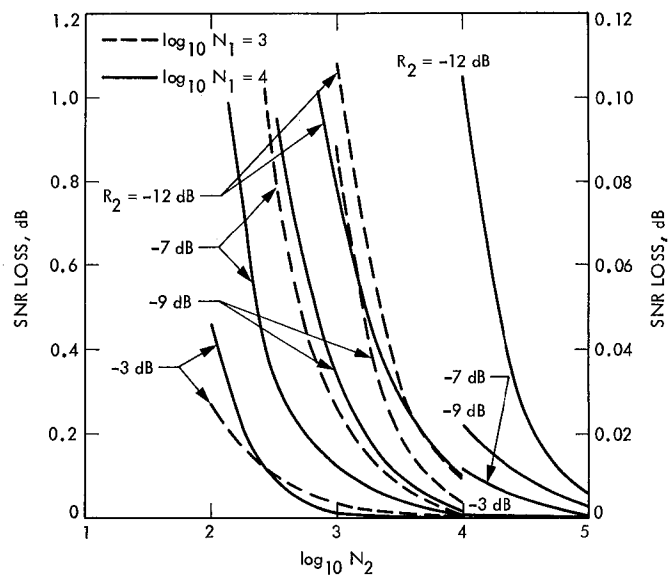


Fig. 10. Simulation of SNR loss vs R_2 and N_2 for $R_1 = -1$ dB (conventional method, with bias removed)

Appendix A

Evaluation of First and Second Moments for the Biased Combiner Weight Estimator

Let

$$x_k = \pm m + n_k$$

where n_k are i.i.d. zero mean Gaussian random variables with variance σ^2.

We want to find $E[\widetilde{\alpha}]$ and $E[\widetilde{\alpha}^2]$ where

$$\widetilde{\alpha} = \frac{\text{sample mean of } |x_k|}{\text{sample variance of } |x_k|}$$

$$= \frac{\frac{1}{N}\sum_{k=1}^{N} |x_k|}{\frac{1}{N-1}\sum_{k=1}^{N}\left(|x_k| - \frac{1}{N}\sum_{i=1}^{N}|x_i|\right)^2}$$

If the sample size is large enough ($N \geq 20$ should suffice), we can use the central limit theorem to express α by

$$\widetilde{\alpha} = \frac{\mu + \sqrt{\epsilon_1^2}\,\xi}{\sigma_v^2 + \sqrt{\epsilon_2^2}\,\psi}$$

where ξ and ψ are independent zero mean Gaussian random variables with unit variances and

$$\mu = E\{|x_k|\}$$

$$\sigma_v^2 = \text{var}\,|x_k|$$

$$\epsilon_1^2 = \frac{1}{N}\sigma_v^2$$

From Appendix A of Ref. 2, the variance of a sample variance is

$$\epsilon_2^2 = \frac{1}{N}\left[E(x_k^4) - 4E(|x_k|^3)\mu + 3\left(E(x_k^2)\right)^2 - \frac{4N-6}{N-1}\sigma_v^4\right]$$

For $\epsilon_2/\sigma_v^2 << 1$ (which is usually the case for large N) we have

$$\widetilde{\alpha} = \frac{\mu}{\sigma_v^2}\left(1 + \frac{\epsilon_1}{\mu}\xi\right)\left(1 - \frac{\epsilon_2}{\sigma_v^2}\psi + \frac{\epsilon_2^2}{\sigma_v^4}\psi^2\right)$$

So

$$E[\widetilde{\alpha}] = \frac{\mu}{\sigma_v^2}(1+B)$$

where

$$B = \frac{\epsilon_2^2}{\sigma_v^4}$$

From Lesh (Ref. 2), we know

$$E\{|x_k|\} = \sqrt{NoT}\left\{\frac{\exp -R}{\sqrt{\pi}} + \sqrt{R}\,\text{erf}(\sqrt{R})\right\}$$

$$E(x_k^2) = NoT\left(R + \frac{1}{2}\right)$$

$$E\{|x_k|^3\} = (NoT)^{3/2}$$

$$\left\{\frac{R+1}{\sqrt{\pi}}\exp -R + \left(R + \frac{3}{2}\right)\sqrt{R}\,\text{erf}(\sqrt{R})\right\}$$

$$E\{x_k^4\} = No^2T^2\left(R^2 + 3R + \frac{3}{4}\right)$$

Since $m = \sqrt{R}\sqrt{NoT}$, we have

$$E[\tilde{\alpha}]\ m = \frac{\sqrt{\frac{R}{\pi}} \exp -R + R\ \text{erf}(\sqrt{R})}{R + \frac{1}{2} - \left[\frac{\exp -R}{\sqrt{\pi}} + \sqrt{R}\ \text{erf}\ (\sqrt{R})\right]^2} (1 + B)$$

$$E(\tilde{\alpha}^2) = \frac{\mu^2}{\sigma_v^4} E\left(1 + \frac{2\epsilon_1}{\mu}\xi + \frac{\epsilon_1^2}{\mu^2}\xi^2\right)$$

$$\times E\left(1 - 2\sqrt{B}\,\psi + 3B\,\psi^2 - 2B^{3/2}\,\psi^3 + B^2\,\psi^4\right)$$

$$= \frac{\mu^2}{\sigma_v^4}(1 + A)(1 + 3B + 3B^2)$$

where

$$A = \frac{\epsilon_1^2}{\mu^2}$$

Thus

$$E(\alpha^2)\,E(x_k^2) = \frac{\left(R + \frac{1}{2}\right)\left[\frac{\exp -R}{\sqrt{\pi}} + \sqrt{R}\ \text{erf}\ (\sqrt{R})\right]^2}{\left[R + \frac{1}{2} - \left[\frac{\exp -R}{\sqrt{\pi}} + \sqrt{R}\ \text{erf}\ (\sqrt{R})\right]^2\right]^2}$$

$$\times\ (1 + A)(1 + 3B + 3B^2)$$

Appendix B

Simulation of the Unbiased Symbol Stream Combining Algorithm

The combining algorithm is summarized as follows:

Step 1: Compute

$$a_1 = \sum |x_k| \quad , \quad s_1 = \sum x_k^2$$

$$a_2 = \sum |y_k| \quad , \quad s_2 = \sum y_k^2$$

Step 2: Compute the intermediate estimates

$$\tilde{\alpha}_1 = \frac{(N_1 - 1)a_1}{N_1 s_1 - a_1^2} \quad , \quad \tilde{R}_1 = \frac{\tilde{\alpha}_1 a_1}{2N_1}$$

$$\tilde{\alpha}_2 = \frac{(N_2 - 1)a_2}{N_2 s_2 - a_2^2} \quad , \quad \tilde{R}_2 = \frac{\tilde{\alpha}_2 a_2}{2N_2}$$

Step 3: Compute the final estimates

$$\hat{\alpha}_1 = h(\tilde{R}_1)\tilde{\alpha}_1, \quad \hat{\alpha}_2 = h(\tilde{R}_2)\tilde{\alpha}_2$$

where the function $h(\cdot)$ is approximated by second-order polynominals given in Section IV. The combined symbol stream will be $z_k = \hat{\alpha}_1 x_k + \hat{\alpha}_2 y_k$.

By means of subroutine NOISE, we generate two independent sequences of symbols,

$$x_k = \pm\sqrt{2R_1} + n_k^1, \; y_k = \pm\sqrt{2R_2} + n_k^2$$

where n_k^1, n_k^2 are independent, identically distributed normal random variables. Thus,

$$\mathrm{SNR}_{x_k} = \frac{m_1^2}{2\sigma_1^2} = R_1$$

and

$$\mathrm{SNR}_{y_k} = \frac{m_2^2}{2\sigma_2^2} = R_2$$

Each time subroutine NOISE is called, it generates two symbols, either x_k, x_{k+1} or y_k, y_{k+1}. In the main program, we gather all the x_k's and y_k's and compute $\hat{\alpha}_1$ and $\hat{\alpha}_2$ according to the above algorithm. To find the output SNR we note that

$$E(z_k) = \hat{\alpha}_1 E(x_k) + \hat{\alpha}_2 E(y_k) = \hat{\alpha}_1\sqrt{2R_1} + \hat{\alpha}_2\sqrt{2R_2}$$

$$\mathrm{Var}(z_k) = \hat{\alpha}_1^2 \mathrm{Var}(x_k) + \hat{\alpha}_2^2 \mathrm{Var}(y_k) = \hat{\alpha}_1^2 + \hat{\alpha}_2^2$$

So,

$$\mathrm{SNR}_{z_k} = R = \frac{\left(\hat{\alpha}_1\sqrt{2R_1} + \hat{\alpha}_2\sqrt{2R_2}\right)^2}{2(\hat{\alpha}_1^2 + \hat{\alpha}_2^2)}$$

We run the same routine 25 times with different sets of samples (i.e., $\{x_k\}$ and $\{y_k\}$). The final SNR loss is computed based on the averaged output SNR:

$$\text{SNR loss} = 10 \log\left(\frac{R_1 + R_2}{R}\right)$$

The simulation is run for different values of R_1, N_1 and R_2, N_2. The results are plotted in Figs. 8 and 9. Listings of the main program, the function $h(\cdot)$, and the subroutine NOISE are also provided for reference in Figs. B-1, B-2, and B-3.

```
   SNR.FOR

   IMPLICIT REAL*8 (A-H,O-Z)
   DIMENSION X(100000),Y(100000)
   COMMON /COM1/B1,BB,CC,VV
   B1=2.**31
   BB=B1-1.
   CC=7.**5
   VV=1099.
   READ(5,10)R1,N1
10 FORMAT(F4.0,I6)
   PRINT 15,R1,N1
15 FORMAT(2X,F4.0,2X,I6)
   R1=10.**(.1*R1)
   AR1=DSQRT(2.*R1)
   XN1=DFLOAT(N1)
   DO 70 K=1,2
   R2=DFLOAT(4*K-11)
   R2=10.**(.1*R2)
   AR2=DSQRT(2.*R2)
   DO 70 L=2,5
   N2=10**L
   XN2=DFLOAT(N2)
   R=0.
   DO 50 I=1,25
   A1=0.
   S1=0.
   DO 20 N=1,N1,2
   CALL NOISE(AR1,X(N),X(N+1))
   A1=A1+DABS(X(N))+DABS(X(N+1))
20 S1=S1+X(N)*X(N)+X(N+1)*X(N+1)
   AT1=(XN1-1.)*A1/(XN1*S1-A1*A1)
   RT1=.5*A1*AT1/XN1
   AT1=H(RT1)*AT1
   A2=0.
   S2=0.
   DO 30 N=1,N2,2
   CALL NOISE(AR2,Y(N),Y(N+1))
   A2=A2+DABS(Y(N))+DABS(Y(N+1))
30 S2=S2+Y(N)*Y(N)+Y(N+1)*Y(N+1)
   AT2=(XN2-1.)*A2/(XN2*S2-A2*A2)
   RT2=.5*A2*AT2/XN2
   AT2=H(RT2)*AT2
   RR=.5*(AT1*AR1+AT2*AR2)**2/(AT1*AT1+AT2*AT2)
50 R=R+RR
   R=R/25.
   RLOSS=10.*DLOG10((R1+R2)/R)
   PRINT 60,N2,R1,R2,R,RLOSS
60 FORMAT(2X,I8,4(2X,F8.4))
70 CONTINUE
   STOP
   END
```

Fig. B-1. Main program

```
   FUNCTION H(X)
   IF(X.GT.0.9000) GO TO 10
   H=-121.49*X*X+221.5*X-100.6772
   RETURN
10 IF(X.GT.0.9600) GO TO 20
   H=-14.6*X*X+29.5263*X-14.4791
   RETURN
20 IF(X.GT.1.0500) GO TO 30
   H=-3.5596*X*X+8.4951*X-4.4639
   RETURN
30 H=-1.0004*X*X+3.0811*X-1.5990
   RETURN
   END
```

Fig. B-2. Function h(·)

```
   SUBROUTINE NOISE(X,E1,E2)
   IMPLICIT REAL*8 (A-H,O-Z)
   COMMON /COM1/B1,BB,CC,VV
10 VV=VV*CC
   VV=DMOD(VV,BB)
   V1=2.*VV/B1-1.
   VV=VV*CC
   VV=DMOD(VV,BB)
   V2=2.*VV/B1-1.
   S=V1*V1+V2*V2
   IF(S-1.) 20,10,10
20 V=DSQRT(-2.*DLOG(S)/S)
   E1=V1*V+X
   E2=V2*V-X
   RETURN
   END
```

Fig. B-3. Subroutine NOISE

A Systolic VLSI Design of a Pipeline Reed-Solomon Decoder

H. M. Shao, T. K. Truong, L. J. Deutsch, and J. H. Yuen
Communication System Research Section

I. S. Reed
University of Southern California

A pipeline structure of a transform decoder similar to a systolic array is developed to decode Reed-Solomon (RS) codes. An important ingredient of this design is a modified Euclidean algorithm for computing the error locator polynomial. The computation of inverse field elements is completely avoided in this modification of Euclid's algorithm. The new decoder is regular and simple, and naturally suitable for VLSI implementation. An example illustrating both the pipeline and systolic array aspects of this decoder structure is given for a (15, 9) RS code.

I. Introduction

A concatenated coding system consisting of a convolutional inner code and a Reed-Solomon (RS) outer code has been adopted as the standard for future space missions by both the European Space Agency (ESA) and NASA (Ref. 1). The convolutional inner code is a (7, 1/2) code which is also used on the Voyager Project. The outer Reed-Solomon code is a (255, 223) block code of 8-bit symbols that is capable of correcting up to 16 symbol errors. The performance of such a concatenated code is investigated in Ref. 2 where it is shown that the concatenated channel provides a coding gain of approximately 2 dB over the channel with only convolutional coding for a decoded bit error rate of 10^{-5}.

Some work has already been done in developing VLSI encoders for RS codes. The algorithm that was developed in Ref. 3 by Berlekamp was used in Ref. 4 to design a single chip (255, 223) RS encoder.

Recently it was suggested by Brent and Kung (Ref. 5) that a pipeline architecture could be used to compute the greatest common divisor (GCD) of two polynomials. In Euclid's algorithm, the GCD of two polynomials can be used to obtain the error locator polynomial of an RS code. It is shown in this article that a modified form of Euclid's algorithm, based on the idea of Brent and Kung, can be used to find the error locator polynomial in a transform RS decoder. This algorithm requires no multiplications by an inverse element in the finite field.

Utilizing the above mentioned modification of Euclid's algorithm, a new pipeline architecture for a transform decoder is developed to decode RS codes. This architecture can be implemented on VLSI chips with nMOS technology. Such a pipeline decoding algorithm can be realized in a systolic-array architecture, which is presented in this article as an example.

II. A VLSI Design of a RS Decoder

Let $GF(2^m)$ be the finite field of 2^m elements. Also, let $N = 2^m - 1$ be the length of the (N, I) RS code over $GF(2^m)$ with minimum distance $d = 2t + 1$, where $I = N - (d - 1)$ denotes the number of m-bit message symbols and t denotes the number of errors that may be corrected in each codeword.

The standard transform decoder for decoding RS codes was described in Ref. 6. In the transform decoding algorithm, the Berlekamp-Massey, Euclid, or continued-fraction algorithm can be used to find the error locator polynomial. It was decided for this design to use the Euclid algorithm because its modularity makes it well suited for VLSI implementation. The operations needed to compute Euclid's algorithm generally require the computation of inverse elements in GF(2^m). To avoid the computation of inverse elements, a modification of Euclid's algorithm can be utilized to find the error locator polynomial.

The transform decoding algorithm, which utilizes the modified form of Euclid's algorithm to find the error locator polynomial, is described in the following four steps:

Step 1: Compute the syndromes

$$S_k = \sum_{n=0}^{N-1} r_n \alpha^{nk}, \quad 1 \leqslant k \leqslant 2t \tag{1}$$

where r_n $(0 \leqslant n \leqslant N - 1)$ is the received code pattern. Note that $E_k = S_k$ $(1 \leqslant k \leqslant 2t)$, where E_k is the transform of the error sequence (Ref. 5).

Step 2: Perform the modified Euclid algorithm (described in the next section) on x^{2t} and the syndrome polynomial

$$S(x) = \sum_{k=1}^{2t} S_k x^{2t-k} \tag{2}$$

to obtain the polynomial,

$$\lambda(x) = \lambda_0 x^t + \lambda_1 x^{t-1} + \cdots + \lambda_t \tag{3a}$$

Divide the polynomial $\lambda(x)$ by its leading coefficient λ_0. This yields the standard error-locator polynomial.

$$\sigma(x) = x^t + \sigma_1 x^{t-1} + \cdots + \sigma_t \tag{3b}$$

where

$$\sigma_i = \frac{\lambda_i}{\lambda_0}, \quad 1 \leqslant i \leqslant t.$$

Step 3: From the known coefficients σ_k in Eq. (3b), compute the remaining elements E_k of the transform of the error sequence for $d \leqslant k \leqslant N$ where $E_0 = E_N$ from the equation (Ref. 5):

$$E_{2t+j} + \sum_{k=1}^{2t} (-1)^k \sigma_k E_{2t+j-k} = 0 \quad \text{for } j \geqslant 1 \tag{4}$$

Step 4: Compute the inverse transform of E_k over $GF(2^m)$ to obtain the estimated error pattern. That is,

$$e_n = \sum_{k=0}^{N-1} E_k \alpha^{-nk}, \quad 0 \leqslant n \leqslant N-1 \tag{5}$$

Finally, the estimate of the original code vector is obtained by subtracting the error pattern from the received pattern.

Figure 1 shows an overall block diagram of this decoder. In the following sections, a VLSI design for each of the functional blocks is described.

III. A VLSI Design for Computing Syndromes

In this section, a VLSI architecture is developed to compute the syndromes for an (N, I) RS code over $GF(2^m)$. The decoding procedure and the VLSI architecture is illustrated in the following simple example:

Example. Consider a (15, 9) RS code over $GF(2^4)$ with minimum distance $d = 7$. In this code, $t = 3$ errors can be corrected.

To compute a generator polynomial for this code (Ref. 7), one can use the representation of the field $GF(2^4)$ given in Table A-1 of the Appendix. One generator polynomial of such a (15, 9) RS code is

$$g(x) = \prod_{i=1}^{6} (x - \alpha^i)$$

$$= x^6 + \alpha^{10} x^5 + \alpha^{14} x^4 + \alpha^4 x^3 + \alpha^6 x^2 + \alpha^9 x + \alpha^6 \tag{6}$$

Assume the message symbols are

$$I(x) = \alpha^{10}x^{14} + \alpha^{12}x^{13} + \alpha^{8}x^{12} + \alpha^{5}x^{11} + \alpha^{6}x^{10} + \alpha^{14}x^{9} + \alpha^{13}x^{8} + \alpha^{11}x^{7} + \alpha^{9}x^{6} \quad (7)$$

The encoded codeword, which is a multiple of $g(x)$, is then

$$c(x) = \alpha^{10}x^{14} + \alpha^{12}x^{13} + \alpha^{8}x^{12} + \alpha^{5}x^{11} + \alpha^{6}x^{10} + \alpha^{14}x^{9} + \alpha^{13}x^{8} + \alpha^{11}x^{7} + \alpha^{9}x^{6} + x^{5} + \alpha x^{4} + \alpha^{2}x^{3} + \alpha^{6}x^{2} + \alpha^{12}x + \alpha^{8} \quad (8a)$$

Written as a vector, the codeword is

$$c(x) = (\alpha^{10}, \alpha^{12}, \alpha^{8}, \alpha^{5}, \alpha^{6}, \alpha^{14}, \alpha^{13}, \alpha^{11}, \alpha^{9}, \alpha^{0}, \alpha, \alpha^{2}, \alpha^{6}, \alpha^{12}, \alpha^{8}) \quad (8b)$$

Suppose that two errors exist in the error pattern as follows:

$$e(x) = (0,0,0,0,\alpha^{11},0,0,0,0,0,0,\alpha^{7},0,0,0) \quad (9)$$

Then the received pattern is

$$r(x) = (r_{14}, r_{13}, \cdots, r_0) = c(x) + e(x)$$
$$= (\alpha^{10}, \alpha^{12}, \alpha^{8}, \alpha^{5}, \alpha, \alpha^{14}, \alpha^{13}, \alpha^{11}, \alpha^{9}, \alpha^{0}, \alpha, \alpha^{12}, \alpha^{6}, \alpha^{12}, \alpha^{8}) \quad (10)$$

The syndromes of this received message are

$$S_k = \sum_{n=0}^{14} r_n \alpha^{nk}$$
$$= \alpha^{7}(\alpha^{3})^{k} + \alpha^{11}(\alpha^{10})^{k} \qquad \text{for } 1 \leqslant k \leqslant 6 \quad (11)$$

This yields $S_1 = \alpha^7$, $S_2 = \alpha^{12}$, $S_3 = \alpha^6$, $S_4 = \alpha^{12}$, $S_5 = \alpha^{14}$ and $S_6 = \alpha^{14}$. To develop a VLSI design to compute S_k, let the first expression in Eq. (11) be rewritten in recursive form as

$$S_k = (\ldots((r_{14}\alpha^{k} + r_{13})\alpha^{k} + r_{12})\alpha^{k} + \cdots)\alpha^{k} + r_0$$
$$= (\ldots((\alpha^{10}\alpha^{k} + \alpha^{12})\alpha^{k} + \alpha^{8})\alpha^{k} + \cdots)\alpha^{k} + \alpha^{8} \quad (12)$$

A structure for computing Eq. (12) is shown in Fig. 2. In this figure, the function of each cell is given by the register transfer relation

$$B_k \leftarrow A_k + B_k\alpha^{7-k} \qquad \text{for } 1 \leqslant k \leqslant 6 \quad (13)$$

where "←" denotes the operation, "is replaced by." The input data are sent to all of the cells simultaneously. Assume initially that all registers are set to zero. After the complete received codeword is entered, the desired syndromes S_k are contained in registers B_{7-k} $(1 \leqslant k \leqslant 6)$. The syndromes computed in this manner are shifted serially from the right end of register B_6 and fed into the next stage, which performs a modification of Euclid's algorithm. The sequence $\{S_k\}$ of syndromes are conveniently represented as in Eq. (2) by what is called the syndrome polynomial,

$$S(x) = \sum_{k=1}^{2t} S_k x^{2t-k}$$

In particular, the syndrome polynomial for this example is

$$S(x) = \alpha^{7}x^{5} + \alpha^{12}x^{4} + \alpha^{6}x^{3} + \alpha^{12}x^{2} + \alpha^{14}x + \alpha^{14} \quad (14)$$

IV. A Systolic Array Architecture to Compute the Error-Locator Polynomial with a Modified Form of Euclid's Algorithm

It is shown (Ref. 7) that the error locator polynomial can be calculated from the syndrome polynomial by means of Euclid's algorithm. Since quotients are needed in the usual form of Euclid's algorithm, inverses of field elements are required at each stage. The successive computation of these inverses is difficult to realize in a VLSI circuit. It was shown in Ref. 5 that Euclid's algorithm could be modified to eliminate the computation of inverse elements. In this section, this new idea is applied to the computation of the polynomial $\lambda(x)$, defined in Eq. (3a).

Consider the two polynomials,

$$A(x) = x^{2t}$$

$$S(x) = \sum_{k=1}^{2t} S_k x^{2t-k}$$

The modified form of Euclid's algorithm is a recursive procedure for finding the i^{th} remainder $R_i(x)$ and the quantities $\gamma_i(x)$ and $\lambda_i(x)$ that satisfy

$$\gamma_i(x)A(x) + \lambda_i(x)S(x) = R_i(x) \tag{15}$$

When the degree of the remainder $R_i(x)$ is less than t, the algorithm stops. The resulting $\lambda_i(x)$ at the termination of the algorithm is the desired polynomial, $\lambda(x)$, in Eq. (3a).

The initial conditions of the algorithm are

$$R_0(x) = A(x) \qquad Q_0(x) = S(x) \tag{16a}$$

$$\lambda_0(x) = 0 \qquad \mu_0(x) = 1 \tag{16b}$$

$$\gamma_0(x) = 1 \qquad \eta_0(x) = 0 \tag{16c}$$

For $i \geqslant 1$, compute recursively

$$R_i(x) = b_{i-1}[\sigma_{i-1}R_{i-1}(x) + \bar{\sigma}_{i-1}Q_{i-1}(x)] - a_{i-1}x^{1_{i-1}}[\sigma_{i-1}Q_{i-1}(x) + \bar{\sigma}_{i-1}R_{i-1}(x)] \tag{17a}$$

$$\lambda_i(x) = b_{i-1}[\sigma_{i-1}\lambda_{i-1}(x) + \bar{\sigma}_{i-1}\mu_{i-1}(x)] - a_{i-1}x^{1_{i-1}}[\sigma_{i-1}\mu_{i-1}(x) + \bar{\sigma}_{i-1}\lambda_{i-1}(x)] \tag{17b}$$

$$\gamma_i(x) = b_{i-1}[\sigma_{i-1}\gamma_{i-1}(x) + \bar{\sigma}_{i-1}\eta_{i-1}(x)] - a_{i-1}x^{1_{i-1}}[\sigma_{i-1}\eta_{i-1}(x) + \bar{\sigma}_{i-1}\gamma_{i-1}(x)] \tag{17c}$$

$$Q_i(x) = \sigma_{i-1}Q_{i-1}(x) + \bar{\sigma}_{i-1}R_{i-1}(x) \tag{17d}$$

$$\mu_i(x) = \sigma_{i-1}\mu_{i-1}(x) + \bar{\sigma}_{i-1}\lambda_{i-1}(x) \tag{17e}$$

$$\eta_i(x) = \sigma_{i-1}\eta_{i-1}(x) + \bar{\sigma}_{i-1}\gamma_{i-1}(x) \tag{17f}$$

where a_{i-1} and b_{i-1} are the leading coefficients of $R_{i-1}(x)$ and $Q_{i-1}(x)$, respectively

$$1_{i-1} = \deg(R_{i-1}(x)) - \deg(Q_{i-1}(x))$$

$$\sigma_{i-1} = 1 \qquad \text{if } 1_{i-1} \geqslant 0$$

$$\sigma_{i-1} = 0 \qquad \text{if } 1_{i-1} < 0$$

The algorithm stops when $\deg(R_i(x)) < t$.

The proof that this algorithm performs the desired computation is similar to that of the standard version of Euclid's algorithm, and it is not included here. Notice that the $2t$ computations of inverse field elements that are needed in the usual form of Euclid's algorithm are not required in the above modified version of Euclid's algorithm. The computational details needed to find the error-locator polynomial of an RS(15, 9) code over $GF(2^4)$ using the modified Euclid's algorithm are given next in a continuation of the previous example. The implementation of this algorithm is accomplished by means of a systolic array cell design. Such a systolic array structure is presented here in considerable detail.

Consider the decoding of the three-error-correcting RS(15, 9) code described previously. The syndrome polynomial $S(x)$, given in Eq. (14), is used in the modified Euclid's algorithm to calculate $\lambda(x)$.

Step 1. By Eq. (17), for $i = 1$: $1_0 = 6 - 5 = 1$ and $\sigma_0 = 1$. Hence let

$$R_1(x) = \alpha^7 R_0(x) + xQ_0(x)$$

$$\lambda_1(x) = \alpha^7\lambda_0(x) + x\mu_0(x)$$

$$Q_1(x) = Q_0(x) \text{ and } \mu_1(x) = \mu_0(x)$$

where

$$R_0(x) = x^6, Q_0(x) = S(x)$$

$$\lambda_0(x) = 0 \text{ and } \mu_0(x) = 1$$

Note that $\gamma_i(x)$ and $\eta_i(x)$ are not needed in the computation of $\lambda(x)$. Thus,

$$R_1(x) = \alpha^{12}x^5 + \alpha^6x^4 + \alpha^{12}x^3 + \alpha^{14}x^2 + \alpha^{14}x$$

$$\lambda_1(x) = x, Q_1(x) = S(x), \text{ and } \mu_1(x) = 1$$

To implement this operation, let the coefficients of $R_0(x)$, $Q_0(x)$, $\lambda_0(x)$ and $\mu_0(x)$ enter the first cell of the systolic array in the manner shown in Fig. 3.

The quantities $d(R_0)$ and $d(Q_0)$ are integers representing the degrees of $R_0(x)$ and $Q_0(x)$, respectively. The "start" signal is a timing signal used to indicate the beginning of the polynomials, i.e., the leading coefficients. Note that to compute $R_1(x) = \alpha^7 R_0(x) + xQ_0(x)$ from $R_0(x)$ and $Q_0(x)$ term by term, the leading coefficients of $R_0(x)$ and $Q_0(x)$ are aligned with the start signal. Such an arrangement corresponds to the simultaneous entry of $R_0(x)$ and $xQ_0(x)$ into cell 1. The inputs $\lambda_0(x)$ and $x\mu_0(x)$ to cell 1 are aligned with the start signal in a similar manner.

The functional block diagram of cell 1 is shown in Fig. 4(a). The start signal, as well as $xQ_0(x)$ and $x\mu_0(x)$, are delayed by one time unit in such a manner that the leading coefficients of $R_1(x)$, $Q_1(x)$, $\lambda_1(x)$ and $\mu_1(x)$ are properly initiated by the start signal at the output of cell 1.

Step 2. Since $1_1 = \deg(R_1(x)) - \deg(Q_1(x)) = 0, \sigma_1 = 1$. Thus,

$$R_2(x) = \alpha^7 R_1(x) + \alpha^{12} Q_1(x)$$

$$= \alpha^{10}x^4 + \alpha^7 x^3 + \alpha^5 x^2 + \alpha + \alpha^{11}$$

$$\lambda_2(x) = \alpha^7 \lambda_1(x) + \alpha^{12}\mu_1(x) = \alpha^7 x + \alpha^{12}$$

$$Q_2(x) = Q_1(x) \text{ and } \mu_2(x) = \mu_1(x)$$

The implementation of cell 2, shown in Fig. 4(b), is essentially identical to cell 1.

Step 3. From cell 2, $d(R_2) = 4$ and $d(Q_2) = 5$. Thus,

$$1_2 = -1 < 0 \text{ and } \sigma_2 = 0$$

Hence, by Eq. (17),

$$R_3(x) = \alpha^{10} Q_2(x) + \alpha^7 x R_2(x)$$

$$= \alpha x^4 + \alpha^{13} x^3 + \alpha^{11} x^2 + \alpha x + \alpha^9$$

$$\lambda_3(x) = \alpha^{10}\mu_2(x) + \alpha^7 x \lambda_2(x)$$

$$= \alpha^{14}x^2 + \alpha^4 x + \alpha^{10}$$

$$Q_3(x) = R_2(x) \text{ and } \mu_3(x) = \lambda_2(x)$$

In Fig. 4(c), the role-switching operation is implemented by a simple crossover at the inputs of cell 3. Following the switching operation, cell 3 is identical in operation to cells 1 and 2.

Step 4. Since $d(R_3) = 4$ and $d(Q_3) = 4$, $1_3 = 0$ and $\sigma_3 = 1$. Thus,

$$R_4(x) = \alpha^{12} R_3(x) + \alpha Q_3(x) = 0x^3 + 0x^2 + \alpha^9 x + \alpha^6$$

$$\lambda_4(x) = \alpha^{10}\lambda_3(x) + \alpha\mu_3(x) = \alpha^9 x^2 + \alpha^6 x + \alpha^7$$

$$Q_4 = Q_3(x) \text{ and } \mu_4(x) = \mu_3(x)$$

The implementation of this step is presented in Fig. 4(d), which is again identical to cells 1 and 2. Note that $d(R_4) = 3$ although the actual degree of $R_4(x)$ is only 1. The reason for this is that $R_4(x)$ can be viewed as a polynomial of degree 3 even though the coefficients of x^3 and x^2 are both zero. Such an arrangement reduces the degree of polynomials regularly by one as the polynomials propagate from cell to cell. At the next stage one will see how zero leading coefficients are treated.

Step 5. Since $d(R_4) = 3$ and $d(Q_4) = 4$, $1_4 = 3 - 4 = -1$ and $\sigma_4 = 0$. However, the computation cannot continue to obtain $R_5(x)$ from $R_4(x)$ and $Q_4(x)$, term-by-term, since the leading coefficient of $R_4(x)$ is zero. Hence the proper treatement is to first assign

$$R_5(x) = Q_4(x), Q_5(x) = R_4(x)$$

$$\lambda_5(x) = \mu_4(x), \mu_5(x) = \lambda_4(x)$$

and to then reduce the degree of $Q_5(x)$ by one and pass the result on to the next cell. That is, let

$$Q_5(x) = 0x^2 + \alpha^9 x + \alpha^6 \text{ and } d(Q_5) = 3 - 1 = 2$$

In this particular example note that after the reduction, $d(Q_5) = 2 < t = 3$. Hence the systolic array should stop calculation at this point and accept $\mu_5(x)$ as the final result, namely,

$$\lambda(x) = \mu_5(x) = \alpha^9 x^2 + \alpha^6 x + \alpha^7 \qquad (18)$$

The functional diagram for the implementation of this step is shown in Fig. 4(e). At the output, the start signal's 1 points to the leading coefficient of 0 of $\lambda(x) = 0x^3 + \alpha^9 x^2 + \alpha^6 x + \alpha^7$.

Since the syndrome polynomial can have degree at most $2t - 1$ and each cell in this modified algorithm reduces the degree of one of the input polynomials by one, it will take $2t$ cells in the worst case to produce an output polynomial of degree less than t. Hence, only $2t$ such cells are needed in the RS decoder. If the algorithm actually concludes after the i^{th} cell, then the $2t - i$ subsequent cells will pass the polynomials unchanged as did the 6^{th} cell in the above example (see Fig. 4(f).

The function for a typical cell of this array is described by the flowchart in Fig. 5. The architecture of the cell is given in Fig. 6. The above systolic array design has been simulated completely on a computer, and it has been validated.

V. The Recursive Circuit for Computing the Remaining Transform Error Pattern

To compute the remaining transform error pattern of an (15, 9) RS code, Eq. (18) is multiplied by α^{-9} to yield

$$\alpha(x) = x^2 + \sigma_1 x + \sigma_2$$

where

$$\sigma_1 = \alpha^{12} \text{ and } \sigma_2 = \alpha^{13}$$

Thus, by Eq. (4), the recursive equation is

$$E_{j+2} = \alpha^{12} E_{j+1} + \alpha^{13} E_j \, (j = 5, \cdots, 13) \qquad (19)$$

The circuit for implementing Eq. (19) is shown in Fig. 7. In Fig. 7, registers R_1 and R_2 are initialized to contain E_5 and E_6, respectively. Thus $E_7 = \alpha^0$, $E_8 = 0$, $E_9 = \alpha^{13}$, $E_{10} = \alpha^{10}$, $E_{11} = \alpha^8$, $E_{12} = \alpha^4$, $E_{13} = \alpha^{11}$, $E_{14} = \alpha^0$, and $E_{15} = E_0 = \alpha^8$ are obtained sequentially. The same circuit extends naturally to the general case given in Eq. (4) for an (N, I) RS decoder over $GF(2^m)$.

VI. A VLSI Design for Computing the Error Pattern

By Eq. (5), the inverse transform of the error pattern obtained in the last section is

$$e_k = \sum_{n=0}^{14} E_n \alpha^{-nk}, \quad 0 \leqslant k \leqslant 14$$

or

$$e_k = (\ldots((E_{14}\alpha^k + E_{13})\alpha^k + E_{12})\alpha^k + \cdots)\alpha^k + \cdots E_0)$$

$$= (\ldots(((\alpha^0 \alpha^k + \alpha^{11})\alpha^k + \alpha^4)\alpha^k \cdots)\alpha^k \cdots)\alpha^k + \cdots \alpha^8) \qquad (20)$$

The VLSI technique for computing Eq. (20) in a (15, 9) RS code over $GF(2^4)$ is similar to that used in the VLSI design for computing syndromes. The number of basic cells needed to compute Eq. (20) is 15. The pipeline structure for computing Eq. (20) is given in Figure 8. The desired result is given in Eq. (9). The same circuit extends to a VLSI design for computing Eq. (5) in a general (N, I) RS code over $GF(2^m)$.

References

1. Kummer, H., *Recommendation for Space Data System Standards: Telemetry Channel Coding*, Issue-1, Consultative Committee for Space Data Systems, Panel 1: Telemetry, Tracking, Command, September, 1983.

2. Miller, R. L., Deutsch, L. J., and Butman, S. A., *On the Error Statistics of Viterbi Decoding and the Performance of Concatenated Codes*, Publication 81-9, Jet Propulsion Laboratory, Pasadena, California, September 1, 1981.

3. Berlekamp, E. R., "Bit-Serial Reed-Solomon Encoders," *IEEE Transactions on Information Theory*, Vol. IT-28, Number 6, November 1982.

4. Truong, T. K., Deutsch, L. J., Reed, I. S., Hsu, J. S., Wang, K., Yeh, C. S., "The VLSI Design of a Reed-Solomon Encoder Using Berlekamp's Bit Serial Algorithm," *Proceedings of the Third Caltech Conference on VLSI*, California Institute of Technology, Pasadena, California, 1983.

5. Brent, R. P., and Kung, H. T., *Systolic VLSI Arrays for Polynomial GCD Computations*, CMU Computer Science Department Report, Carnegie-Mellon University, Pittsburgh, Pennsylvania, 1982.

6. Reed, I. S., Scholtz, R. A., Truong, T. K., and Welch, L. R., "The Fast Decoding of Reed-Solomon Codes Using Fermat Theoretic Transforms and Continued Fractions," *IEEE Transactions on Information Theory*, Vol. IT-24, pp. 100-106, 1978.

7. McEliece, R. J., *The Theory of Information and Coding*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1977.

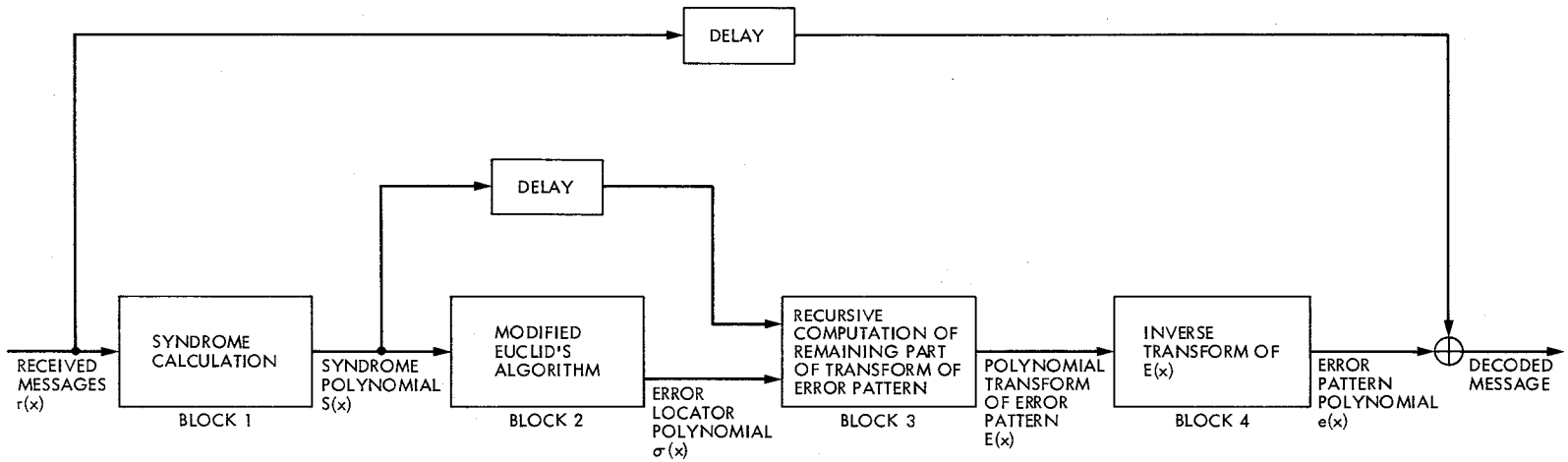


Fig. 1. The overall block diagram of A pipeline (*N*, *I*) RS decoder

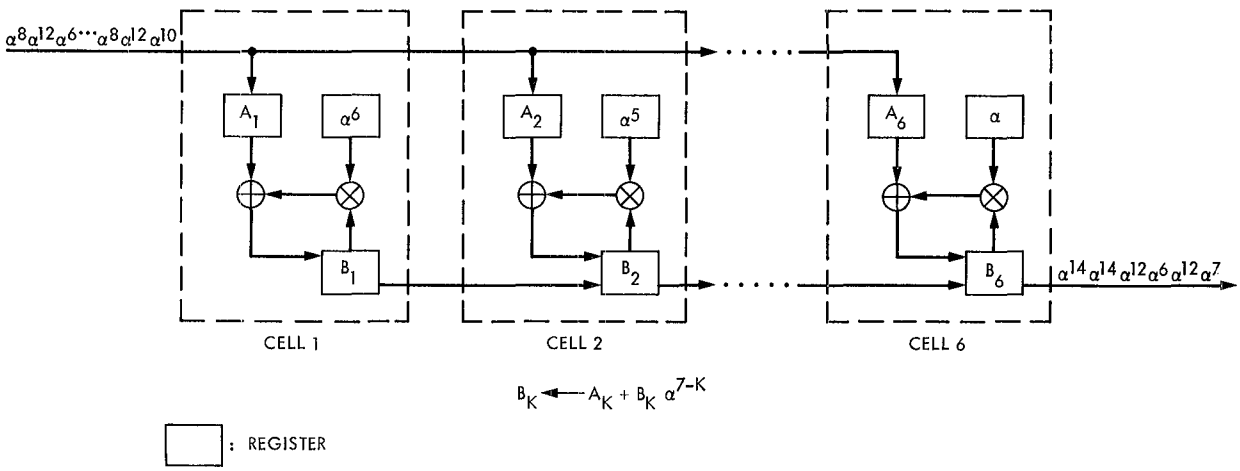


Fig. 2. The systolic array to compute syndromes of a (15, 9) RS code over *GF*(2^4)

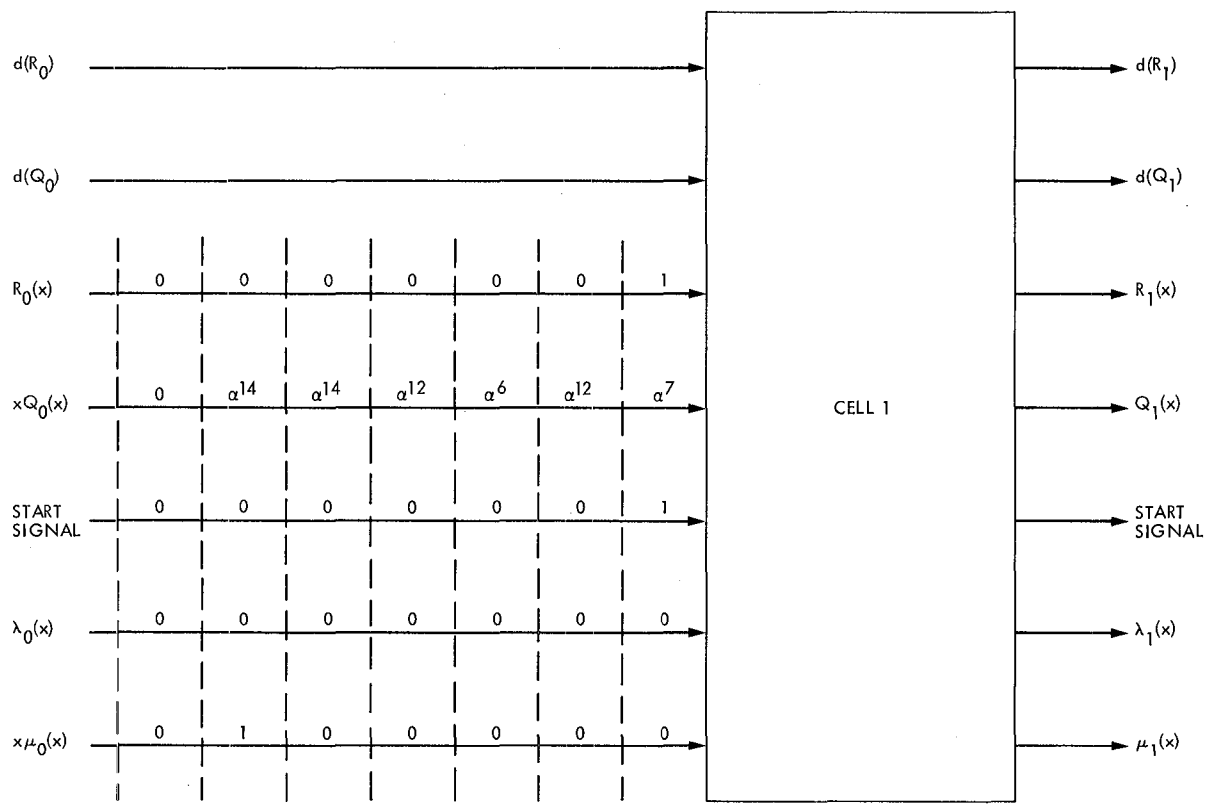


Fig. 3. The timing sequence of the input data

This Page Intentionally Left Blank

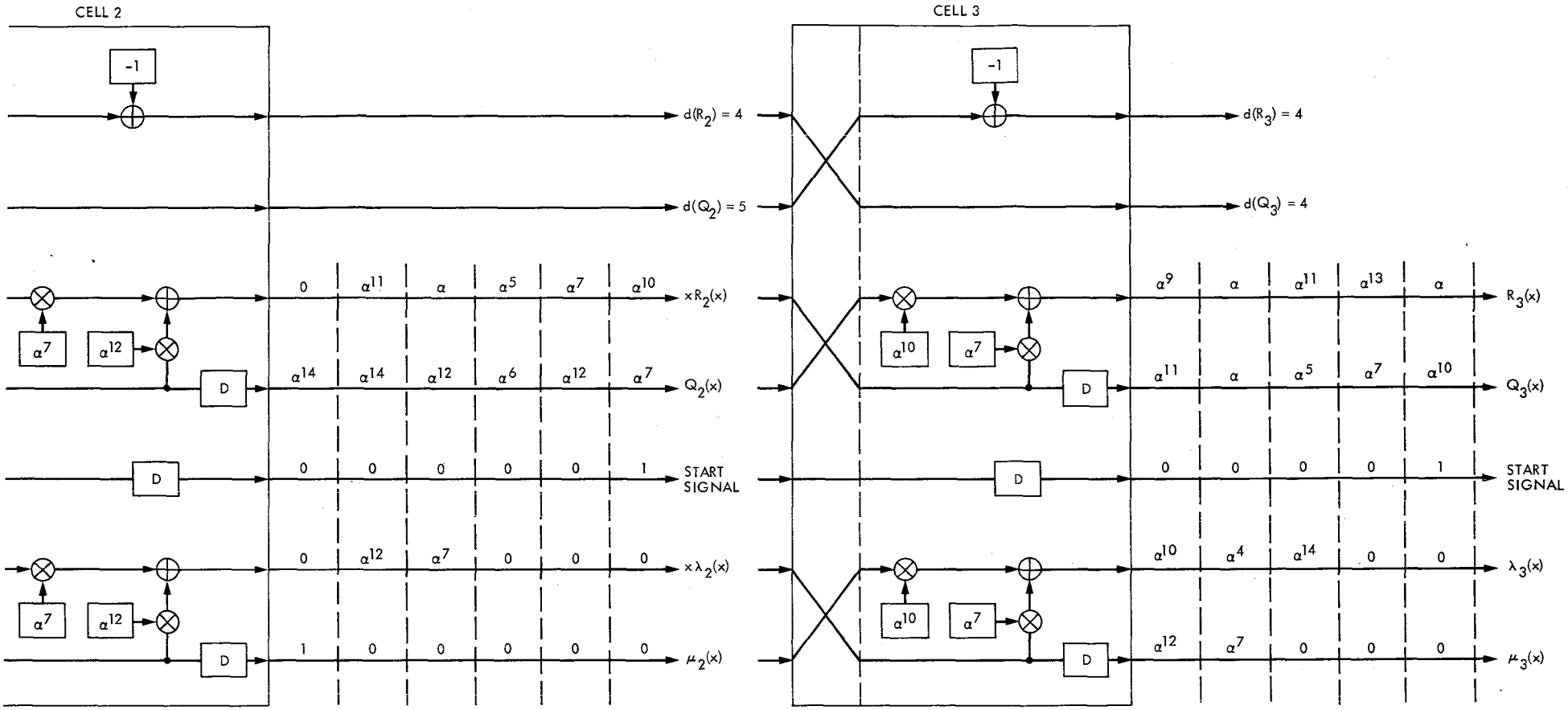


Fig. 4. The modified GCD algorithm for the example in the text

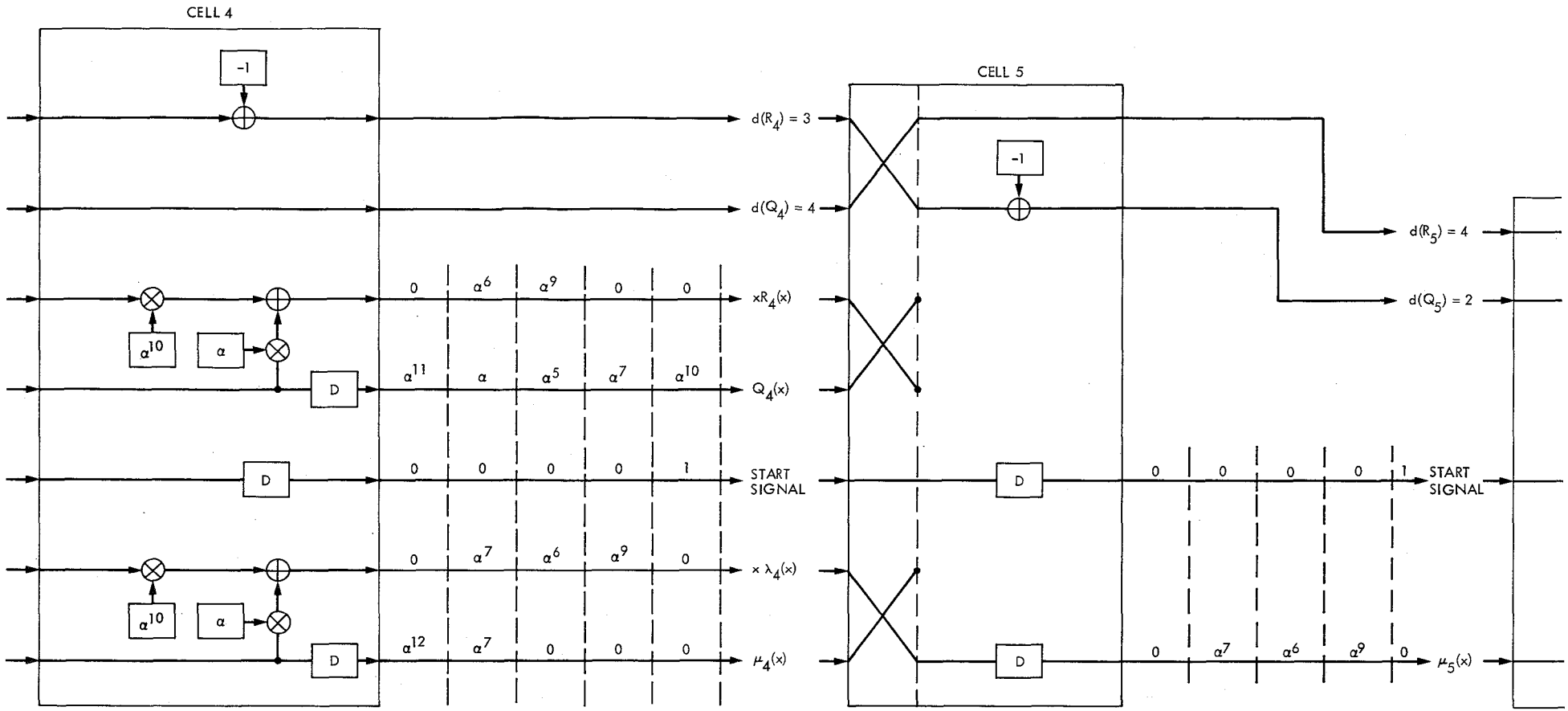

CELL 4
CELL 5
-1
D
α^{10}
α
$d(R_4) = 3$
$d(Q_4) = 4$
$xR_4(x)$
$Q_4(x)$
START SIGNAL
$\times \lambda_4(x)$
$\mu_4(x)$
$d(R_5) = 4$
$d(Q_5) = 2$
$\mu_5(x)$

Fig. 4 (contd)

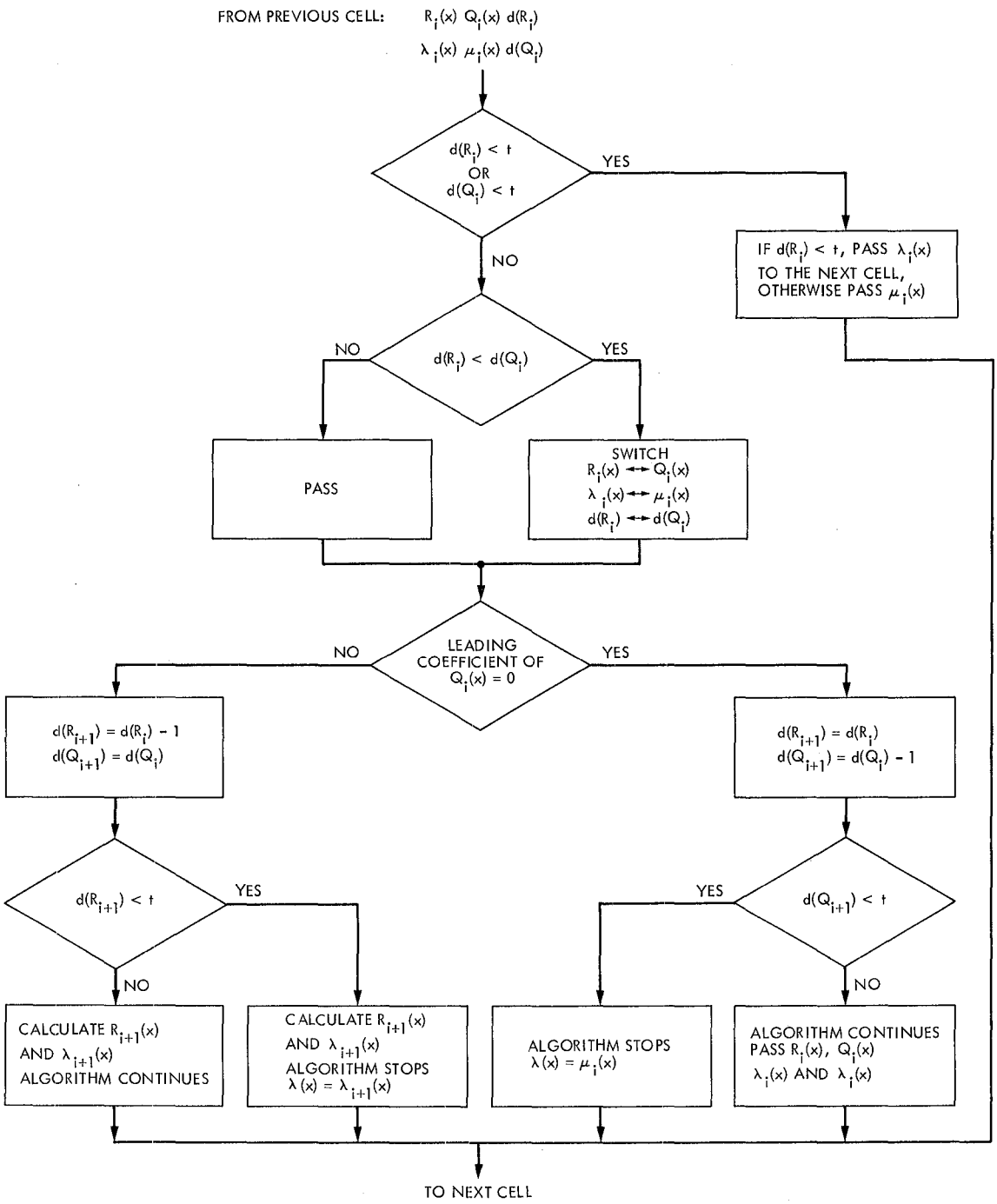


Fig. 5. The flowchart of the function of each cell

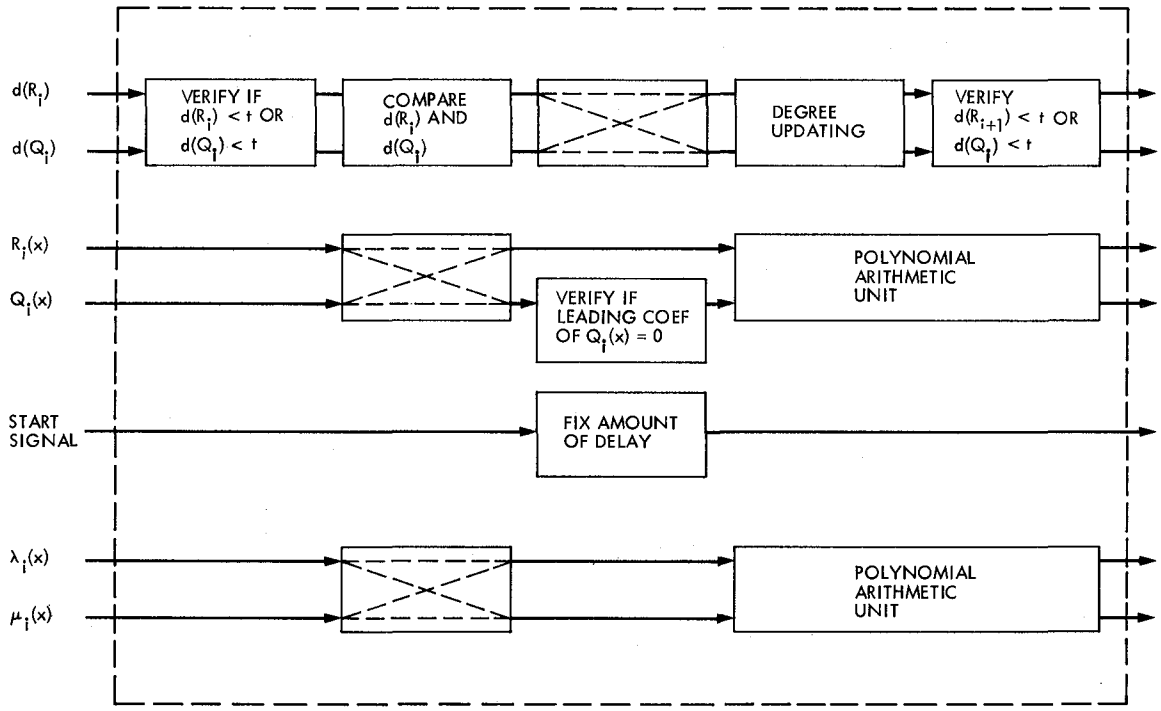


Fig. 6. The architecture of one cell of a systolic array to compute polynomial with modified Euclid's algorithm

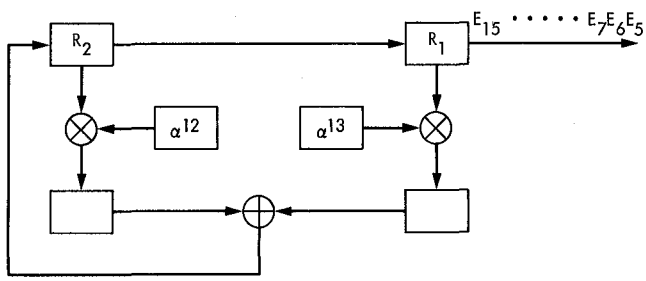


Fig. 7. A recursive circuit for computing the transform error pattern

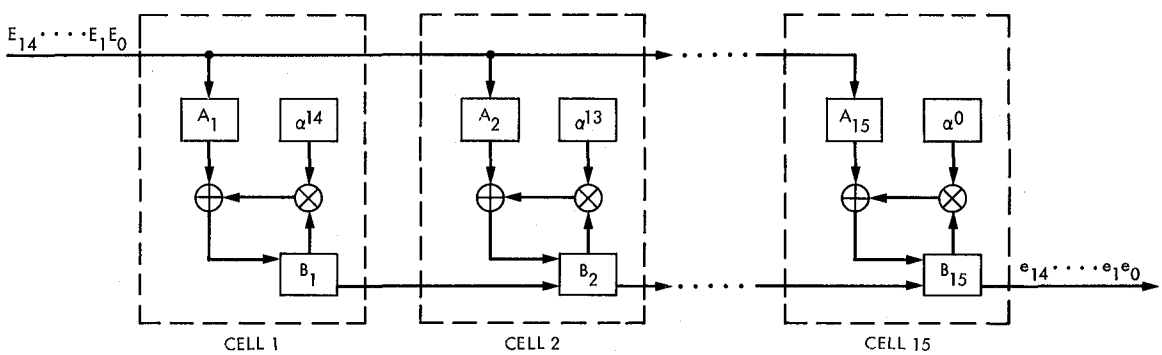


Fig. 8. The pipeline structure for computing error pattern in a (15, 11) RS code

Appendix A

The Construction of GF (2^4)

In this appendix, the construction of a finite field of 2^4 elements is given. To construct this field consider the irreducible polynomial $x^4 + x + 1$ over $GF(2)$. Let $\alpha \epsilon GF(2^4)$ be a root of $x^4 + x^2 + 1$, i.e., $\alpha^4 + \alpha^2 + 1 = 0$. The 15 nonzero field elements are given in Table A-1 in terms of root α.

Table A-1. Representations of the elements of $GF(2^4)$ generated by $\alpha^4 + \alpha + 1 = 0$

Element	α^3	α^2	α	α^0
α^0 =	0	0	0	1
α^1 =	0	0	1	0
α^2 =	0	1	0	0
α^3 =	1	0	0	0
α^4 =	0	0	1	1
α^5 =	0	1	1	0
α^6 =	1	1	0	0
α^7 =	1	0	1	1
α^8 =	0	1	0	1
α^9 =	1	0	1	0
α^{10} =	0	1	1	1
α^{11} =	1	1	1	0
α^{12} =	1	1	1	1
α^{13} =	1	1	0	1
α^{14} =	1	0	0	1

ISEE-3 Microwave Filter Requirements

J. L. Galvez, H. Marlin, and P. Stanton
Radio Frequency and Microwave Subsystem Section

The 64-m subnet is committed to support the ISEE-3 spacecraft. The uplink and one of the downlink frequencies will be, respectively, 2090 and 2217 MHz. As these two frequencies fall outside the normal DSN transmit and receive bands, the 64-m antennas present new filter requirements, which will be analyzed here.

I. Introduction

The frequencies used to support ISEE-3 mission will be 2090 MHz for the uplink, 2270 and 2217 MHz for single and dual downlink.

Uplink power in excess of 20 kW (perhaps 150 kW) will be required at DSS-63 and at DSS-14. The S-Band Preamplifier Filter (SPF), which is a five-cavity waveguide filter, has to be retuned in order to prevent the power in the new transmitter frequency from going into the receiver path. Analysis shows that when using only one spacecraft transmitter, the signal level received is inadequate to attain the desired telemetry data rate; therefore, there is a requirement to provide dual downlink capability at DSS-14 and at DSS-63 when they return to operation in Mark IV configuration. As the existing Megawatt Transmitter Filter (MTF), which is a six-cavity waveguide filter, does not present any isolation at 2217 MHz, a new filter, designated MTF-II, has to be added in the transmitter path to stop the noise generated by the Klystron at that frequency.

Figure 1 shows the transmitter and receiver paths where the new filter requirements have to be met and the major components to be either added or modified.

II. S-Band Block V Maser

A new maser, S-Band Block V TWM, will be installed at DSS-14 and DSS-63. It will provide the capability of receiving a Left Circular Polarized (LCP) signal at 2217 MHz through the SPF. This signal will be combined with the Right Circular Polarized (RCP) signal at 2270 MHz to attain the adequate level of the telemetry data rate required.

The roll-off characteristic of the low frequency side of the new Maser bandwidth is shown in Fig. 2.

The threshold noise density for this Maser, for an estimated system noise temperature of 15 K, is

$$N_{th} = KT = -186.6 \text{ dBm/Hz}$$

III. Klystron Noise

The 20 kW Klystron generates an output noise with spectrum typified in Fig. 3. It can be seen that close to the Klystron passband the three-cavity roll-off effect dominates, while at frequencies far from the passband, a single cavity tends to become the dominant factor. The effective gain of the last cavity is -45 dB relative to the full Klystron gain at the center of the band; therefore, the remote noise spectrum can be modeled as a single-pole filtered white noise with a band center noise density 45 dB below that of the Klystron at band center. Thus the noise density is given by the following (see "Final RFI Report," Ford Aerospace and Communication Corporation, 333-561-ER-03, Dec. 7, 1981):

$$N(f) = KG\,(NF - 1)\,T_o A \frac{1}{(f/f_o)^2}$$

where

K = Boltzmann's constant (-198.6 dBm/Hz, K)

G = Klystron gain at small signal (56 dB)

NF = Tube noise figure ($\leqslant$40 dB)

T_o = Ambient temperature, K

$(NF-1)$ = 64.6 dB (Ref. 1)

A = Relative gain of single cavity (-45 dB)

f_b = Single cavity "corner" frequency (center frequency plus 21.1 MHz)

At 2217 MHz,

$$\left[A\frac{1}{(f/f_b)^2}\right]\text{dB} = -45 + 10\log\frac{21.1^2}{(2217 - 2215)^2} = -58.6\text{ dB}$$

Hence, the noise density is

$$N_{20\text{ kW, }2217\text{ MHz}} \simeq -198.6 + 56 + 64.6 - 58.6$$

$$\simeq -136.6\text{ dBm/Hz}$$

For the 100 kW transmitter this value can be scaled, in the worst case, to account for the difference between output powers. Therefore, the noise density from the 100 kW tube will be

$$N_{100\text{ kW, }2217\text{ MHz}} \simeq -136.6 + 7 = -129.6\text{ dBm/Hz}$$

IV. Transmitter Noise Filtering

From all the above it may be inferred that the minimum isolation needed to prevent the Klystron receiver-band noise from entering into the receiver path is

$$N_{100\text{ kW, }2217\text{ MHz}} - N_{th} = -129.6 - (-186.6) = 57\text{ dB}$$

Considering a typical safety margin of 25 dB, the isolation desired will be 82 dB.

In order to possibly eliminate the need for a new filter design, a spare MTF was retuned with the isolation curve and VSWR within the transmit band being measured. Figures 4 and 5 show the rejection of the MTF before and after it was retuned for 90 dB rejection over the 2200 to 2230 MHz band.

This exercise proved the feasibility of using the present MTF retuned to provide the isolation needed at the new receiver frequency. It also demonstrated that it has to be matched to reduce the transmit band VSWR, which increased when retuned. Note that the maximum measured isolation shown in Figs. 4 and 5 is limited by the 70 dB dynamic range of the instrumentation used. Additional isolation of some 25 to 30 dB between ports of the Orthomode will result in achieving the required total system isolation, including the 25 dB safety factor.

V. Transmitter Power Filters

The S-Band Block V Maser also has to be isolated from the transmitter output at 2090 MHz, which for the 100 kW Tube is 80 dBm.

This function will be provided by the existing S-Band Preamplifier Filter (SPF), after being retuned for 60 dB rejection over the 2090 to 2120 MHz band, and by the Orthomode.

Figure 6 illustrates the rejection before and after the SPF was retuned. The SPF insertion loss and input VSWR in the receiver bandpass were not changed as a result of the retuning.

VI. Conclusion

As a result of the previous considerations, the following modifications are being implemented:

(1) Install a new filter, in addition to the existing MTF, between the transmitter and the diplexer. This filter is the same design as the present MTF but tuned to

provide more than 90 dB of isolation at 2217 MHz. This filter (designated the MTF-II) will be matched to present a VSWR low enough so it does not disturb the transmitter path.

(2) Retune the existing SPF to obtain a minimum of 60 dB of isolation at 2090 MHz.

With these two changes, plus the addition of a matching section at the diplexer transmitter port (the subject of a later report), the transmitter and receiver paths of the 64-m antennas will meet the ISEE-3 requirements. Figure 7 compares the receiver and transmitter bandwidths and the MTF-II and SPF isolation curves. Figure 8 is a layout of the S-Band Polarization Diverse (SPD) cone components.

References

1. Krauss, John D., *Radio Astronomy*, McGraw-Hill, 1966.

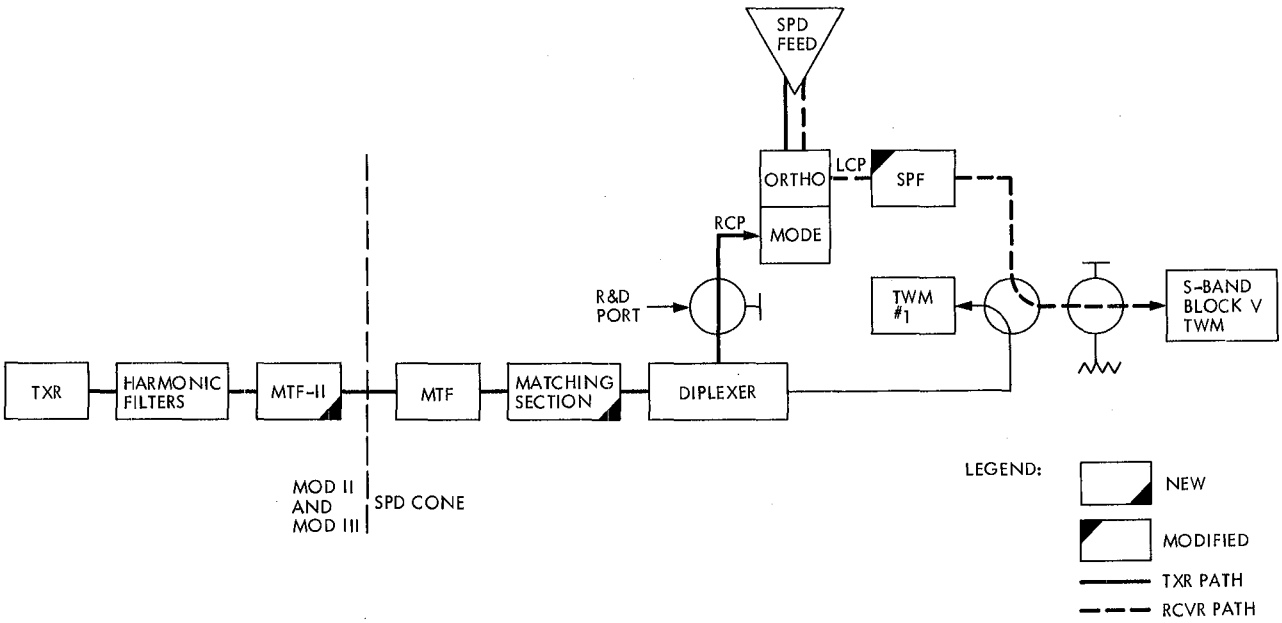


Fig. 1. 64-meter antenna: ISEE-3 transmit and receive paths

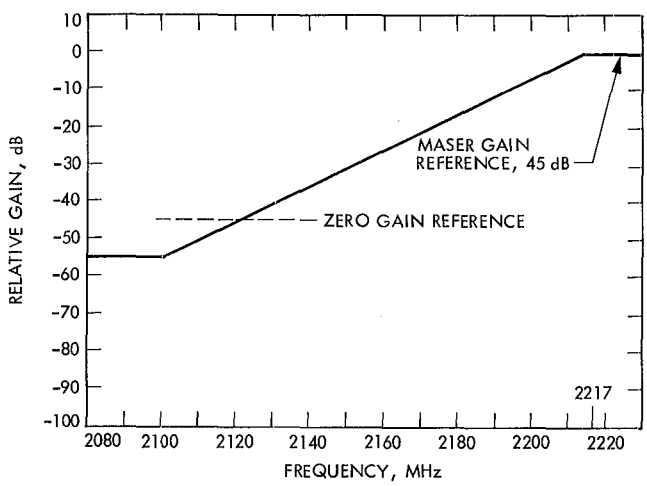


Fig. 2. S-band Block V TWM bandwidth

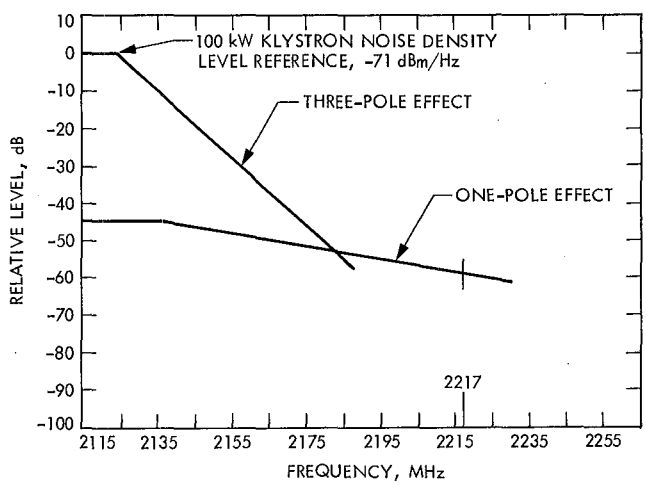


Fig. 3. Klystron noise characteristic

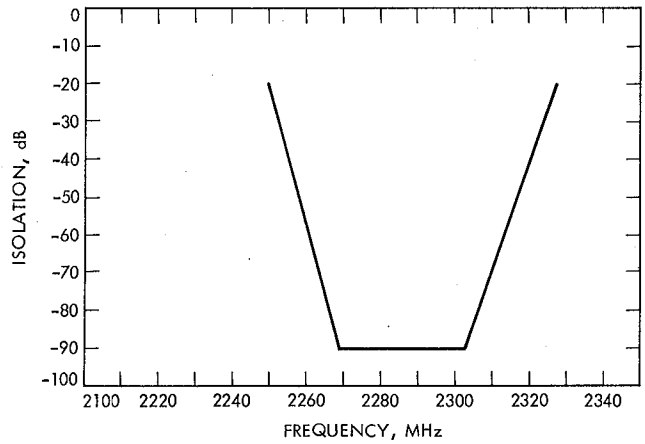


Fig. 4. MTF isolation before retuning

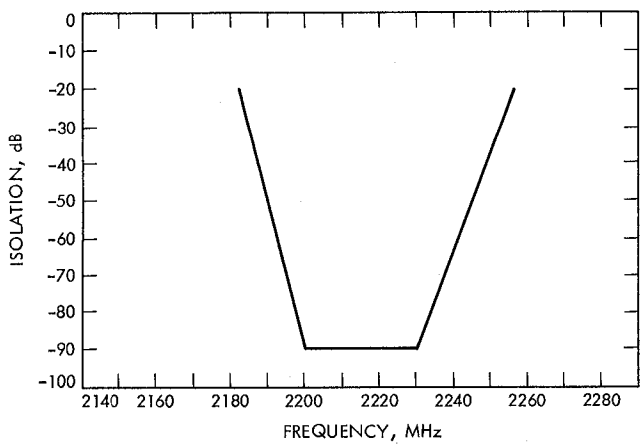


Fig. 5. MTF isolation from 2200 to 2230 MHz

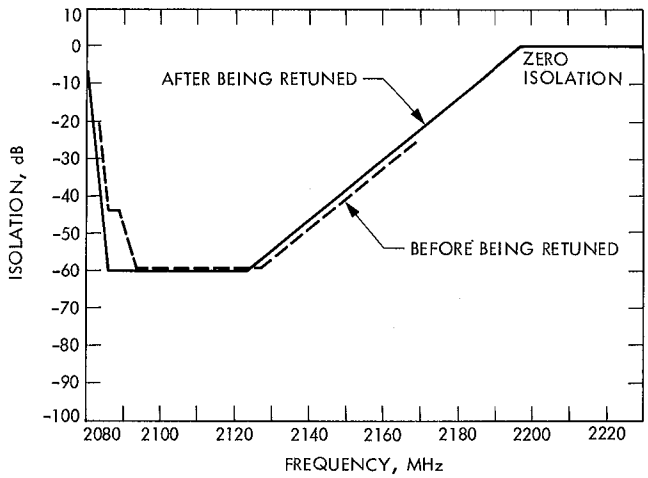


Fig. 6. SPF isolation curve

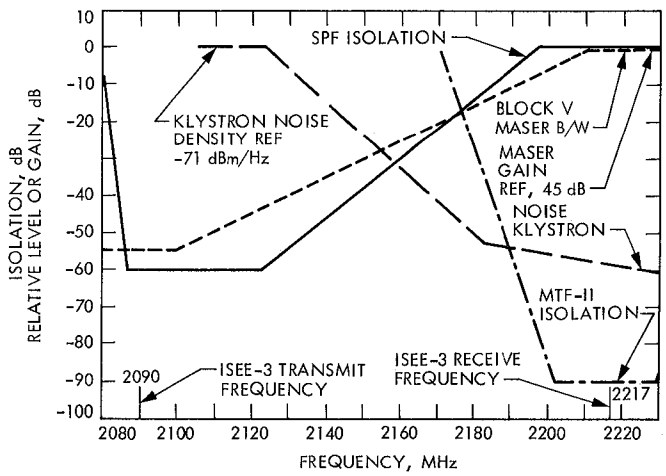


Fig. 7. Combination of transmit and receive bandwidths and isolation of SPF and MTF-II

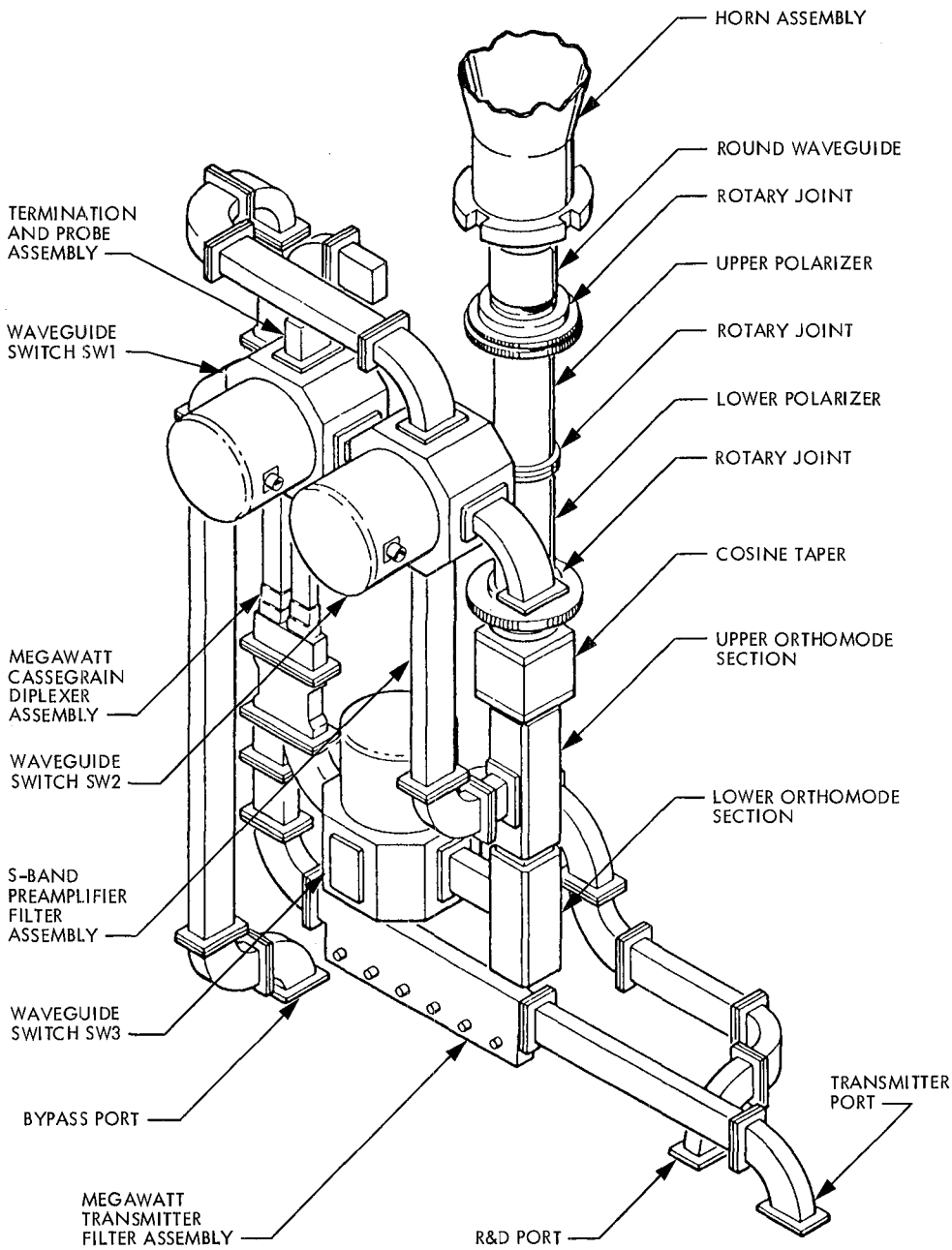


Fig. 8. SPD cone assembly microwave components

Telemetry Simulation Assembly Implementation in the DSN

M. E. Alberda
DSN Data Systems Section

An improved telemetry simulation capability is being implemented, starting in 1983, as part of the Mark IV A Network Implementation Project. The Telemetry Simulation Assembly (TSA) is replacing the Simulation Conversion Assembly (SCA) throughout the DSN. The development of the TSA is discussed, and the design is described to the block diagram level. The testing and evaluation completed so far are summarized.

I. Introduction

Beginning in late 1983, a new telemetry simulation capability is being implemented as a part of the Mark IV A Network Implementation Project throughout the Deep Space Network (DSN). The Telemetry Simulation Assembly (TSA) is a major assembly within the DSN Test Support Subsystem (DTS), and it is required to support testing requirements associated with an expanded set of spacecraft missions.

The design implemented provides simulated spacecraft and related reference telemetry test signals that are routed to other DSN subsystems, as shown in Fig. 1.

The TSA is a computer controlled data generation and transmission system that is based on the Modular Computer Systems Classic 7845 minicomputer. Also included are various peripheral equipment and the JPL designed Signal Generation Assembly (SGA) equipment. The SGA equipment consists of large plug-in circuit boards, an output amplifier unit, a power supply, a cable interface panel, and multiple external function generators.

The TSA was developed to replace the existing Simulation Conversion Assembly (SCA) currently in use throughout the DSN.. The SCA contains a limited capability computer (XDS-910) that is virtually obsolete based on today's technology. The TSA provides for very much higher data and symbol rates, more extensive encoding capability, and greatly increased storage and memory capacity, including disc and random access memory (RAM).

Following an in-house prototype fabrication and evaluation effort, a production contract was awarded to an electronics manufacturer in 1982 covering four systems of SGA equipment plus extensive spares and documentation. This equipment was successfully completed and acceptance tested in 1983. Then two complete systems were delivered to the DSN Systems Integration Facility (SIF) where systems integration and testing are in progress. The balance of two systems plus all spares will be delivered to the SIF in early 1984.

This report describes the equipment built and summarizes the testing completed to date. Lab and acceptance testing have

been done by utilizing a test program developed for this purpose. The test program is loaded into the computer from disc, and operator commands are entered from a local terminal. This terminal is provided for lab use only and has not been provided for use in the operational mode, where control will be remote, from another subsystem.

A TSA operational program is under development and is currently in the development test phase. A description of this program is not included in this report.

II. Functional Characteristics

The missions supported are listed in Table 1, along with the telemetry characteristics associated with each mission. However, significantly higher capability was incorporated in the new TSA above that needed for the new mission set, including added output channels and delay sections; much higher data and symbol rate capability (up to 2 M b/s or 4 M s/s) was also incorporated. This was done to satisfy anticipated requirements associated with potential new missions, such as Venus Radar Mapper (VRM).

The TSA was designed to be remotely operated and monitored, with local operation to be done only for checkout and maintenance. The remote operation is accomplished from a manual console, which is a part of the Link Monitor and Control Subsystem (LMC). This unit also contains an MC7845 minicomputer which is linked with several other subsystem minicomputers via a Local Area Network (LAN). These subsystems are located in a Signal Processing Center (SPC), which is part of a Deep Space Communications Complex (DSCC). Three DSCC's are contained within the DSN, one at Goldstone, California, one in Spain, and one in Australia.

Six independent data output channels have been provided, all operating simultaneously. Each channel can operate in either of two modes – a computer data source mode or a self-contained data generation mode. Each channel provides data formatting (NRZ-L, NRZ-M, or Biphase-L), data encoding (either long- or short-constraint length codes), and frame-length counting with coder-reset capability.

Six delay sections have been provided in order to support antenna array testing. In the DSN, either a single antenna or a group (two or more) of arrayed antennas can be used to track one spacecraft. Arraying provides a significant gain in signal-to-noise ratio (SNR) over that available from a single antenna. If arrayed, a signal reaches each antenna at a slightly different time due to the physical distance between antennas. The TSA generates and sends a test-signal stream to the Test Transmitter (TT) associated with each antenna; the test-signal stream is delayed a different magnitude for each antenna in order to simulate the difference in arrival time of a signal at each antenna. A fixed spacecraft position is assumed for each array test, and a delay resolution down to 0.1 μs is provided by each delay.

III. Detailed Hardware Design Description

A brief summary of the subsystem design is presented below.

A. Subsystem Block Diagram

The interfaces to other subsystems are shown in Fig. 2. These consist of the following:

(1) Local Area Network (LAN) ties in with other computers, e.g., the Link Monitor (LMC) and the Complex Monitor (CMC). Data, command, and status transfers are transmitted between selected computers on the bus, as required.

(2) The TSA receives multiple inputs of a 10 MHz precision reference from the Frequency and Timing Subsystem (FTS). These references are routed to each one of the 12 function generators and also to the SGA planes containing the digital channels and the delay sections. Also received from the FTS are 1, 10, 100, and 1000 pulses/s timing signals that are used by the computer for real-time interrupt generation, as required.

(3) The TSA provides up to seven independent output baseband signals, one signal to each Test Transmitter (or Test Translator) via an Exciter, contained within a Receiver-Exciter subsystem associated with each antenna (see Fig. 1). Each TSA signal consists of a square wave subcarrier onto which data (or symbols) have been bi-phase modulated. The subcarrier is also amplitude controlled to represent a desired mod index attenuation for a given test.

(4) The TSA provides multiple reference output signals (symbols, symbol clock, data, and modulated subcarrier) to telemetry link assemblies where error rates are determined. These are Baseband Assembly (BBA), Subcarrier Demodulator Assembly (SDA), Symbol Synchronizer Assembly (SSA), and Telemetry Processor Assembly (TPA).

The subassembly units contained within the TSA consist of the 7845 computer with peripherals, the SGA circuit planes which plug into a Peripheral Controller Enclosure (PCE) associated with the computer, a removable cartridge disc, an output amplifier unit (OAU), twelve function generators, and a power

supply. This equipment is contained in four racks as shown in Fig. 3.

B. TSA Internal Organization

The computer is used to remotely control and monitor all equipment within the TSA, as shown in Fig. 4. The function generators (FGs) are remotely controlled and monitored via an IEEE-488 bus driven from a controller card located in the IOIS section of the computer. Waveform (square wave), frequency, phase angle, amplitude, and DC offset are set by computer commands. Self-test of each FG can also be commanded, and the result is fed back to the computer via the bus.

Control words are sent out from the computer and loaded into holding registers that provide groups of control bits, as required, to control all operating features of each digital channel, delay section, and modulator-output amplifier section, including signal routing in and/or out of each of those circuits.

Long-loop simulation capability is provided for data rates up to 30 Kb/s. Data are received at an SPC via a ground communications link and sent to the TSA via the LAN. The TSA stores, formats, and sends the data to the SGA where the modulated signal is used to test the end-to-end Telemetry System.

The computer and its I/O configuration are shown in Fig. 5, where each line represents a plug-in circuit board. An I/O Processor, called the internal IOP, is provided on the CPU board. Another IOP, called the external IOP, is provided on a separate board. A separate bus is provided for each of these IOPs, and maximum data transfer rates are as shown. The SGA planes are loaded from the external IOP when operating one (or more) digital channels in the computer mode. An aggregate digital channel output rate of 6 Mb/s has been demonstrated, i.e., 1 channel at 6 Mb/s or 2 channels at 3 Mb/s each, etc., using test software. The SGA planes have been built from partially populated circuit planes, called General Purpose Interface Controllers, with custom circuitry (about 400 ICs) added on each plane.

C. Digital Channel Description

Six digital channels are provided, three on each of two SGA-1 type boards. Fig. 6 shows the functions and data flow, in block diagram form, for a typical channel. In the computer mode, each of two FIFO memory buffers is alternately loaded from the computer in a burst mode, up to 64 16-bit words per FIFO. Each FIFO is alternately emptied into a parallel-to-serial converter, where a continuous serial data bit stream is generated. When one FIFO is empty, data are obtained from the other FIFO, while at the same time a Data Request is sent to the computer to refill the empty FIFO. In the other data source mode, simulated data are generated by hardware contained within each channel. Any one of the following data forms is available: square wave (1, 0, 1, 0, . . .), all 1's, all 0's, a repetitive pseudo-noise (PN) sequence 2^{11} -1 bits long, or an operator-selected repetitive pattern up to 6-bytes long. Then the serial data stream, selected from any of the desired sources, is routed to a data format generator (NRZ-L, NRZ-M, or bi-phase L) and then to a selected encoder, if coding is required, where symbol bits (symbols) are generated. A separate FG provides the clock input for each channel, from which symbol and data clocks are derived. A frame counter is also provided that generates a signal to reset the encoder shift register at the end of each frame, if required by the selected code. The outputs of each digital channel are data, symbols, data clock, and symbol clock.

The outputs from any one digital channel can be routed to any one (or more) delay section by computer control of the multiplexer switching circuitry provided.

D. Delay Section Description

As mentioned previously, the sole purpose for the delay sections (Fig. 7) is to support arrayed antenna testing. When non-arrayed tests are conducted, the delay sections serve no function. However, the digital channel outputs are routed through the delay section anyway enroute to the modulator-amplifier and reference output sections, with essentially zero delay introduced (nominally 0.4 μs maximum).

For array testing, outputs from only one digital channel are routed to multiple delay sections (one delay section per arrayed antenna), and different magnitude delay intervals are generated by each delay section, depending on the assumed spacecraft position and the known physical location of each antenna. Delay interval D1 is provided to simulate the delay in arrival time of a spacecraft signal to an antenna which is farther away from the spacecraft than is the antenna in the array group nearest to the spacecraft. Delay D2 is provided to simulate the time delay required for the signal transmission delay through the cable from an antenna to the Signal Processing Center (SPC). Delay D3 is provided to simulate the processing time required for the Baseband Assembly (BBA) to combine its multiple baseband inputs and develop its single improved SNR output signal to the TPA.

Within each delay section, FIFO buffer memories are provided that store symbols and data for the required delay intervals. Symbol-clock and data-clock signals are also regenerated to simulate the required delays. One output from a delay section is a symbol (data) stream occurring after D1, which is routed to the modulator-amplifier section. The other outputs

from a delay section are reference outputs, occurring after D2 + D3, that are routed to the reference output section. The required delay intervals are developed within each delay section by counting down, at a 10 MHz rate, counters that have been initially loaded from the computer with a binary number corresponding to the desired delay interval. The reference outputs from only one delay section within an array group are used and routed to the other assemblies in a telemetry link, i.e., from the delay section with the longest total delay interval, DT. There are five variable and one fixed delay sections provided, totaling six.

E. Modulator-Amplifier Section

There are six bi-phase modulator (mod) sections provided. Each mod has two inputs, a symbol (data) stream from one of the six delay sections and a simulated subcarrier input from one of six FGs. The FG signal is fed to both positive and negative inputs of a high-speed operational amplifier (op amp) through high-speed switching transistors (FETs). One FET is conducting while the other is not. Thus, the op amp output is either inverted or noninverted, depending on which FET is turned on. One FET is controlled by noninverted symbols while the other FET is controlled by inverted symbols, derived from the same symbol stream. Thus, the op amp output provides the desired bi-phase modulation. The simulated subcarrier input amplitude is set, at the FG, to provide an output amplitude proportional to the desired mod-index value. Subcarrier phase is also set at the FG to simulate the desired D1 delay.

Each mod output is connected to one of seven output amplifier units through an 8-to-1 analog multiplexer switch. Each output amplifier is a high-speed, high-power hybrid op amp coaxial cable driver whose output is connected to an assigned Test Transmitter (TT), or Test Translator, in a Receiver-Exciter Subsystem. Sufficient speed and power are available from these amplifiers so that an excellent square wave shape is preserved and delivered to a TT up to a frequency of 5 MHz, through a 50 ohm coaxial cable up to 75 m long.

The analog multiplexer switches are provided to achieve overall improved availability so that any digital channel and any delay section can drive a specified TT without requiring any manual cable switching.

All modulators, analog switch multiplexers, and output amplifiers are contained on one large printed circuit board mounted in the Output Amplifier Unit (OAU).

F. Reference Output Section

The reference outputs from any one delay section can be routed, through digital multiplexer switching, to any one of four telemetry links in the SPC. There are two types of links utilized, one containing a dual channel BBA and a dual channel TPA, and the other type containing two SDAs, two SSAs and a dual channel TPA.

The reference outputs provided from a delay section are

(1) Modulated subcarrier (fixed level) to an SDA

(2) Symbols to an SSA

(3) Symbols, symbol clock, and modulated subcarrier (fixed level) to a BBA

(4) Data to a TPA

Three delay sections, a fixed delay section, a control register section, and the Reference output section are contained on an SGA-2 type circuit board.

G. Power Supply

The power supply unit provides + and - 5V and + and - 15V power to the OAU and + and - 15V power to the SGA boards, as required.

H. Overall Performance Characteristics

The variable ranges and resolution of the signals from digital channel and delay sections are indicated in Table 2.

The wave shape observed at the load end of a coaxial cable to a TT shows rise and fall times of about 10 to 15 ns.

IV. Testing and Evaluation

A. Acceptance Testing

The TSA Test Program was developed to support TSA testing both for individual subassemblies and for the complete system, operating in a typical telemetry link environment. For both types of testing, a Classic 7845 computer is assumed available to run the test program. A CRT terminal (VT100) is used for both types of testing, for manual entries and message display.

The test program provides a menu of commands that allows for selection and control of all operating features of each digital channel, delay section, function generator, and signal routing between all sections.

A prototype TSA system, located in a development laboratory, was used for initial checkout and acceptance testing of all SGA-1 and SGA-2 circuit boards as well as all power supply and output amplifier units. The Acceptance Test Procedure

(ATP) and test data sheets are included within the Operation and Maintenance Manual (JPL TM 08301).

The final portion of the ATP was completed, for each TSA system, using equipment located in the Verification Test Facility, where TSA system inputs and outputs were verified using telemetry-link equipment.

The prototype system has been used for TSA operational program development and testing, in addition to its use for hardware checkout and testing.

B. System Integration Facility (SIF) Testing

After all parts of the ATP were successfully completed, the TSA systems were delivered to the SIF where integration testing with other subsystems was completed. Two TSA systems have been delivered to the SIF so far, and one system to the VTF, which meets project schedules. The remaining system is fully tested and will be delivered in the near future, along with the spare subassemblies scheduled for delivery to each of three DSCCs and to the VTF.

C. First Installation Testing

After testing was completed at the SIF, the first TSA system was delivered to the Goldstone DSCC (SPC-10), along with the other subsystems that are parts of the Mark IV A project. After installation here, extensive TSA testing and evaluation will be performed while using the other required subsystems, including calibration procedures as required, for operating the TSA delay sections to support array testing.

V. Conclusion

The prototype development phase and the production fabrication and test phases of the TSA project have proceeded in an orderly fashion with relatively few changes being required. The effort has closely followed the original development plan, including cost, schedule, and the meeting of performance requirements.

Table 1. Telemetry characteristics

Mission Data Channel	PCM Type	Coding Characteristics	Data Rates, bits per sec	Subcarrier, kHz	Mod Index 0°	RF Band	Frame Size	Frame Sync
PIONEER 10/11								
1	NRZL	Uncoded	8-2048	32.768	66	S	192 or 384	
1	NRZL	Conv. K=32; R=1/2	8-2048	32.768	66	S	192 or 384	
PN-Venus								
1	NRZL	Uncoded	8-4096	16	37, 67	S		
1	NRZL	Conv. K=32; R=1/2	8-2048	16		S	512	$F8549_{(16)}$
Space Telescope Biϕ-L			4000					
Voyager 2								
Uranus 1	NRZL	Conv. K=7, R=1/2	10-29, 900	360	70	X	1792	
2	NRZL	Uncoded	40	360	55	X	2200	$03915ED3_{(16)}$
2	NRZL	Uncoded	40	22.5	55	S		
Galileo								
1	NRZL	Conv. K=7, R=1/2	7,680–134,400	360		X	5120, 1120	
1	NRZL	Conv. K=7, R=1/2	Up to 40,000	22.5/360	20 to 88	S	1920, 4480	$03915ED3_{(16)}$
2	NRZL	Uncoded	40	22.5		S	7680	
ISEE-C								
1	NRZL	Conv. K=24, R=1/2	64	1.024	40	S	1024	$FAF320_{(16)}$
1	Biϕ-L	Conv. K=24, R=1/2	512-2048	N/A	63	S		EC819FBE
ISPM-ESA 1	NRZL	Conv. K=7; R=1/2	1024-8192	131.072	32	X	1024	$EB90_{(16)}$
AMPTE-CCE								
1	Biϕ-L	Conv. K=7, R=1/2	101,000	NA		S		
2	NRZL	Uncoded	3300	16		S		
AMPTE-IRM								
1	Biϕ-L	Conv. K=7, R=1/2	1000-8000	NA	74	S		
1	NRZL	Conv. K=7, R=1/2	2048, 4096, 8192	131.072		S		
AMPTE-UKS								
1	Biϕ-L	Conv. K=7, R=1/2	1000-32000	NA				
GIOTTO	NRZL	Conv. K=7, R=1/2	20000	327.7				

Table 2. SGA performance characteristics

Parameter	Range	Resolution
Delays	1.0 - 537.6 μs	0.1 μs
Subcarrier	50 Hz - 5 MHz	0.01 Hz
Symbol rate (NRZ WO S/C)	6 SPS - 5 MSPS	0.01 Hz
Symbol rate (NRZ W S/S)	4 SPS - 5 MSPS	0.01 Hz
Symbol rate (BIO-L)	100 SPS - 5 MSPS	0.01 Hz
Data rate (7-1/2, 7-1/3 CR)	10 BPS - 1.7 MBPS	0.01 Hz
Data rate (24-1/2, 32-1/2 CR)	8 BPS - 2 MBPS	0.01 Hz

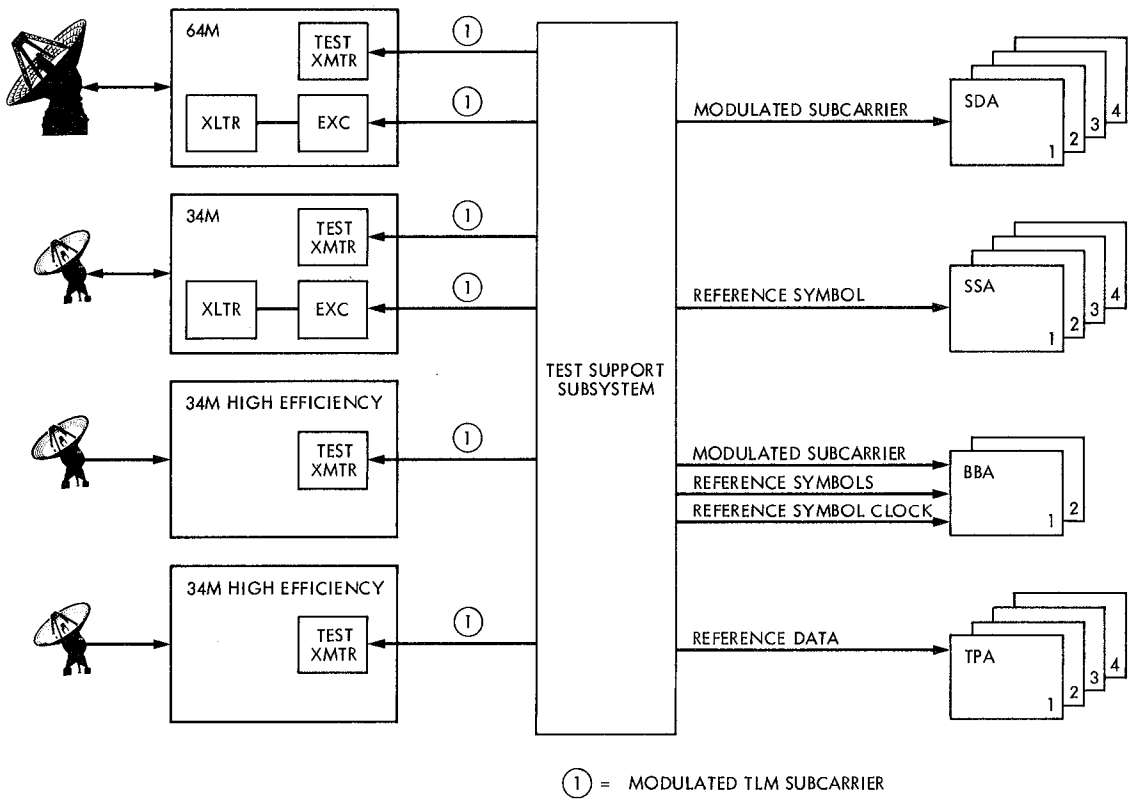
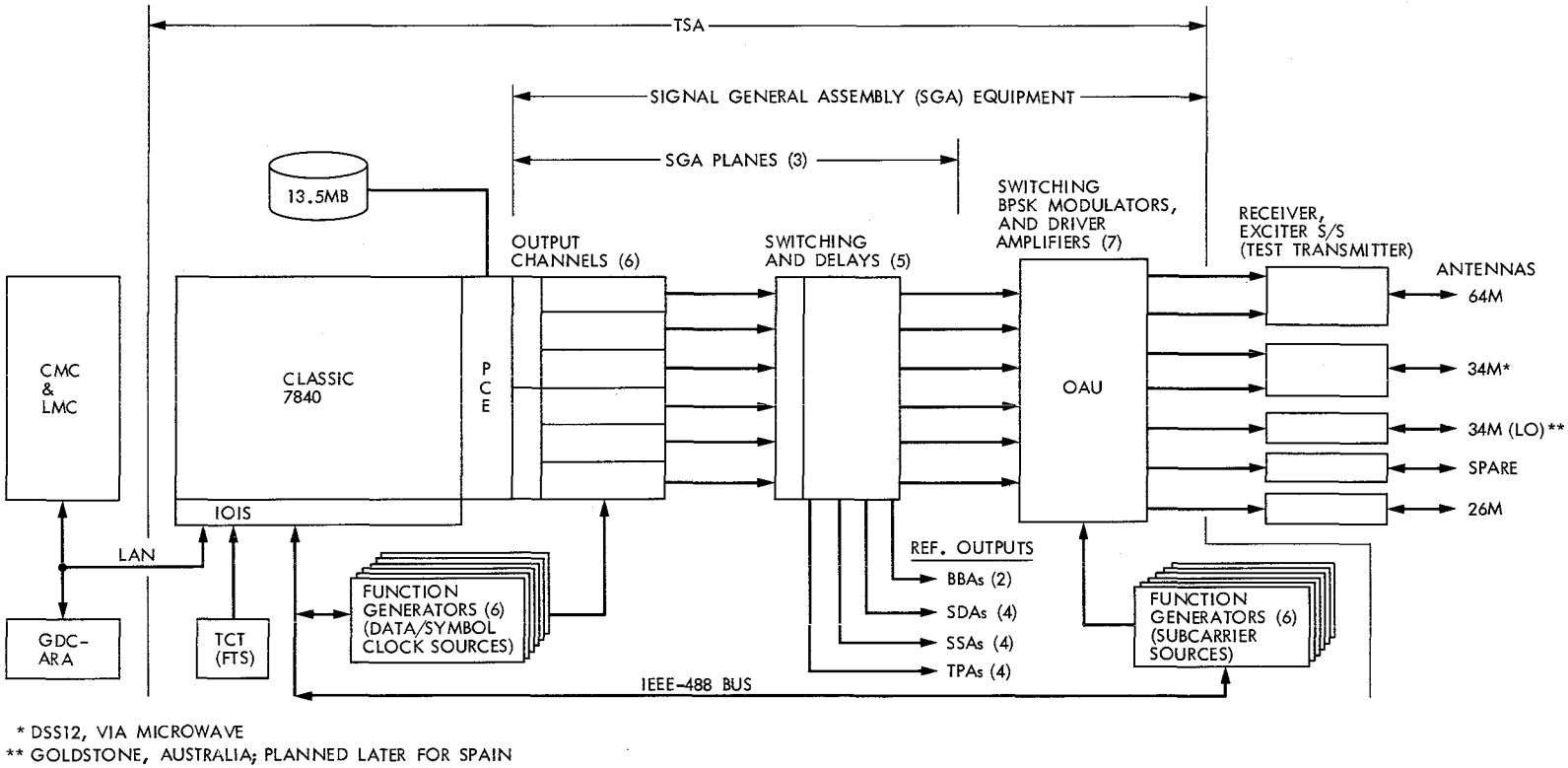


Fig. 1. Test signal distribution



* DSS12, VIA MICROWAVE
** GOLDSTONE, AUSTRALIA; PLANNED LATER FOR SPAIN

Fig. 2. TSA subsystem block diagram

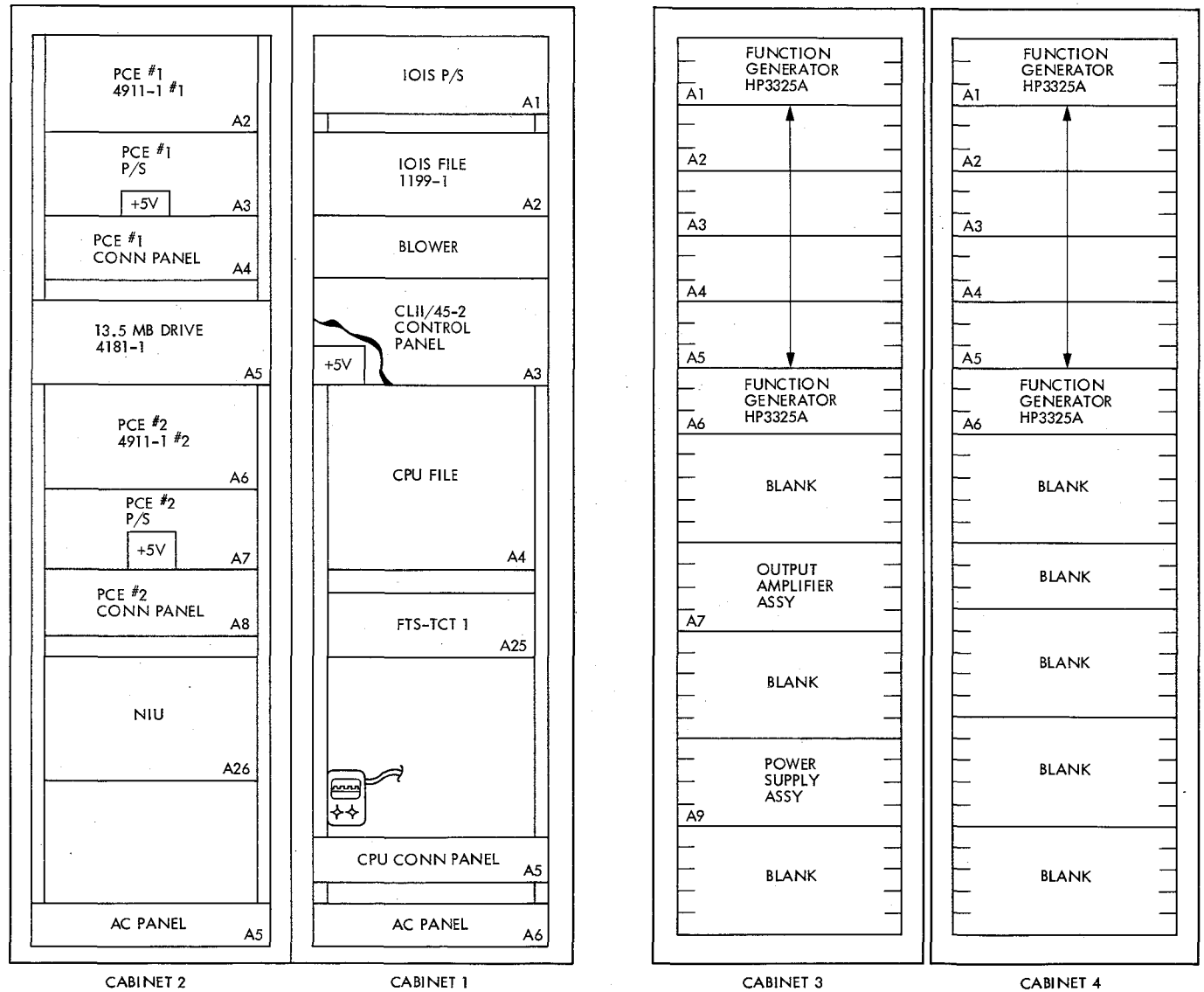


Fig. 3. TSA rack configuration

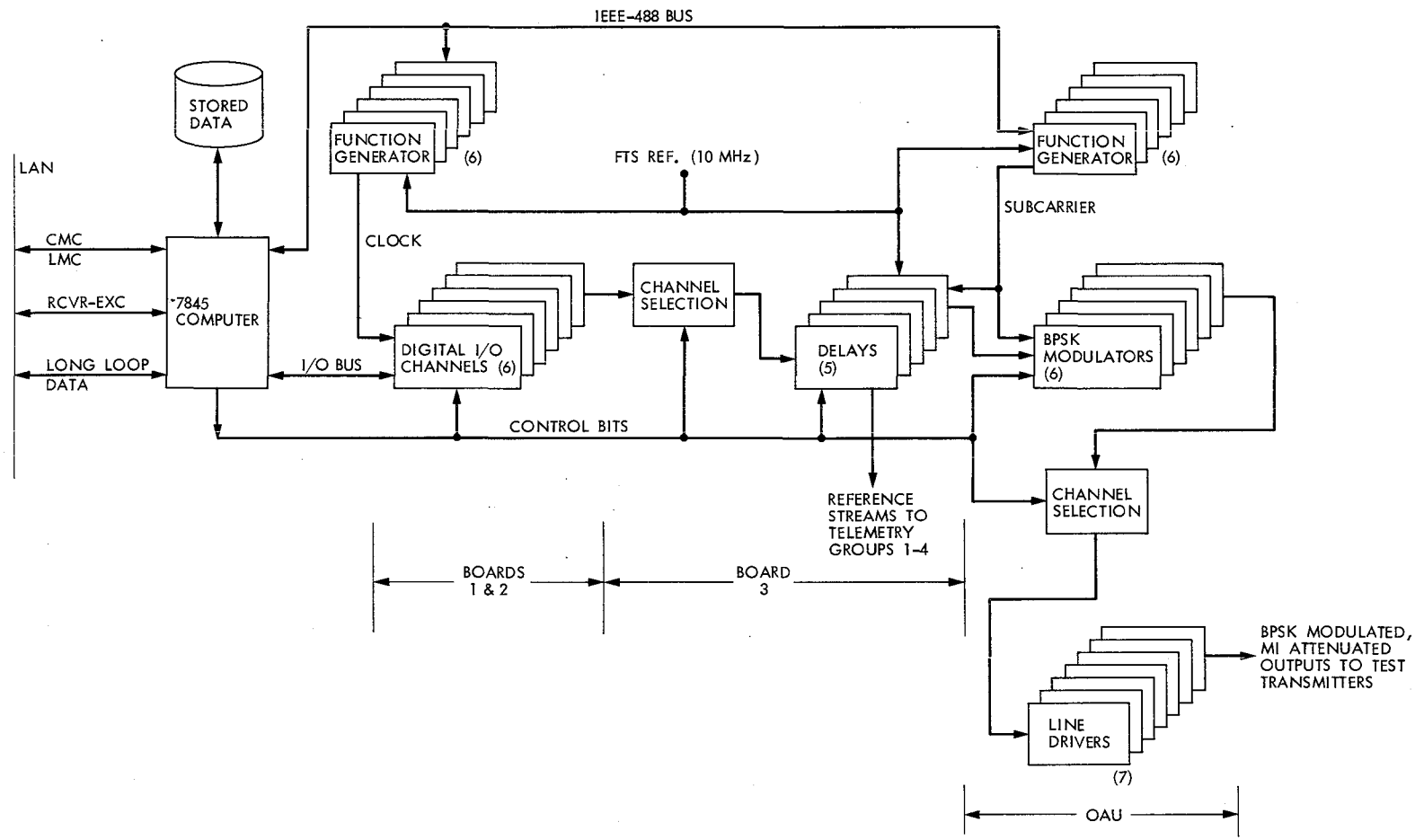


Fig. 4. TSA hardware configuration

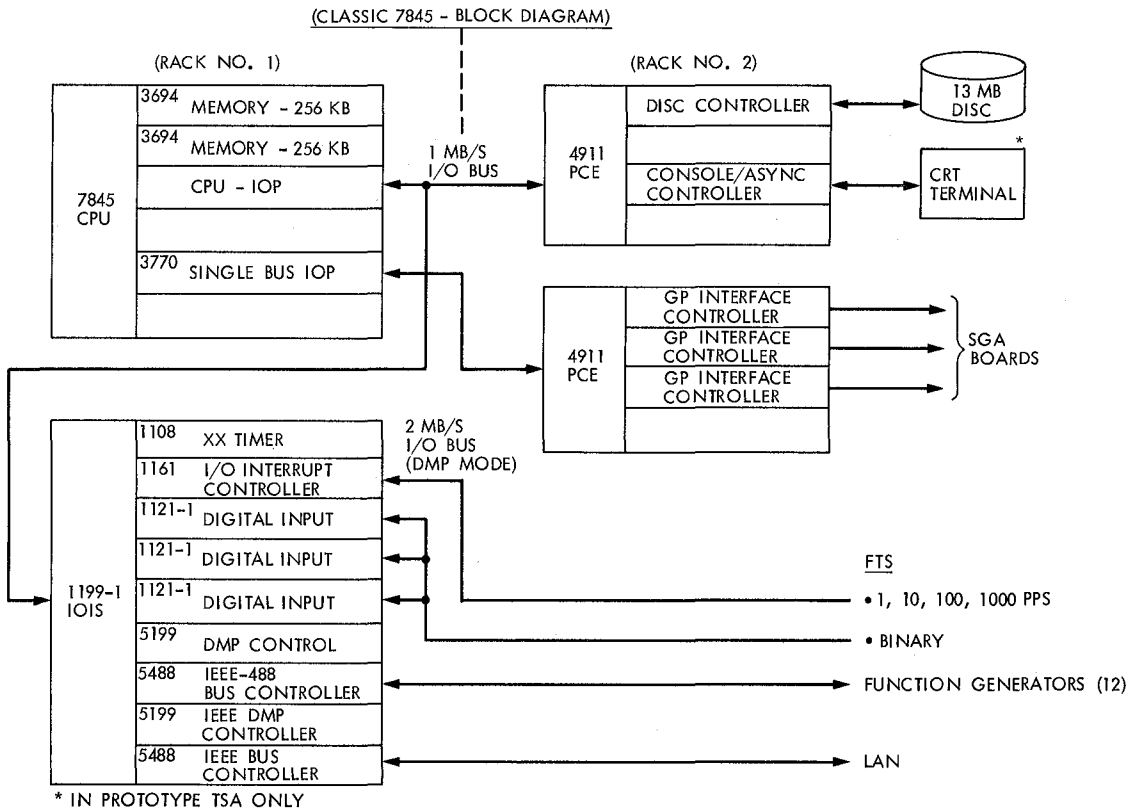


Fig. 5. TSA computer configuration

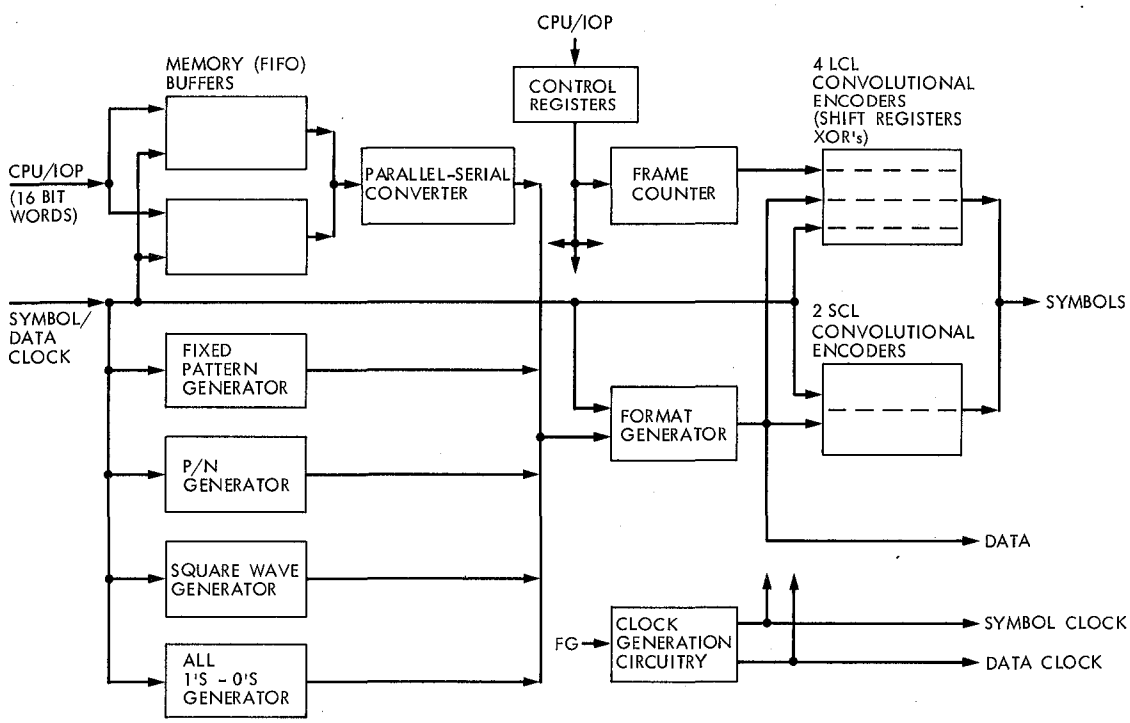


Fig. 6. Digital channel configuration

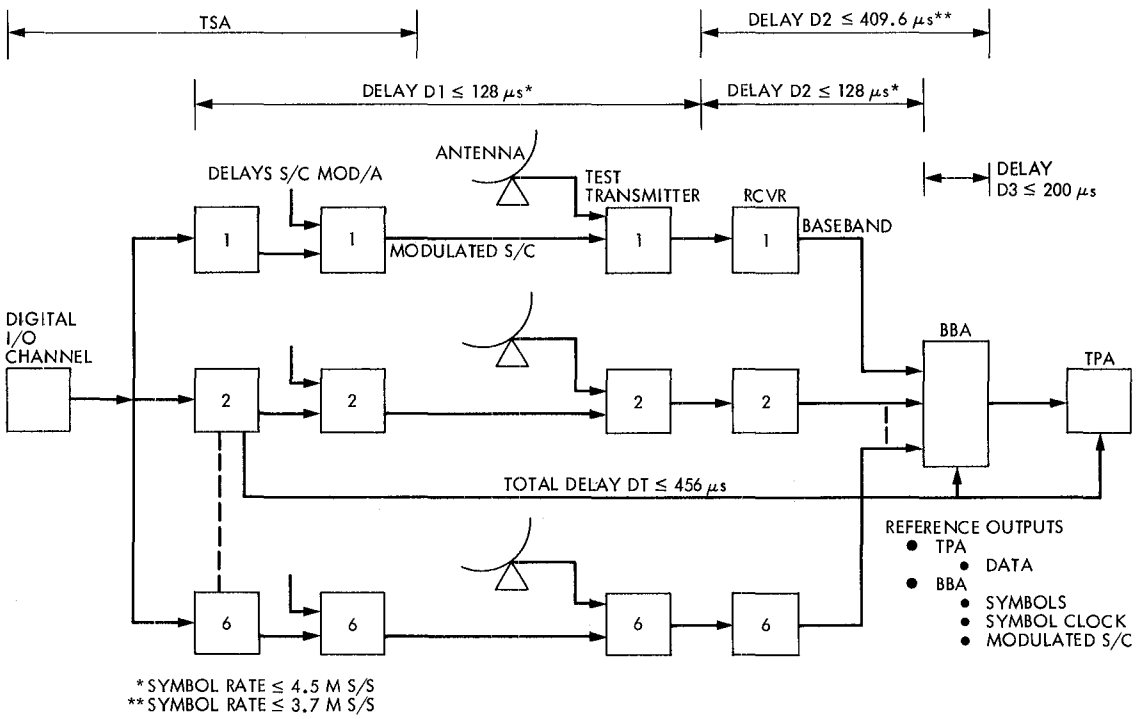


Fig. 7. Antenna array block diagram

Operator Workload Measurement Validation for the Mark IV-A DSCC Monitor and Control Subsystem

M. Le May
Montclair State College, New Jersey

E. E. Hird
Operations Sustaining Engineering Section

B. Y. Rodriquez
Goldstone Operations Section

The validation of some proposed measures of operator workload in a large interactive computer system is described. The tests were conducted on Mark III Deep Space Network (DSN) equipment using experienced operators as subjects. In addition to identifying the significant operator workload measures, some useful guidelines have also been obtained on operator task scheduling.

I. Introduction

The updating and increasing automation of large, interactive computer systems is a common process where such systems are in use. The DSN is currently undertaking such an update by introducing a new Mark IV-A monitor and control subsystem at its three station complexes located approximately 120 degrees longitude apart around the earth. The purpose of the Mark IV-A system is to reduce operating costs by centralizing operations at each station, and increasing the amount of automation (Ref. 1). This will result in new combinations of operators, and new ways of operating the system, with as yet unknown workload demands.

Therefore, it was considered necessary to develop measures of operator workload that would underscore those subtasks on which the load is likely to exceed scheduled times and increase operator stress, especially if anomalies occur during the task performance. The extensive literature on human workload measurement (Refs. 2, 3, and 4) yielded a number of possible techniques. An evaluation of their effectiveness in various circumstances, also based on the literature, led to a choice of three measures that might prove useful. These are:

(1) Sequential operator ratings of workload on subtasks.

(2) The ratio of the time required to the time available to do a particular task.

(3) Initialization time, which is defined as the time taken to initiate the physical (keystroke) portion of a subtask.

These measures are described in detail in Ref. 1. The salient feature of the particular technique used here is that the operator ratings are taken at the end of each small subtask of a whole task (e.g., at the end of the load of a particular program such as the Metric Data Assembly (MDA), in a total "load and go" task) so that ratings and times are obtained for each subtask. In this way, subtasks that are likely to exercise an operator's capacity near its limit are rated and timed in the context of the whole task. This is especially useful during the design and operational testing phase of a new large interactive computer system such as the Mark IV-A, since it may be possible to make design changes or devise workarounds to alleviate difficulties with particular subtasks during this phase.

Although the three proposed measures were culled from an extensive literature on the subject, and should be usable on this basis, their use in large interactive computer systems has not been tested. Moreover, the technique of inserting ratings at intervals in the task and using ratings to time subtasks is also unique in this context. Therefore, the study described here was undertaken to assess the validity and reliability of the proposed workload measures in the operational environment of an existing Mark III Deep Space Station (DSS). The experiment used a "5 X 3 within subjects" design, in which each subject completed a whole computer "load and go" task consisting of five subtasks under each of three workload conditions: low, medium, and high workload. (The low workload condition involved only four subtasks.) In this way, the sensitivity of the measures to an objective increase in workload, and to differences among subtasks, could be determined since, if the measures were valid, they could be expected to rise with actual workload increases and with more difficult tasks.

In addition to the rating of each subtask, operators were asked to rate a number of aspects of the working environment on a separate questionnaire.

II. Method

A. Subjects

Nine operators of the ground tracking and data acquisition equipment at DSS 12 at Goldstone, California, served as subjects. All had some experience and were thoroughly familiar with the job.

B. Apparatus

The operating elements of the DSN include the Network Operations Control Center (NOCC) at JPL and three Deep Space Communications Complexes (DSCCs) located at Canberra, Australia; Goldstone, California; and Madrid, Spain. The NOCC controls, monitors, and coordinates the DSCCs' operational activities in response to the requirements placed upon the DSN by the various Flight Projects. The DSCCs provide ground-based communications links with Flight Projects' spacecrafts in interplanetary flight or in highly eccentric earth orbit. In addition, they support some radio astronomy tasks and very long baseline interferometry measurements. Each of the DSCCs currently operates a 64-m and a 34-m antenna. The 64-m and 34-m antennas at Goldstone are operated independently of each other at separate locations, while similar antennas at Canberra and Madrid are colocated and conjointly operated. A number of computerized subsystems, some of which were used in collecting data for the study reported here, are used in a partially automated manner in conjunction with some other manually operated subsystems and a selected antenna to acquire data from the various spacecraft in accordance with a predetermined schedule.

Time was scheduled in four-hour blocks at DSS 12 at Goldstone to conduct the prepared operator workload trials. One operator could always be tested in a time block, and it was sometimes possible to test two. A simulated "load and go" task that required one operator was designed for the study. The task involved initializing seven different computers so that they were ready to track a designated spacecraft. A "load and go" may be characterized as the reinitialization of each subsystem computer for the next spacecraft to be tracked during the period of time needed to repoint the antenna. It was assumed that subsystem elements that need calibration will not change significantly during the upcoming support activity.

C. Procedure

There were three versions of the task, designed to impose a low, medium, or high workload on the operator. The low workload condition was a simple "load and go" task with no complications; its scheduled time for completion was 12 minutes. In the medium workload condition, a ranging operation requiring additional equipment was added. The scheduled time for this was 8 minutes in addition to the 12 minutes required for the original task. In the high workload condition the same task was used (with ranging included), and two anomalies on two different computers were added by the investigators. The scheduled times were set by the experimenters, and provided just enough time to complete each task under normal working conditions.

Before any data was collected, the purpose and methods of the study were explained to all the potential subjects in a written form, and in a brief discussion session. Then, at the start of each session, subjects read a standard set of instructions which directed them to rate, on a scale from one to seven, the amount of mental effort (how hard they had to think) and the amount of time stress (how rushed they felt) involved in the subtask they had just completed. Each of the ratings was given a verbal label. These labels were printed on a

card that was mounted on the console for easy visibility during the procedure. A reproduction of the card is illustrated in Fig. 1. They were also informed of the scheduled time of 12 minutes, and of the 8-minute time "pad" allowed for the medium and high workload conditions.

Following the instructions, subjects were given a short practice session in which to familiarize themselves with the procedure, so that the workload measures themselves would not intrude on the timing of the subtasks.

Then, when they were sure they understood the procedure thoroughly, the high, medium, and low workload conditions were arranged in a random order and the trials were begun. Each operator input and computer response was obtained on a printout, along with its time, so that operator ratings and times could be read for each subtask. In a continuous task such as the one studied here, the end of one subtask signals the start of the next; thus, the operator's ratings could be used to obtain the total time for that subtask and the time for the operator to start the next one. This latter was the time between the keystroke that gave the rating and the next operator input. Total subtask time was the time between one rating and the next.

The low workload condition was chosen so that operators could accomplish the task easily, with no time pressure; it was possible to accomplish the medium workload task in the allotted time, given the 8-minute "pad", and it was possible to complete the task in the time allowed under the high workload condition if the operator understood the system thoroughly and worked fast.

After the completion of each condition, the operator was asked to complete a 14-item workload questionnaire on which a number of aspects of the working environment were also rated on a scale from one to seven. This is reproduced in Fig. 2.

Following the completion of all of the conditions, a debriefing session was held in which the operators were asked to comment on the study and explain the ratings they had just given.

III. Results

Three measures of operator workload were obtained from the printouts. The first was, of course, the operator's rating of subtask difficulty and time stress. This was obtained directly from the printout. The notation on the printout of the times when these were given made it possible to calculate the amount of time the operator actually required to perform the subtask. This was then used to obtain the second measure, which was the ratio of the time required to the time allowed.

$$W = \frac{T_r}{T_a}$$

where

W = workload ratio

T_r = actual time required by the operator

T_a = time allowed, or scheduled time

The scheduled time is set by NOCC for each whole task, e.g., one-half hour is allowed for a complete "load and go" operation. Actual times to perform the task were obtained from printouts for four recent operational runs on which no anomalies were encountered. The task was found to occupy an average of 38 percent of the scheduled time under these conditions. Average times were then obtained for each subtask under operational conditions, and these were divided by 0.38 to obtain the scheduled time, T_a for that subtask. Thus, the scheduled time for each subtask was assigned on the basis of its average proportion of the scheduled time for the whole task. It was necessary to do this because scheduled or allowed times are given only for whole tasks, and detailed data on subtasks was desired.

The third measure, initialization time, was the time between the rating of one task and the first input on the next.

Each of the measures was subjected to a "3 (workload conditions) by 5 (subtasks)" analysis of variance to determine whether differences observed among the conditions were significant. The two ratings (mental effort and time stress) were analyzed separately. The time stress ratings and the workload ratio varied significantly with both subtask and condition; the mental effort rating varied only with subtask, and the initialization time was not related to workload variations. No significant interactions were observed, i.e., the effect of increasing workload was the same over all subtasks.

The means of each of the measures for all subtask and workload conditions are shown in Tables 1 through 4, where each of the processors used in program loading are indicated. The mean ratings for "mental effort" rise with increasing workload, and vary with the subtask, but these differences are not significant. The means of the ratings for each load condition over all subtasks for the "time stress" measurement, however, do rise significantly with the objective increase in workload, and also change with the subtask. This indicates that different subtasks impose differing amounts of time stress, and that the measure is sensitive to objective workload increases. Figure 3 illustrates the overall rise in the ratings for low, medium, and high workload conditions. The data has

been normalized for easier comparison with the workload ratio data presented in Fig. 4.

The means of the workload ratio (W) for each condition are presented in Table 3, and the significant rise in workload ratio with increasing objective workload is illustrated in Fig. 3. The workload ratio is somewhat more sensitive to workload conditions than is the "time stress" rating, although both measures rise and fall similarly with the same subtasks. This is consistent with the lack of significant interaction terms. The Spearman rho correlation between "time stress" ratings and workload ratio is 0.78. It appears that both measures are measuring the same subtask characterisics, and both are sensitive to workload increases.

Variations in initialization time did not prove significant, either for subtask or for imposed workload. These are presented in Table 4.

The number of errors (see Table 5) made by operators during task performance was relatively small (43 in almost 8 hours of operation) so no statistical analysis could be done, but they increased from 9 under the low workload condition to 17 under the medium and high workload conditions. The number of operator inputs relative to the minimum number of required inputs also increased from 1.09 at low workload to 1.15 at medium workload and 1.34 at high workload.

The mean ratings for each of the statements on the Operator Workload Questionnaire are presented in Fig. 1. None of the ratings seem to indicate that operators have particular problems with the working environment of the system as it now stands. During the exit interview, the operators agreed that the task chosen was representative of typical tasks at the DSS. They also felt that the taking of the workload measures was somewhat intrusive on task performance. No other problems in operating the system were mentioned.

IV. Discussion

The significant findings for the time stress ratings and the workload ratio mean that these two measures do, in fact, vary with objective workload and can therefore be used together to measure operator workload at the DSN monitor and control consoles. Moreover, the fact that the subsystem has a significant effect on both measures means that subtasks vary in the amount of workload. Tasks with higher measured workloads are those on which errors leading to data loss might be expected to occur. This knowledge itself should help operators in becoming oriented to a new system (operators probably are aware of "trouble spots" in an older system without any measures), and may provide information useful for hardware and software improvements in the system during operational testing.

The "mental effort" ratings did rise with increasing workload, but differences were too small to be significant. This is not too surprising, since it is not usually necessary to exert much mental effort in the performance of a very familiar operation. This measure probably would be sensitive to workload during the operation of a new system. At any rate, since the sensitivities of time stress ratings and workload ratio were demonstrated with the mental effort measure in use, it should be left in place in future use, in order not to change the cognitive aspects of the situation, i.e., its removal might have an effect on the other two measures, which were significant.

The initialization time measure was an attempt to measure the time required for mental processing of the information needed to perform a control function. The reasons for its failure to respond to objective workload increases are beyond the scope of this paper. However, in view of the recent finding of Wierwille and Connor (1983) that only 3 out of 20 workload measures were actually sensitive to workload increases on a particular task, the lack of sensitivity of 1 such measure is not too surprising.

The high correlation between the rating and the workload ratio measures indicates that they are essentially measuring the same thing, and they therefore act as a check on each other's reliability. They are not redundant, however. The rating is extremely easy to use and interpret, and can be obtained immediately for use in possible "on the spot" diagnosis of troublesome tasks. Moreover, it is necessary in order to obtain the times used in calculation of the workload ratio.

The workload ratio is a particularly useful measure, since it provides information on the time to perform each subtask as well as the whole task, and on the proportion of task time needed for each subtask. It is the measure most sensitive to workload variations. Moreover, although the mean ratios appear rather low for all the conditions, there is considerable variability. In particular, the presence of anomalies, even simple ones such as those introduced by the experimenters, can raise the ratio near or over 1.00. This means that the scheduled time has been exceeded on that task. This happened seven times, or five percent of the total number of subtasks during the medium and high workload conditions, and the ratio was over 0.80 on another seven occasions. An examination of the raw data showed that these were the occasions on which errors were likely to be made. Moreover, nine "volunteer" equipment failures (not induced by the experimenters) occurred during the testing, and these also seemed related to increased workload, since seven occurred during the high workload condition. Thus, both equipment and operator failures seem to be related to operator workload. In addition, the simple fact of exceeding the scheduled time can mean lost data. Thus, it is important for system operation that the

operator workload be kept at a moderate level and the first step towards achieving this goal is the use of the techniques developed here for operator workload measurement.

V. Conclusions

1. Two workload measures that are sensitive to increasing objective workload were developed for use in operating a large, interactive computer system (the DSN monitor and control subsystem). They are:

   a. Operator rating, on a scale from one to seven, of the "time stress" encountered in a particular subtask.

   b. The ratio of the time required to do a particular subtask to the time allowed for it.

2. The average workload ratio in the (typical) task studied was approximately 0.40; i.e., the tasks occupied 40% of the scheduled time. This ratio led to exceeding the scheduled time on five percent of the subtasks.

3. Time stress ratings vary from one individual to another. However, subtasks that have a mean rating of three or above from at least three operators are considered problematic.

4. To ensure that the 0.40 workload ratio is not exceeded, workload ratings and ratios can be taken whenever new equipment or a new program or procedure is introduced, so that reasonable scheduled times can be established. This work would be done after operators are somewhat familiar with the system. Also it may need to be done more than once, as time (T_r) will become shorter as operators become more familiar with the system.

References

1. Le May, M., Chafin, R., and Hird, E. E., "The Measurement of Operator Workload in the Mark IV-A DSCC Monitor and Control Subsystem." *The Telecommuncations and Data Acquisition Progress Report, 42-72*, October-December 1982, pp. 177-185, Jet Propulsion Laboratory, Pasadena, California, February 15, 1983.

2. Wierwille, W. W., Williges, R. C., and Schiflett, S. G., "Aircrew Workload Assessment Techniques," in Hartman, B. O. and McKenzie, R. E., *Survey of Methods to Assess Workload*. AGARDograph No. 246, August, 1979.

3. Moray, N., *Mental Workload: Its Theory and Measurement*. Plenum Press, New York, 1979.

4. Wierwille, W. W. and Connor, S. A., Evaluation of 20 Workload Measures Using a Psychomotor Task in a Moving-Base Aircraft Simulator. *Human Factors*, Vol. 25, No.1, pp. 1-16, 1983.

Table 1. Mean operator ratings of "mental effort" required for each subtask under low, medium, and high workload conditions

Load Condition	Subtask				
	CPA[a]	MDA[b]	TPA[c]	DCO[d]	PRA[e]
Low	1.67	1.55	1.78	1.67	
Medium	2.11	1.77	1.77	1.77	2.77
High	2.22	1.77	1.77	1.66	3.00

[a]Command Processor Assembly (two processors were initialized).
[b]Metric Data Assembly.
[c]Telemetry Processor Assembly (two processors were initialized).
[d]Digitally controlled oscillator.
[e]Planetary Ranging Assembly.

Table 2. Mean operator ratings of "time stress" imposed by each subtask under low, medium, and high workload conditions

Load Condition	Subtask				
	CPA	MDA	TPA	DCO	PRA
Low	2.00	1.67	1.78	1.55	
Medium	2.22	1.55	1.66	1.77	2.88
High	2.22	1.77	1.66	1.88	3.77

Table 3. Mean workload ratio measure ($W = T_r/T_a$) for each subtask under low, medium, and high workload conditions

Load Condition	Subtask				
	CPA	MDA	TPA	DCO	PRA
Low	0.56	0.29	0.32	0.39	
Medium	0.59	0.19	0.24	0.56	0.41
High	0.65	0.27	0.31	0.59	0.60

Table 4. Mean initialization time in seconds for each subtask under low, medium, and high workload conditions

Load Condition	Subtask				
	CPA	MDA	TPA	DCO	PRA
Low	14.4	6.5	14.9	11.3	
Medium	17.4	6.5	16.4	7.5	14.3
High	17.4	11.8	14.9	13.3	15.7

Table 5. Number of errors and ratio of actual inputs (I_a) by operators to minimum number of required inputs (T_r) under each workload condition

Parameters	Load condition		
	Low	Medium	High
Errors	9	17	17
I_a/T_r	1.09	1.15	1.34

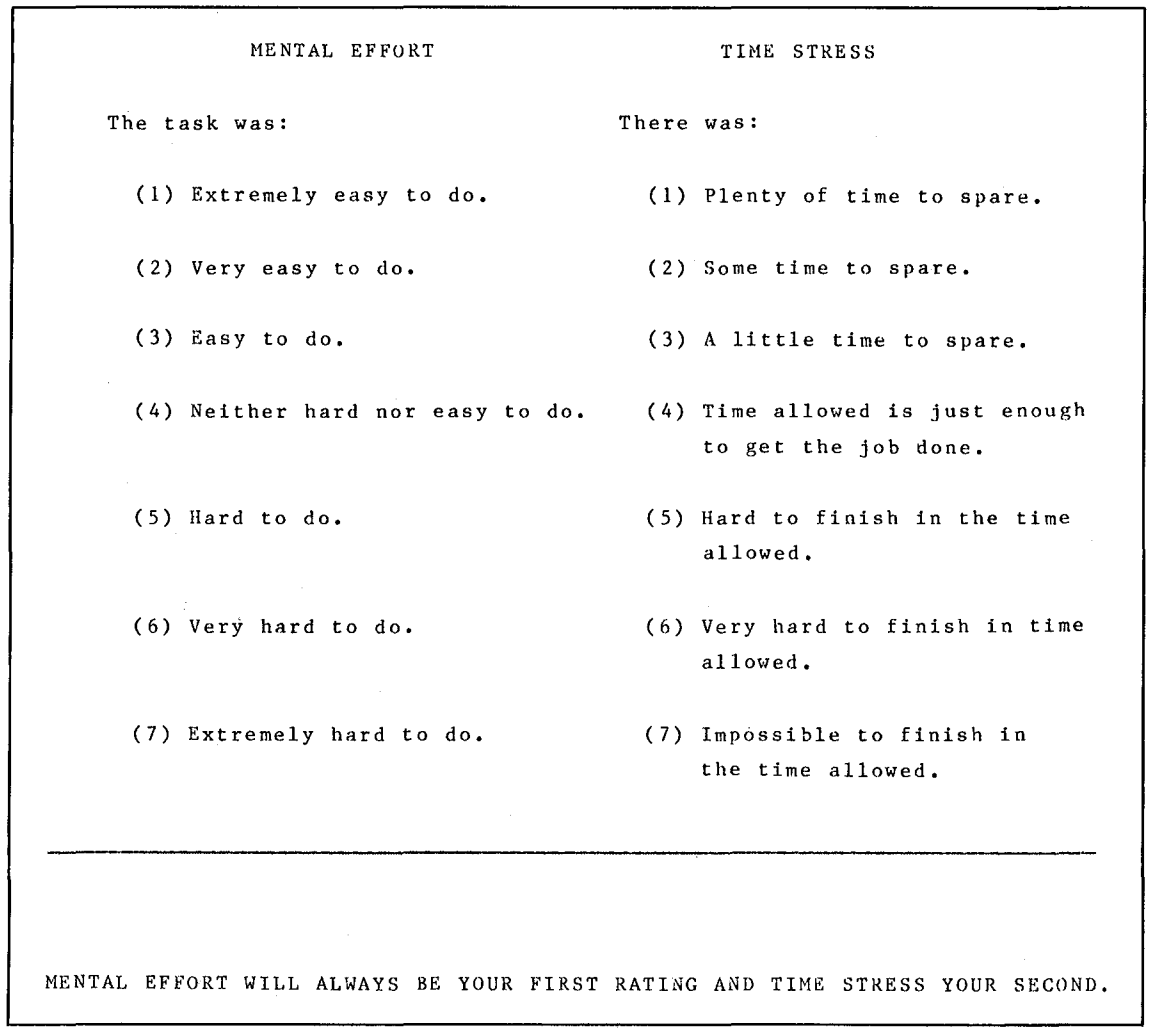


Fig. 1. Verbal labels for rating scales given to operators during the testing procedure

OPERATOR WORKLOAD QUESTIONNAIRE

Please indicate, in the space provided, your agreement or disagreement with the following statements on a scale from 1 to 7 where:

1 = very strongly disagree
2 = strongly disagree
3 = disagree
4 = neither agree nor disagree
5 = agree
6 = strongly agree
7 = very strongly agree

		Mean Operator Rating
1.	The workload required of the operators of this system is too high.	[ ] 2.78
2.	The tasks in this run were excessively complex.	[ ] 3.00
3.	I was familiar with all the commands I needed to use on this task.	[ ] 5.88
4.	It was easy to find the information I needed in the system documentation.	[ ] 5.00
5.	I had confidence in the actual equipment in the string that was made available for this task.	[ ] 4.77
6.	The lighting arrangements were appropriate for the tasks.	[ ] 4.77
7.	The heating and/or air conditioning were adequate.	[ ] 4.77
8.	The physical arrangement of the equipment was appropriate.	[ ] 4.0
9.	The system provides all the displays that are needed for the task.	[ ] 4.77
10.	Each display provides adequate information for its task.	[ ] 4.89
11.	Prompts make clear what action is to be taken.	[ ] 4.44
12.	Subsystem error messages describe each fault in sufficient explanatory detail.	[ ] 3.67
13.	Subsystem error messages make clear what action is to be taken.	[ ] 3.89
14.	There were other problems in operating the system (specify).	[ ]

Fig. 2. Operator workload questionnaire sample with mean ratings of nine operators for each statement

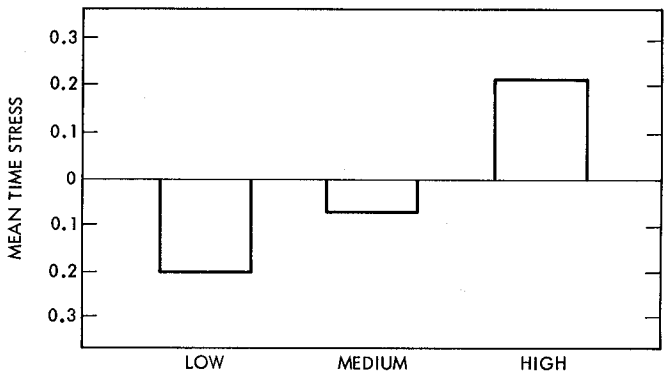


Fig. 3. Mean normalized operator "time stress" rating under low, medium, and high workload conditions

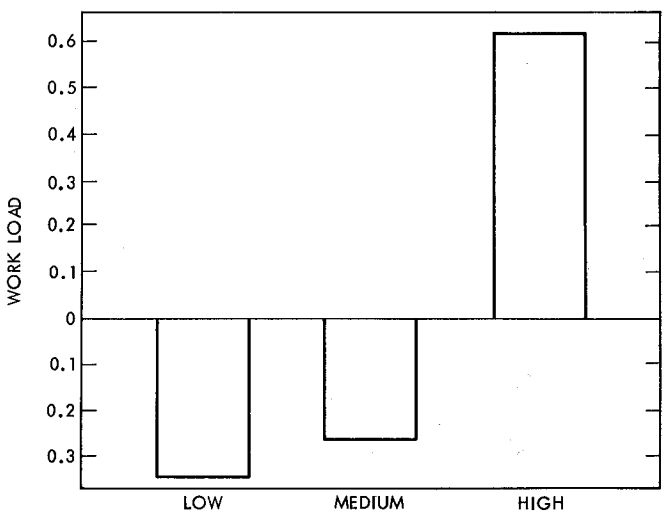


Fig. 4. Mean normalized workload ratio ($W = T_r/T_a$) under low, medium, and high workload conditions

First Results of GPS Time Transfer to Australia

J. McK. Luck and J. R. Woodger
Division of National Mapping
Canberra, ACT, Australia

J. E. Wells and P. N. Churchill
DSS42/43
Tidbinbilla, ACT, Australia

P. A. Clements
Communications Systems Research Section

A GPS time transfer unit built by NBS under contract to JPL was installed at Tidbinbilla Deep Space Communications Complex of the DSN in June 1983. It has been used to estimate the relationship to UTC(USNO MC) of the Tidbinbilla frequency and time system TID(FTS) based on a hydrogen maser, and thence to estimate the performance of the Australian free-running time scale UTC(AUS). Data from the first three months has been analyzed three ways: by two-hop "common view" using JPL as intermediary; by "long-arc" interpolation of measurements against space vehicle clocks; and by "long-arc" interpolation of GPS-Time results. Residuals from a single quadratic fit through three months of UTC(USNO MC) – TID(FTS) results were white noise with standard error 15 ns, and a flying clock measurement gave 70 ns agreement. A straight line fit through results UTC(USNO MC) – UTC(AUS) gave 90 ns standard error and 120 ns agreement. It is proposed to use the GPS measurements to steer UTC(AUS) to UTC(BIH), and to rename the existing time scale TA(AUS).

I. Introduction

A Global Positioning System (GPS) Time Transfer Unit (TTU) built by U.S. National Bureau of Standards (NBS) for the DSN was installed at Deep Space Communications Complex (DSCC) 42/43 at Tidbinbilla (TID) in late June 1983. It was turned on in July and has operated correctly from that moment. Its principal function of interest here is to monitor the performance of the Tidbinbilla Frequency and Time System (TID(FTS)), which was derived from a SAO Model VLG11 hydrogen maser P14 during the period to November 1983. The basic results recorded are the time intervals between TID(FTS) 1 pps and the timing marks broadcast from the clocks on board each GPS space vehicle (SV), corrected in real time for propagation delay calculated from the on-board ephemeris transmitted by each SV and an assumed position for the antenna:

λ = 148°58′ 48″2018 E

ϕ = 35°24′ 8″0444 S

h = 665.54 m

Also available is the time comparison between TID(FTS) and the GPS master clock (GPS TIME) in the GPS Control Segment (Ref. 1). Generally two passes from each of space vehicles 5, 6, 8, and 9 are observed each day, of which only the higher altitude passes are used in this analysis.

Similar results are obtained on each SV at the US Naval Observatory (USNO), referred to their Master Clock (UTC(USNO MC)), although at different times in the day, and made available through the bulletin Series 4: Daily Phase Values, and otherwise. Time transfer between UTC (USNO MC) and TID(FTS) is obtained by linear interpolation to the time of a Tidbinbilla observation between two adjacent reported USNO observations, usually but not always one sidereal day apart. In this paper, data gaps up to two days are tolerated, and only linear interpolations are employed; hence it is tacitly assumed that on-board clock behaviour is linear over a day or two and that receiving antennae locations and broadcast orbital parameters are accurate. Because of the appreciable interpolating factors, this method is here designated "long arc" and is applied to results obtained both from SV observations and GPS TIME results.

The "common view" method in which both stations take measurements simultaneously (Refs. 2 and 3) is impossible between Tidbinbilla and USNO and difficult between Tidbinbilla and Jet Propulsion Laboratory (JPL). Nevertheless a two-hop quasi-"common-view" experiment has been attempted, with a GPS TTU at JPL taking nearly simultaneous observations with Tidbinbilla on the one hand, and nearly simultaneous observations with USNO on the other. The effectiveness of the time transfer then depends on the behaviour of the JPL Frequency and Time System between the two sets of observations which may be several hours apart.

Since four satellites were available, and more can be expected, the opportunity exists for averaging. This has been done by interpolating linearly to 0^h UTC between adjacent time transfer results from each satellite prior to taking the mean. This process yields a "consolidated" result.

II. The Tidbinbilla GPS System

A. Equipment

The GPS equipment installation at the Canberra Deep Space Communication Complex (CDSCC), Tidbinbilla, is depicted in Fig. 1. All of the equipment with the exception of the antenna unit is mounted in a short rack adjacent to the CDSCC frequency and timing system (FTS) monitor panel in the operations building (Fig. 2). The antenna unit, which also includes a low noise amplifier and mixer, is a sealed enclosure located on the roof of the operations building (Fig. 3).

The antenna position was obtained by carrying out a traverse from the surveyed ground monument position beneath the DSS 42 34-metre antenna. A receiver offset of 272 nanoseconds was inserted in the GPS software to correct for the delay due to cable length between antenna unit and receiver.

A 5 MHz signal is provided from the FTS as a reference frequency from which is derived the receiver 100 MHz to supply the first I.F. mixer in the antenna unit. D.C. power is also run through the 100 MHz coax cable, avoiding the requirement to provide a separate power cable. A 1 PPS signal is also connected from the FTS to provide the CDSCC clock reference to the system.

A modem phone is connected to an RS 232 port of the microprocessor to allow data acquired and stored by the system to be transmitted periodically by telephone to JPL, usually once a week. Data from a remote receiver may also be printed out locally using this modem link.

B. Operation of GPS Receiver

The GPS receiving system is capable of running in an automatic mode, once all relevant parameters have been entered. In this mode, the system will acquire chosen space vehicles, lock onto the downlink, and accumulate relevant data. The system is normally operated in the automatic mode at DSS 43.

The software/operator interface consists of a "user friendly," menu-driven parameter selection matrix, arranged in the form of one main menu and a number of submenus.

In order to bring the system up to a functioning condition, it is necessary to perform a "cold start" operation. This consists of instructing the system to acquire an almanac from a satellite when such a vehicle is within reception range (usually above horizon). It is also necessary to set the system's internal UTC clock by responding to the system prompts. Under normal operating conditions, a battery back-up supply maintains the system's RAM such that a cold start is not necessary even in the event of a power failure.

The system software provides an aid to the selection of suitable satellites for tracking purposes. If selected, the system will print a graph of elevation and azimuth versus time for up to five vehicles, dependent, of course, on the system's knowledge of its present location. This parameter may be entered by calling up the NEW RECEIVER COORDINATES feature on the display menu.

Once suitable satellites to track have been selected, appropriate track times may be entered, derived from the aforementioned graph. The facility exists to cause the system to decrement the track start times by 4 minutes per day to account for the fact that the satellites are in sidereal orbits. The satellite may then be tracked in the same position in the sky each day.

The system may be commanded to perform position location calculations by one of two methods. The first method requires that four satellites be in view so that data of sufficient precision may be obtained for the navigation solution. The second method requires that the system lock sequentially onto four satellites every two minutes. This second method produces a solution in a shorter time than the former and has the advantage of eliminating much of the short term noise in the local clock system.

The GPS measurement computation is performed by calculating a pseudorange value from system counter measurements and then computing the slant range, relativistic and ionospheric corrections based on data obtained from the satellite ephemeris. These corrections, along with the local receiver delay, are subtracted from the pseudorange estimate to obtain the local clock minus satellite clock value. The satellite clock correction, transmitted from the vehicle is then added to this value to produce a figure for local clock minus GPS clock.

III. Space Vehicle and GPS Times

The results produced by Tidbinbilla's GPS TTU are, for each SV(i), numbers in the form TID(FTS) – SV(i) and TID(FTS) – GPS TIME via SV(i). The corresponding numbers disseminated by USNO are UTC(USNO MC) – SV(i) and UTC (USNO MC) – GPS TIME via SV(i). It has been found that the "raw" results TID-SV and USNO-SV are adequately modelled by quadratic curves over the interval MJD 5528-5618 for SV(5) and SV(9), while cubic or higher-order fits would be required for SV(6) and SV(8). The parameters of the quadratics are given in Table 1 in the form:

$$\text{TID(FTS)} - \text{SV}(i) = a(i) + b(i) \times (t - \bar{t}) + c(i) \times (t - \bar{t})^2 \qquad (1)$$

and similarly for UTC(USNO MC) - SV(i), where $a(i)$ is the offset at the mean time $\bar{t}$, $b(i)$ is the rate at $\bar{t}$ and $c(i)$ is half the drift rate. The standard errors of residuals $\sigma(i)$ are also tabled, and the residuals are displayed in Figs. 4 and 5. Both graphs have many features in common, demonstrating that vagaries in on-board clock behaviour are readily detectable.

Several rate changes were observed to occur in GPS time in the interval considered, so the results were broken into four segments and straight lines fitted through each as shown in Tables 2 and 3. The residuals are displayed all together in Figs. 6 and 7, where it can be seen that results are rather better at USNO than at Tidbinbilla which is possibly a consequence of efforts made at the GPS master station to follow USNO time.

IV. Long-Arc Time Transfer to USNO

Time transfer from USNO to Tidbinbilla was achieved by linearly interpolating between successive daily results UTC (USNO MC) – SV(i), and subtracting the observed value of TID(FTS) – SV(i) from it. It was felt that it was better to use the USNO results to interpolate on since its time scale is the reference being accessed and is therefore to be considered definitive for these purposes. Linear interpolation was adequate and in fact desirable since the effects of drift rate over one or two days are swamped in the random noise. The "consolidated" result is shown in Fig. 8, on which is also shown the value

$$\text{UTC(USNO MC)} - \text{TID(FTS)} = -11.030\ \mu\text{s}$$

by USNO/Bendix flying clock on 1 October 1983.

Quadratic curves in the same form as Eq. (1) were fitted to the outcomes using each space vehicle separately. Their parameters are given in Table 4, and the residuals therefrom are shown in Fig. 9. Residuals from the "consolidated" time transfer are shown in Fig. 10, and the Allan variances of these residuals are displayed in Fig. 11 in which the slope is close to -1, indicating that the residuals after removal of the quadratic are very nearly random uncorrelated "white" phase noise.

Almost identical results are obtained when transfer is effected via GPS TIME instead of SV TIME. Table 4 contains their quadratic parameters, Fig. 12 shows their consolidated Allan variances, Fig. 13 the individual residuals and Fig. 14 the consolidated residuals.

From this analysis, the drift rate of the Tidbinbilla hydrogen maser with respect to UTC(USNO MC) is +4 parts in 10^{15} per day, and is undoubtedly now well measured: The drift rate itself has not changed during three months.

V. Common View Measurements

The results given above show quite clearly that the GPS receivers can detect anomalies in the on-board clocks and in GPS TIME as small as 10 ns or less, so simultaneous observa-

tions should remove their effects. The geographical locations of Tidbinbilla and USNO make "common view" observations impossible, so a two-hop scheme is necessary. In this, the GPS TTU at JPL's Goldstone Radar Net (GRN) has been taking measurements at the same time as USNO (to within ten minutes) and also at the same time as Tidbinbilla (to within one minute) each day. The hydrogen maser based timing system at GRN has a rate of about 20 ns/day but is assumed here to be error-free in relating the two sets of measurements. Then,

$$\text{UTC(USNO MC)} - \text{TID(FTS)} = [\text{UTC(USNO MC)} - \text{JPL}] - [\text{TID(FTS)} - \text{JPL}] \quad (2)$$

Results from GPS TIME averaged over all space vehicles are shown in Fig. 15, and residuals from the quadratic fit whose parameters are given in Table 4 are in Fig. 16. Figure 17 shows the results in the vicinity of the 1 October flying clock measurement, while Fig. 18 gives Allan variances after the quadratic curve has been removed.

Figure 19 copies Figs. 10, 14 and 16 to show the residuals by each of the three methods on one graph. It can be seen that the "common view" results have some spikes not visible in the other results, but that otherwise the results are quite similar. It is noteworthy that the statistics given in Table 4 for "long-arc" using SV TIME and using GPS TIME are very similar, and differ by about 170 ns in offset and 2 ns/day in rate from the "common view" results whose standard error is also somewhat larger. It is therefore evident that the greater atmospheric effects caused by the "common view" method in this case outweigh the clock modelling errors of the "long arc" method.

VI. Transfer to UTC(AUS)

The free-running time scale UTC(AUS) is calculated from TV comparisons between caesium standards and hydrogen masers located mainly in Canberra, Sydney and Melbourne (Refs. 4 and 5). The clocks contributing to UTC(AUS) in the three month period under consideration are summarised in Table 5. Until the GPS receiver was put in to use at Tidbinbilla, the only regular means of comparing UTC(AUS) with adequate precision to the outside world was by flying clock trips every three or four months organised by USNO and Bendix Corporation. It is now possible, however, to measure the relationship on a daily basis, using TID(FTS) as the intermediary.

The results UTC(AUS) – TID(FTS) as published in NATMAP's Bulletin E during July-October are shown in Fig. 20. Combining these with UTC(USNO MC) – TID(FTS) as in Fig. 8 gives UTC(USNO MC) – UTC(AUS) shown in Fig. 21. Residuals from a straight line fit (see Table 4) are shown in Fig. 22 and their Allan variances in Fig. 23. It is immediately seen that the standard errors are almost an order of magnitude greater, which is directly attributable to TV noise. The flying clock trip error is 120 ns of which a substantial proportion can be attributed to the flying clocks themselves.

There is no doubt that, for as long as the GPS TTU remains at Tidbinbilla, a regular, reliable and accurate avenue is available for comparing clocks in Eastern Australia to clocks and time scales overseas. Conversely, it is now possible for 16 or more Southern Hemisphere clocks to be used in the computation of International Atomic Time (TAI) if so desired.

VII. Use of Results to Steer UTC(AUS)

Time scales such as UTC(AUS), UTC(BIH) and UTC(USNO) are calculated in batch mode after-the-event; for example, UTC(AUS) is calculated monthly about a fortnight in arrears. It is thus feasible to accommodate the delays in gathering GPS data and reducing them to a common time such as 0^h UTC in order to include it in the time scale algorithm. This can be done in such a way that the time scale generally follows the external clock or scale to which the GPS measurements are referred, yet continues virtually uninterrupted if the GPS results are unavailable.

Let the time shown by clock i be denoted x_i at a given time, and let the measurement against the local TV (corrected for propagation delay) be l_i:

$$l_i = x_i - \text{TV}, \qquad i = 1, 2, \ldots, n \quad (3)$$

It is desired to calculate a time scale X, e.g., X = UTC(AUS), which appears in the form of a set of results z_i:

$$z_i = X - x_i, \qquad i = 1, 2, \ldots, n. \qquad z_o = X - \text{TV} \quad (4)$$

taking into account weights p_i for each clock according to some prearranged criteria. This is achieved by invoking the fundamental time scale equation (Ref. 6):

$$\sum_{i=1}^{n} p_i (z_i - \hat{z}_i) = 0 \quad (5)$$

where $\hat{z}_i$ is an unbiassed prediction of UTC(AUS) - x_i at the time of observation (Ref. 7). There are n observations and one condition available to solve for the $n + 1$ unknowns z_0, $z_1, z_2, \ldots, z_n$.

Now suppose that an external observation l_T is available on one of the clocks $x_T(T \epsilon_i)$ against the reference time scale R, e.g., R = UTC(USNO MC):

$$l_T = R - x_T \tag{6}$$

and impose the steering condition:

$$X = R \tag{7}$$

i.e., UTC(AUS) = UTC(USNO MC) at the time of measurement. The full set of equations to be solved can then be put into the form of observation equations with a condition, from Eqs. (6), (3) and (5) respectively:

$$\begin{aligned} l_T &= z_T, \qquad T \epsilon i \\ l_i &= z_o - z_i, \qquad i = 1, 2, \ldots, n \end{aligned} \tag{8}$$

$$\sum_{i=1}^{n} p_i (z_i - \hat{z}_i) = 0$$

which can be treated by standard least squares methods; a weight p_T should be assigned judiciously to the external direct observation l_T. This formulation can be readily extended to cater for the hypothetical situation in which several GPS units, or indeed any other international time transfer systems, were operating in Australia.

When this suggestion has been investigated and implemented, UTC(AUS) will be truly a coordinated scale of Universal Time. Removal of the direct external observation(s) l_T would restore it to its current position as a true free-running time scale, more appropriately designated perhaps as TA(AUS) yet regularly monitored against TAI. It is proposed to run both solutions once certain prediction and weighting biases in the UTC(AUS) algorithm have been removed.

VIII. Conclusion

It has been demonstrated that the GPS TTU at Tidbinbilla DSCC can achieve time transfer half way around the world with 20 nanosecond precision over extended periods and with accuracy inside the measurement capability of flying clocks. Little degradation occurs by employing the "long-arc" method rather than the "common view" method because "common view" requires low observing altitudes over these distances, and use of USNO GPS TIME values is almost as good as USNO SV values, besides being easier to obtain.

The first practical application of the method has been the determination of the drift rate of Tidbinbilla's hydrogen maser as 4 parts in 10^{15} per day.

Because the user data processing has proved to be very straightforward, it is highly feasible to include the results in the computation of UTC(AUS) and indeed to steer this time scale to UTC(USNO MC) or to UTC(BIH). It is also now possible for a wider community to use Australian clocks for scientific purposes. We hope that other Southeast Asian and Pacific countries will be encouraged by our results to examine the method very closely.

Acknowledgments

The authors thank Mr. D. Abreu (National Mapping) and Mr. J. Hoyland (Tidbinbilla) for assistance in data processing and communications; Mr. M. Miranian (U.S. Naval Observatory) for sending much USNO space vehicle data; and Mr. H. Sadler (Bendix Corporation) and his associates for their frequent visits with flying clocks.

References

1. Kovach, K. L. (1981): "Using the NAVSTAR Global Positioning System as a Global Timing System," *Proc. Thirteenth Annual PPTI*, Naval Research Lab., NASA Conference Publ. 2220, pp. 133-163.

2. Davis, D. D., M. Weiss, A. Clements, and D. W. Allan (1981): "Unprecedented Syntonization and Synchronization Accuracy via Simultaneous Viewing with GPS Receivers and Construction Characteristics of an NBS/GPS Receiver," *Proc. Thirteenth Annual PTTI*, Naval Research Lab, NASA Conference Publ. 2220, pp. 527-543.

3. Clements, P. A. (1982): "Intercontinental Time and Frequency Transfer Using a Global Positioning System Timing Receiver," *Proc. Fourteenth Annual PTTI*, Goddard SFC, NASA Conference Publ. 2265, pp. 517-527. Also published in the *Telecommunications and Data Acquisition Progress Report*, Vol. 42-75, July – September 1983, p. 98.

4. Luck, J. McK. (1979): "Comparison and Coordination of Time Scales," *Proc. Astronomical Soc. Australia*, Vol. 3, No. 5, pp. 357-363.

5. Woodger, J. R. (1980): "Australian Time and Frequency Standards," *IREE Aust. Precise Time and Frequency Conference Record*, The Institution of Radio and Electronics Engineers Australia, Sydney.

6. Percival, D. B. (1978): "The U.S. Naval Observatory Clock Time Scales," *IEEE Trans. Instrumentation and Measurement*, Vol. IM-27, No. 4, pp. 376-385.

7. Luck, J. McK. (1983): *Construction and Comparison of Atomic Time Scale Algorithms*, Division of National Mapping Tech. Report 32, Canberra.

Table 1. Quadratic fits to observations on satellite clocks

Observation	Offset (a), μs	Rate (b), μs/d	Drift/2 (c), μs/d/d	Average time ($\bar{t}$), MJD	Standard error (σ), μs
TID(FTS-SV(5)	188.236	0.213022	0.00009463	5571.18	0.01
TID(FTS)-SV(9)	15.809	0.075300	0.00022562	5572.31	0.01
UTC(USNOMC)-SV(5)	180.642	0.144689	-0.00009541	5575.53	0.05
UTC(USNOMC)-SV(9)	7.562	0.007514	0.00004740	5575.20	0.03

Table 2. Straight line fits to TID(FTS) – GPS TIME

SV No.	Offset (a), μs	Rate (b), μs/d	Average time ($\bar{t}$), MJD	Standard error (σ), μs	Smoothed	Between
5	5.861	0.08233	5530.75	0.02	5528	5532
6	5.874	0.05275	5530.78	0.03	5528	5532
8	5.887	0.07441	5531.21	0.03	5528	5532
9	5.812	0.07190	5530.86	0.02	5528	5532
5	6.282	0.05981	5537.74	0.04	5532	5542
6	6.296	0.06118	5537.77	0.02	5532	5542
8	6.293	0.06123	5537.69	0.02	5532	5542
9	6.287	0.06314	5538.17	0.02	5532	5542
5	7.269	0.08851	5549.99	0.01	5542	5556
6	7.280	0.09221	5550.01	0.02	5542	5556
8	7.265	0.09207	5549.66	0.02	5542	5556
9	7.221	0.08997	5549.80	0.03	5542	5556
5	8.979	0.06995	5573.14	0.04	5556	5589
6	9.010	0.07002	5573.16	0.02	5556	5589
8	9.017	0.09669	5572.96	0.02	5556	5589
9	8.962	0.07134	5573.24	0.03	5556	5589
5	10.339	0.06487	5592.08	0.03	5589	5594
6	10.368	0.07126	5592.11	0.01	5589	5594
8	10.377	0.06552	5592.04	0.02	5589	5594
9	10.327	0.06830	5592.19	0.02	5589	5594
5	11.060	0.14007	5598.07	0.03	5594	5601
6	11.094	0.13991	5598.09	0.03	5594	5601
8	11.102	0.14669	5598.02	0.03	5594	5601
9	11.047	0.14017	5598.17	0.03	5594	5601
5	11.763	0.08500	5604.35	0.02	5601	5607
6	11.806	0.08724	5604.37	0.02	5601	5607
8	11.846	0.09814	5604.44	0.02	5601	5607
9	11.754	0.09100	5604.45	0.02	5601	5607
5	12.401	0.07943	5612.08	0.02	5607	5618
6	12.431	0.07593	5612.25	0.02	5607	5618
8	12.455	0.07508	5612.18	0.02	5607	5618
9	12.384	0.07606	5612.33	0.01	5607	5618

Table 3. Straight line fits to UTC(USNOMC) – GPS TIME

SV No.	Offset (a), μs	Rate (b), μs/d	Average time ($\bar{t}$), MJD	Standard error (σ), μs	Smoothed	Between
5	0.054	0.02183	5529.91	0.02	5528	5532
6	0.075	0.01957	5530.83	0.01	5528	5532
8	0.078	0.02276	5531.24	0.02	5528	5532
9	0.054	0.02202	5529.99	0.01	5528	5532
5	0.115	0.00188	5537.00	0.01	5532	5542
6	0.127	0.00565	5537.76	0.01	5532	5542
8	0.108	0.00003	5537.70	0.03	5532	5542
9	0.114	0.00349	5536.98	0.01	5532	5542
5	0.360	0.03048	5549.50	0.01	5542	5556
6	0.388	0.02952	5550.24	0.01	5542	5556
8	0.392	0.03226	5549.92	0.02	5542	5556
9	0.394	0.03135	5549.98	0.02	5542	5556
5	0.630	0.00270	5582.29	0.01	5556	5589
6	0.610	0.00207	5582.30	0.02	5556	5589
8	0.619	0.00197	5582.36	0.02	5556	5589
9	0.619	0.00197	5582.35	0.02	5556	5589
5	0.637	0.00948	5592.45	0.03	5589	5594
6	0.613	0.00622	5592.39	0.03	5589	5594
8	0.639	-0.01150	5592.44	0.01	5589	5594
9	0.620	0.00693	5592.44	0.03	5589	5594
5	0.911	0.05304	5598.45	0.02	5594	5601
6	0.886	0.05647	5598.37	0.02	5594	5601
8	0.920	0.05879	5598.43	0.06	5594	5601
9	0.897	0.05507	5598.42	0.02	5594	5601
5	1.101	0.00862	5604.94	0.02	5601	5607
6	1.087	0.00853	5604.85	0.01	5601	5607
8	1.109	0.00625	5604.86	0.01	5601	5607
9	1.090	0.00695	5604.91	0.01	5601	5607
5	1.084	-0.00690	5613.41	0.01	5607	5618
6	1.074	-0.00405	5613.61	0.02	5607	5618
8	1.091	-0.00438	5613.64	0.01	5607	5618
9	1.081	-0.00383	5613.35	0.02	5607	5618

Table 4. Quadratic fits to UTC(USNO,MC) − TID(FTS)

Medium	Offset (a), μs	Rate (b), μs/d	Drift/2 (c), μs/d/d	Average time ($\bar{t}$), MJD	Standard error (σ), μs
SV(5)	-8.485	-0.06867	-0.0001853	5574.96	0.019
SV(6)	-8.334	-0.06841	-0.0001663	5572.08	0.010
SV(8)	-8.478	-0.06904	-0.0001788	5574.46	0.019
SV(9)	-8.510	-0.06924	-0.0001817	5575.86	0.022
All SV	-8.387	-0.06849	-0.0001764	5573.39	0.017
GPS TIME, SV(5)	-8.483	-0.06893	-0.0001872	5574.96	0.023
GPS TIME, SV(6)	-8.326	-0.06851	-0.0001752	5572.08	0.016
GPS TIME, SV(8)	-8.472	-0.06925	-0.0001896	5574.46	0.022
GPS TIME, SV(9)	-8.523	-0.06969	-0.0001830	5575.86	0.017
GPS TIME, all SV	-8.385	-0.06872	-0.0001869	5573.39	0.015
Common view	-8.559	-0.07054	-0.0001614	5573.50	0.023
UTC(USNO, MC) − UTC(AUS) via GPS TIME, all SV	-12.19	-0.01680	0.0	5576.92	0.088

Table 5. Contributors to UTC(AUS), July–October 1983

Organization	Location	Time standards
Tidbinbilla DSCC	Canberra	1 hydrogen maser 2 HP caesium standards
Division of National Mapping	Canberra	2 HP caesium standards
Orroral Valley STDN	Canberra	1 HP caesium standard
CSIRO National Measurements Lab	Sydney	2 hydrogen masers 3 HP caesium standards
Royal Australian Navy	Sydney	1 HP caesium standard
TELECOM Australian Research Lab	Melbourne	5 HP caesium standards

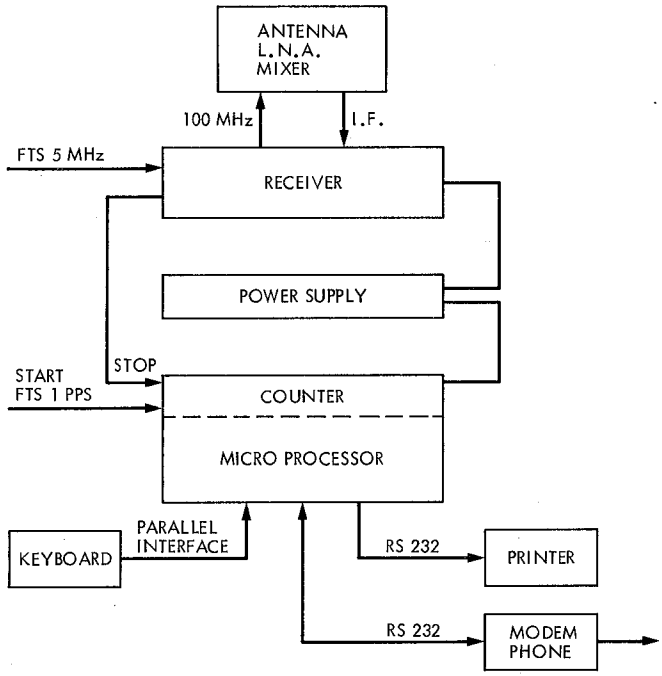


Fig. 1. GPS equipment configuration at Canberra Deep Space Communications Complex, Tidbinbilla

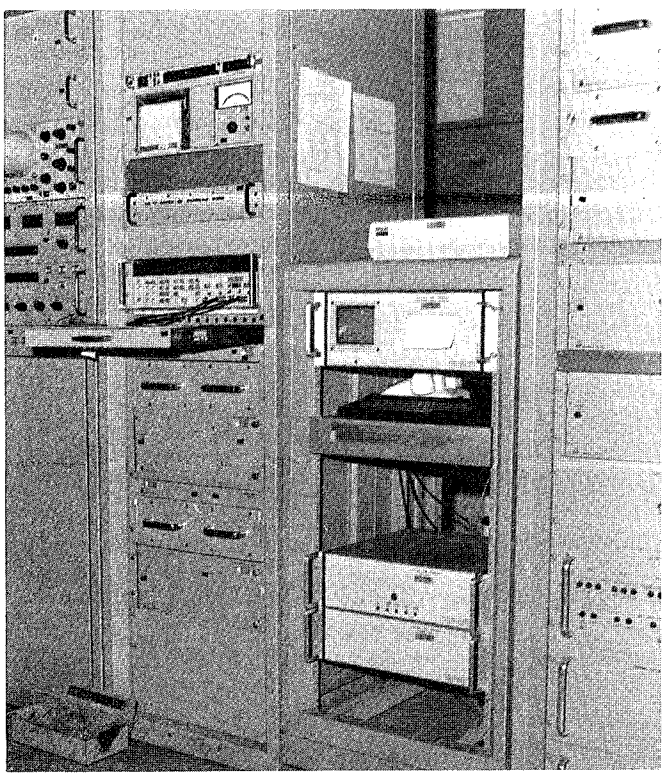


Fig. 2. GPS receiver installation at Tidbinbilla

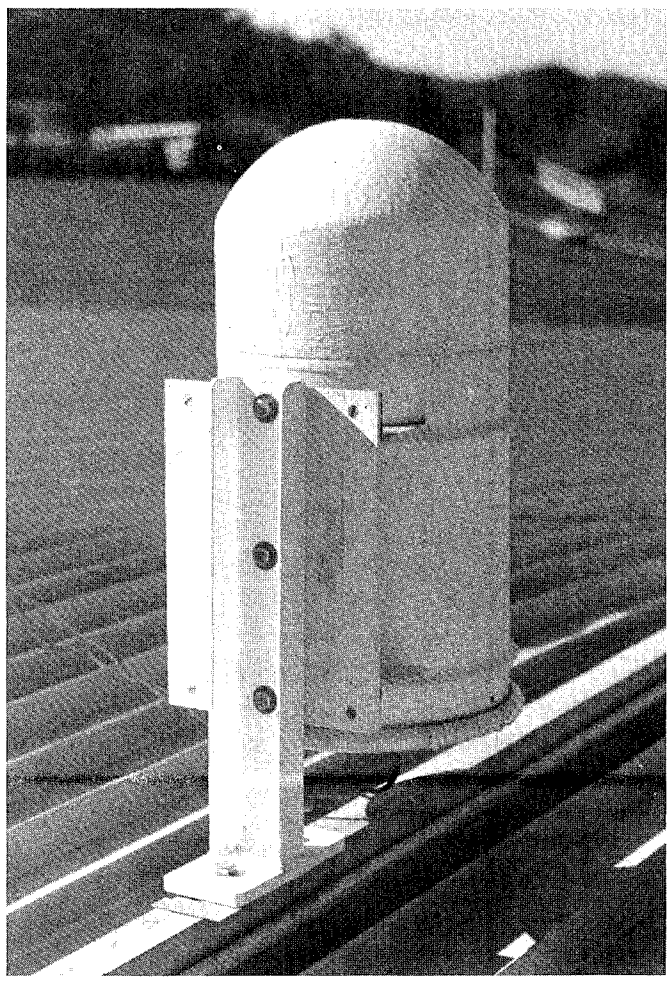


Fig. 3. GPS antenna housing on roof of CDSCC

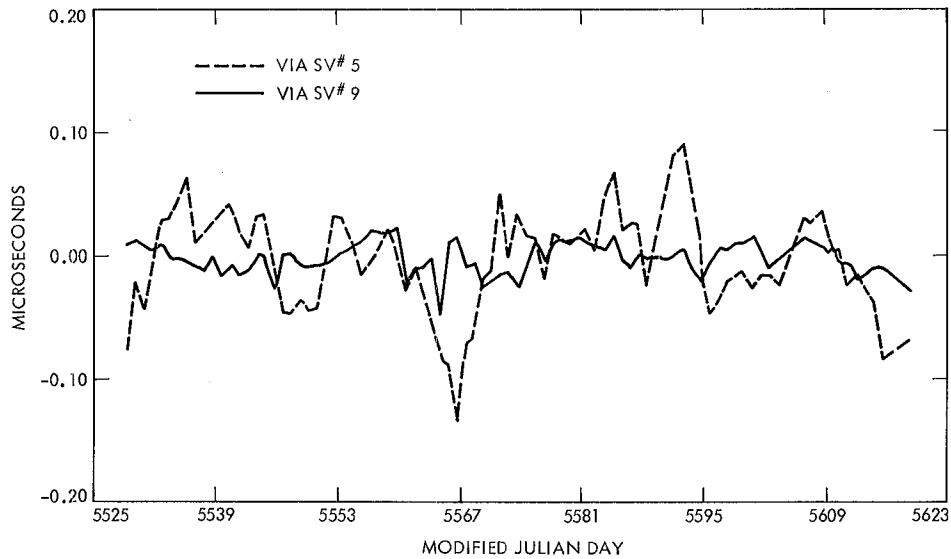


Fig. 4. Residuals from quadratic fits through raw data TID(FTS) – SV TIME, space vehicles 5 and 9 (see Table 1)

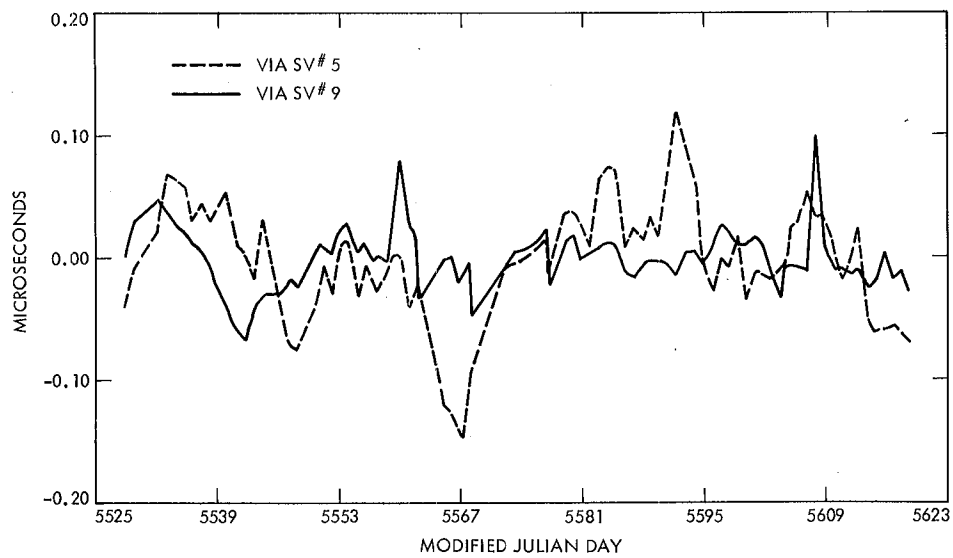


Fig. 5. Residuals from quadratic fits through raw data UTC(UNSNO,MC) – SV TIME, space vehicles 5 and 9 (see Table 1)

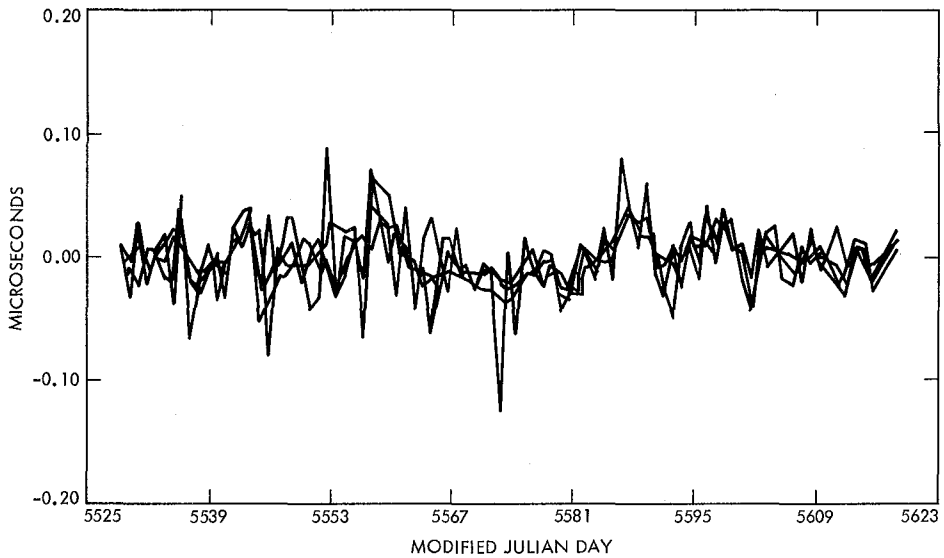


Fig. 6. Residuals from straight line fits through raw data TID(FTS) – GPS TIME, space vehicles 5, 6, 8, 9 segmented at MJDs 5532, 5542, 5556, 5589, 5594, 5601, 5607 and 5618 (see Table 2)

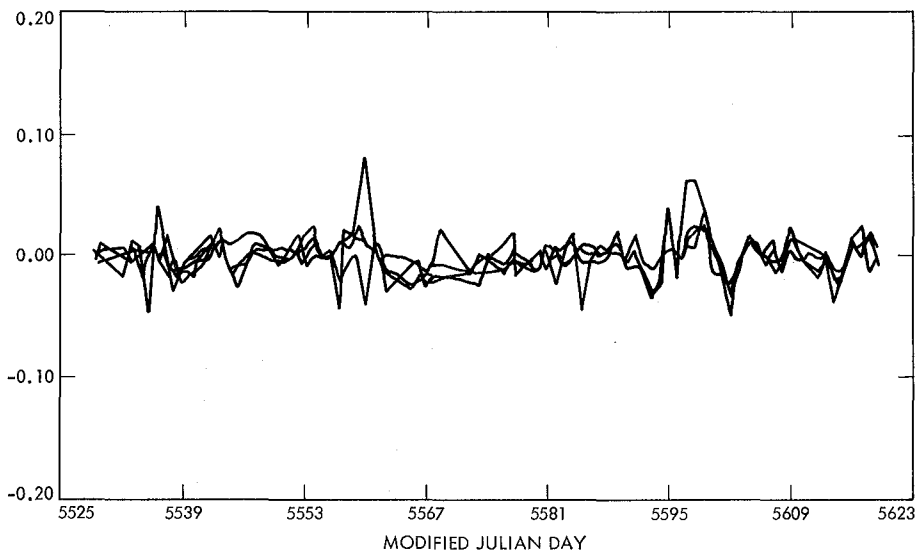


Fig. 7. Residuals from straight line fits through raw data UTC(USNO,MC) – GPS TIME via space vehicles 5, 6, 8, 9, as in Fig. 6 (see Table 3)

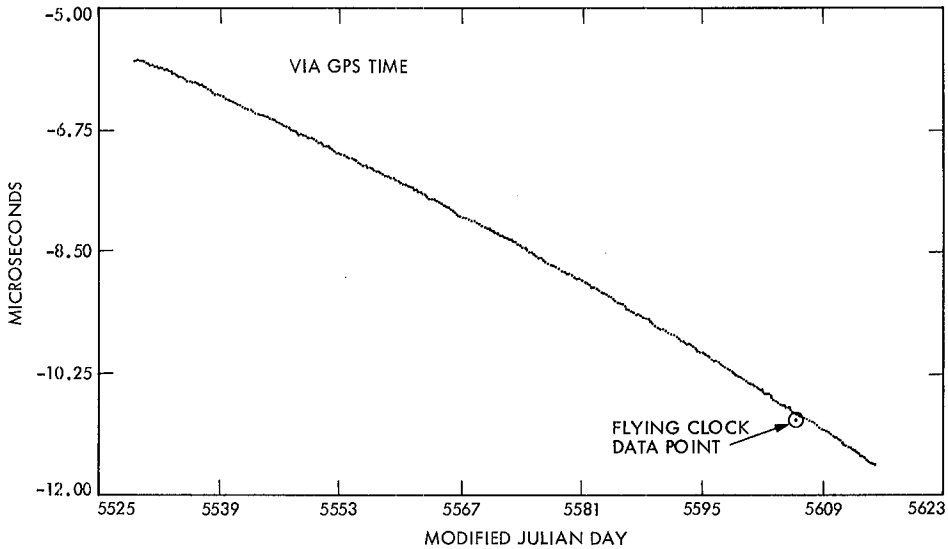


Fig. 8. Interpolated UTC(USNO,MC) – TID(FTS) using GPS TIME from space vehicles 5, 6, 8, 9 together

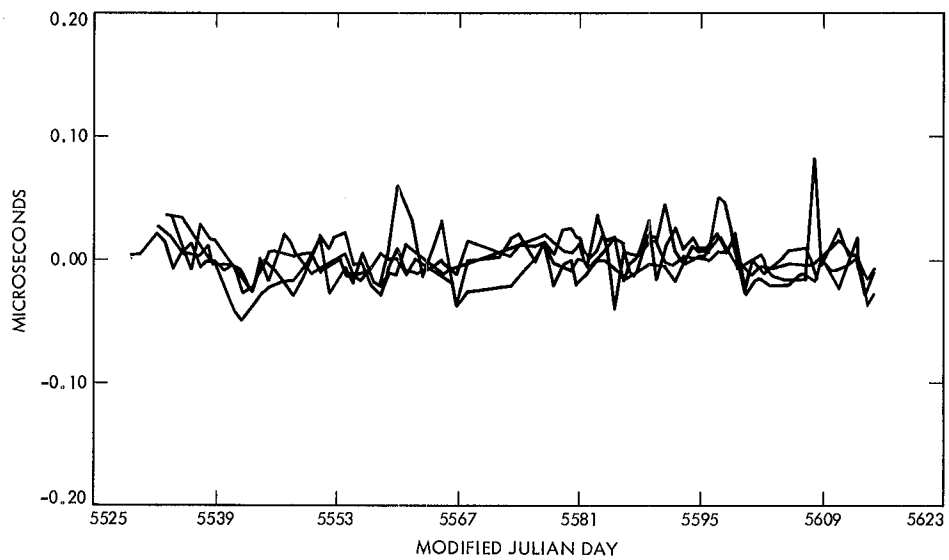


Fig. 9. Residuals from quadratic fits through interpolated data UTC(USNO,MC) – TID(FTS), using SV TIME from space vehicles 5, 6, 8, 9 (see Table 4)

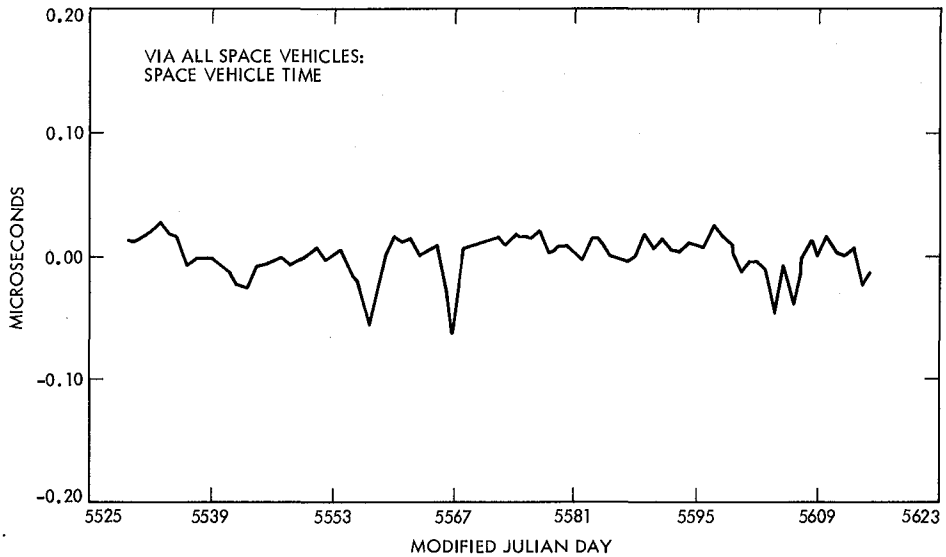


Fig. 10. Residuals from quadratic fit through interpolated data UTC(USNO,MC) – TID(FTS) reduced to O*h* UTC and averaged over SV TIME from space vehicles 5, 6, 8 and 9 (see Table 4)

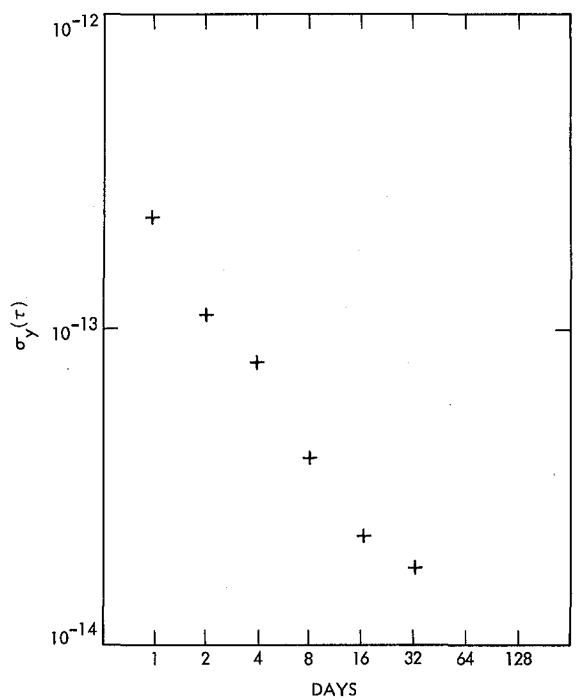


Fig. 11. Allan variances of residuals in Fig. 10

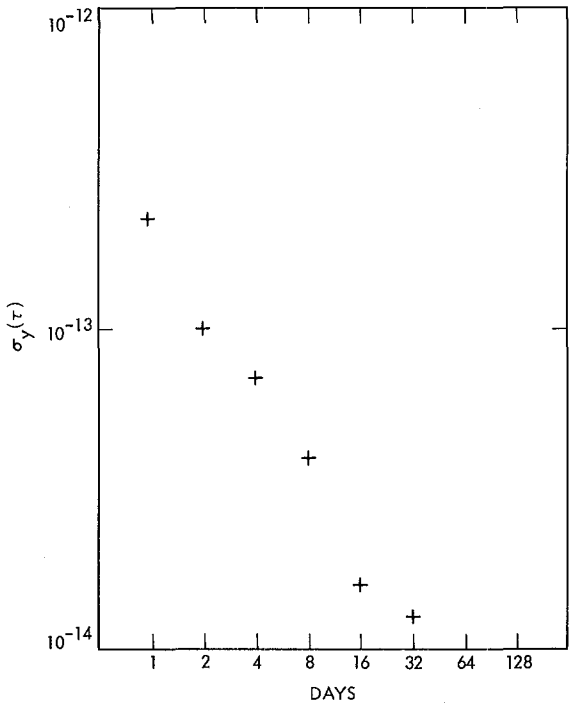


Fig. 12. Allan variances of residuals in Fig. 14

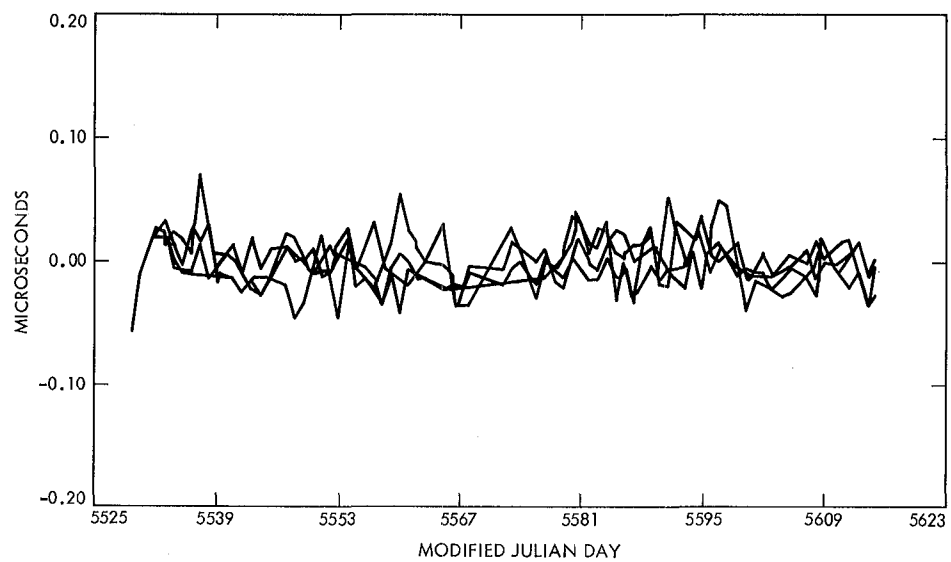


Fig. 13. Residuals from quadratic fits through interpolated data UTC(USNO,MC) – TID(FTS) using GPS TIME from space vehicles 5, 6, 8 and 9 (see Table 4)

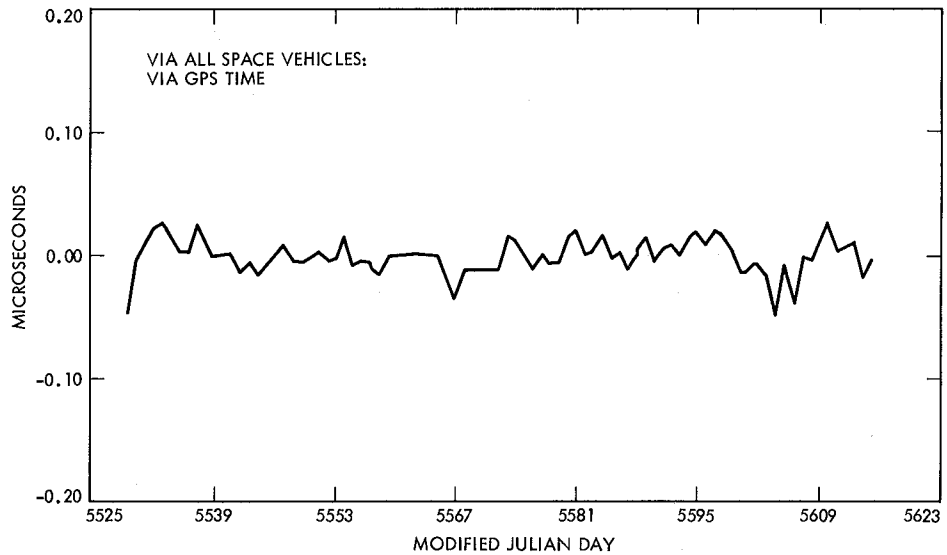


Fig. 14. Residuals from quadratic fit through interpolated data UTC(USNO,MC) – TID(FTS) reduced to O*h* UTC and averaged over GPS TIME from space vehicles 5, 6, 8 and 9 (see Table 4)

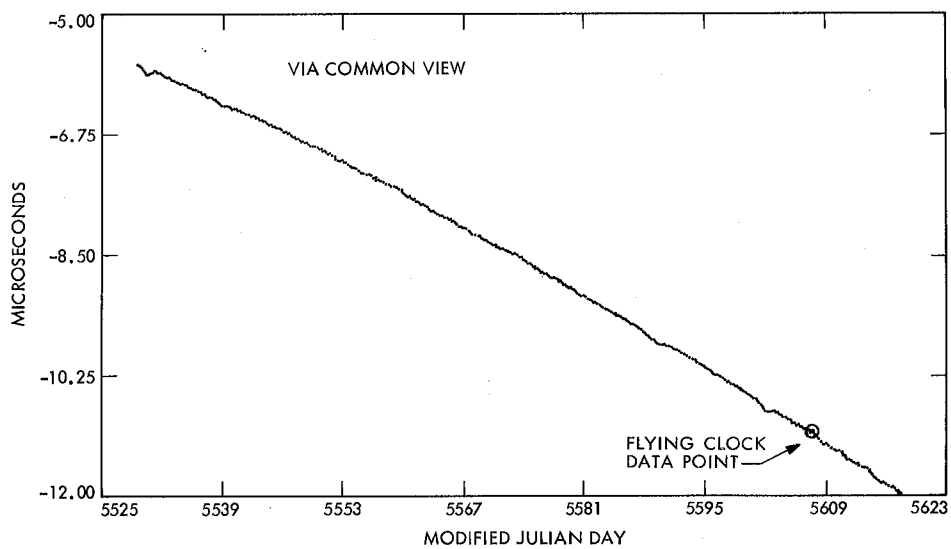


Fig. 15. Common view results UTC(USNO,MC) – TID(FTS) showing flying clock result on 1 October 1983

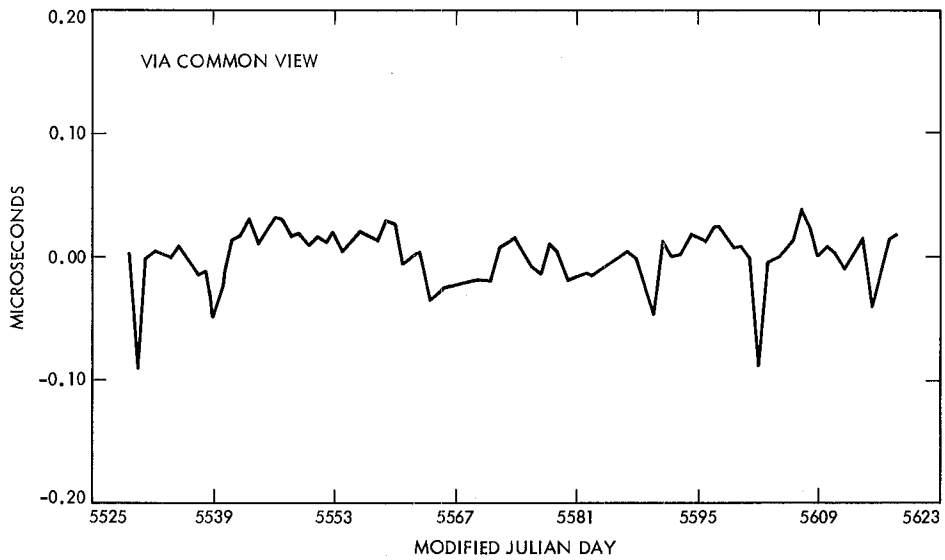


Fig. 16. Residuals from quadratic fit through common view results UTC(USNO,MC) – TID(FTS), averaged over GPS TIME from space vehicles 5, 6, 8 and 9 (see Table 4)

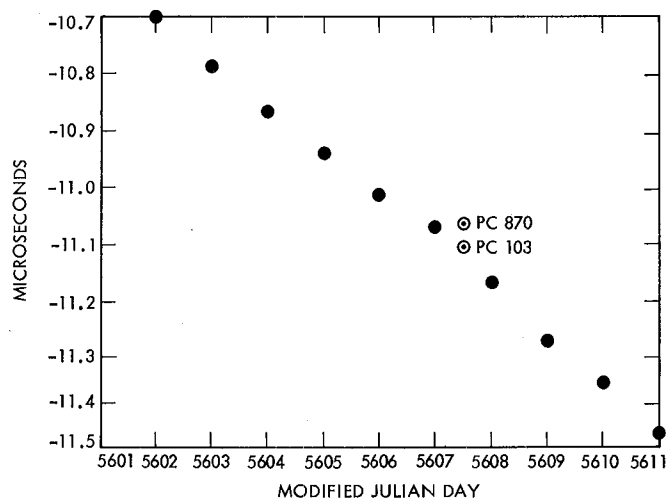


Fig. 17. Common view UTC(USNO,MC) – TID(FTS) in vicinity of travelling clock data points on 1 October 1983; "consolidated" from GPS TIME results on 0*h* UTC

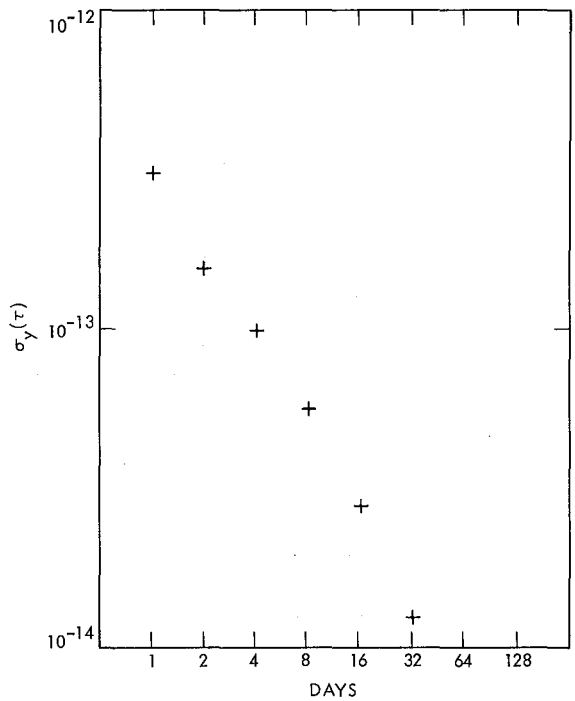


Fig. 18. Allan variances from quadratic fit through data of Fig. 17 (see Table 4)

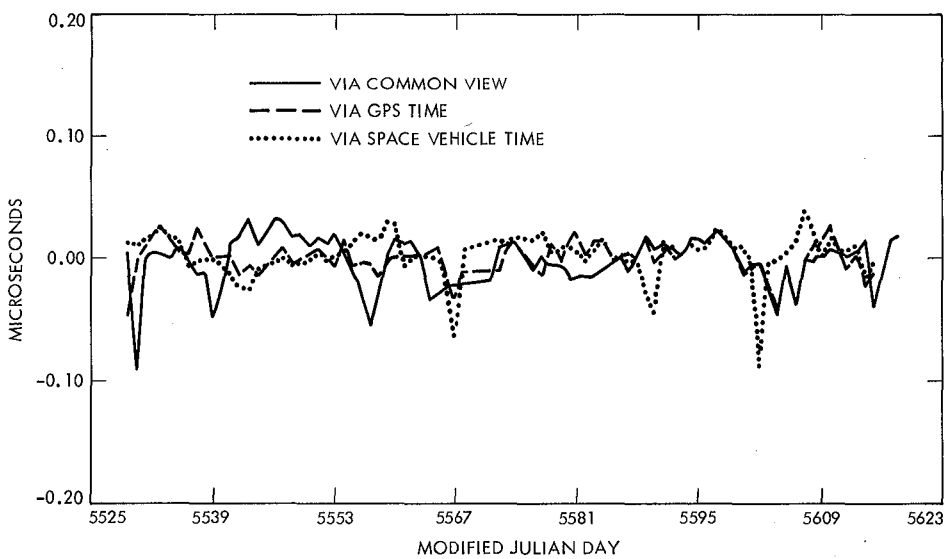


Fig. 19. Composite of residuals from quadratic fits through UTC(USNO,MC) – TID(FTS) via common view (Fig. 16), long-arc GPS TIME (Fig. 14) and long-arc SV TIME (Fig. 10)

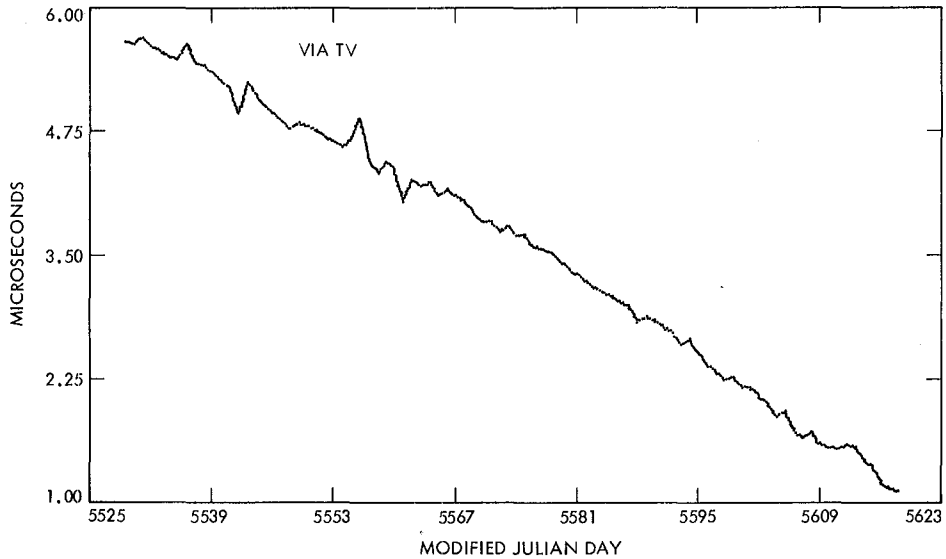


Fig. 20. UTC(AUS) – TID(FTS) via TV as published in NATMAP Bulletin E

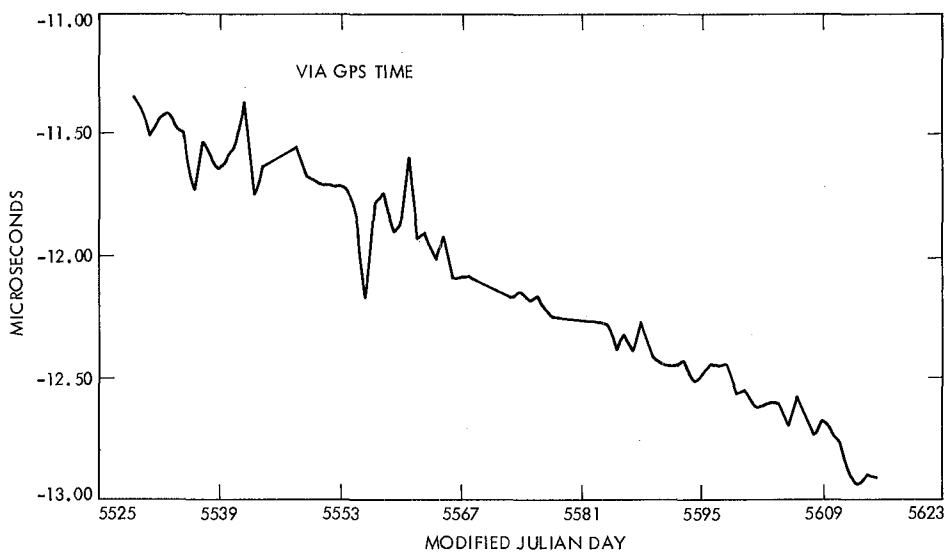


Fig. 21. UTC(USNO,MC) – UTS(AUS) derived from long-arc GPS TIME results to Tidbinbilla (Fig. 8) and Bulletin E (Fig. 20)

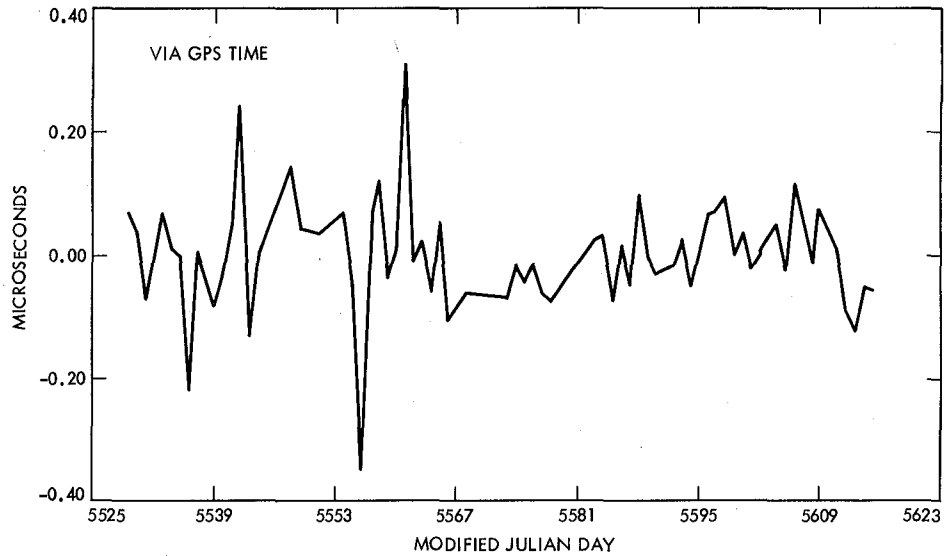


Fig. 22. Residuals from straight line fit through UTC(USNO,MC) – UTC(AUS) shown in Fig. 21 (see Table 4)

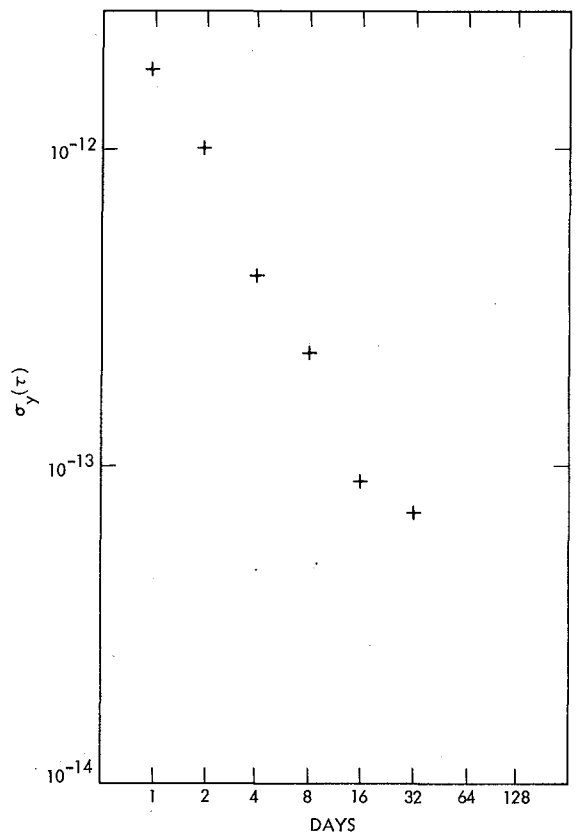


Fig. 23. Allan variances of the residuals UTC(USNO,MC) – UTC(AUS) shown in Fig. 22

Implementation of Large Antennas for Deep Space Mission Support

R. Stevens
Office of Telecommunications and Data Acquisition

The large antennas of the DSN support reception of low-power telemetry signals from spacecraft (S/C), transmission of high power commands to S/C, and navigation of S/C by precision radio metric data. The specification and design of the antennas have been driven by the requirement to support those functions with high reliability. The number of antennas required in the DSN is determined by the number of S/C to be supported and their level of activity. A given size antenna aperture can be realized with a single element or by arraying smaller elements with the same total area. That approach can be applied to meeting many DSN requirements. There is a cost vs capability trade-off in arrayed vs single element designs. The operating microwave frequency is an important parameter for the antenna. Some of the communications and radio metric data functions are much improved at higher frequencies, e.g., X- vs S-Band. Current technology allows arraying for reception by antennas that are far apart. Thus, for example, major radio astronomy antennas can be arrayed with DSN antennas to increase reception capability during important scientific events, e.g., Voyager encounters with Uranus and Neptune. This article examines the specification, design, and development of DSN antennas in the context of the above considerations.

I. Introduction

This report describes the evolution of DSN large antenna capability for deep space mission support. The reasoning behind making particular design or configuration choices is given.

II. Deep Space Mission Generic Requirements

Deep space S/C missions require support of these principal technical functions:

High rate telemetry

Low rate telemetry

Routine command

Emergency command

Emergency telemetry/weak signal search

Generation of precision radio metric data

Sophisticated flight radio science experiment instrumentation

Some of the functions require support more often than others, but all of them must be supported at some time or other. Most of the functions put extraordinary requirements on the technical performance of the ground antennas.

In addition to deep space S/C missions, the DSN supports, or will support, other users whose antenna technical performance requirements are similar to those required for deep space mission support:

Planetary Radar

Radio Astronomy, especially VLBI, including earth orbital missions

The Search for Extraterrestrial Intelligence (SETI) Program

The mission support requirements have driven the DSN designs, as will be illustrated in following sections of this report.

III. Antenna Capacity to Meet Mission Support Requirements

The capacity, i.e., the number of separate antennas required, is determined by the number of individual supported missions and their required support level.

A. S/C Generic Requirements

All currently active deep space S/C require continuous two-way communications in high-activity phases, or daily downlink and frequent two-way passes in cruise phase. Thus, each high-activity S/C requires about 90 antenna passes per month (two or three times that when arraying is required), and each cruise S/C requires about 30 to 60 passes per month.

B. Support Capacity of DSN Antennas

A S/C support pass takes an average of 10 station hours, including pre- and post-pass calibrations (some passes are as short as two hours; some are as long as 13 hours). Thus, depending on the view periods of S/C supported, an antenna can support one to two passes per day (approximately 30 to 60 per month) plus allow time for station maintenance and repair. Based on empirical experience, DSN scheduling uses, for planning projections, an average capacity of 46 passes per antenna-month. For example, the average support capacity of the mid-80s (Voyager Uranus era) eight-antenna network of three 64-m and five 34-m antennas is about $8 \times 46 = 368$ passes per month; that of the proposed nine-antenna late-80s network (Voyager Neptune era) is about $9 \times 46 = 414$ passes per month.

C. Example of Antenna Support Capacity

As a simple illustration of network capacity vs requirements arithmetic, consider the following scenario, representative of DSN mission support in the late 80s prior to operational support of the SETI Program. There are eight missions supported – Pioneers 10, 11, 12; Voyagers 1, 2; Galileo; ISPM; VRM. Of these, assume one is in a high-activity phase and the other seven are in cruise. The total of passes required per month is approximately $1 \times 90 + 7 \times 45 = 405$ (an average number of 45 passes per month is used for the cruise missions, approximately satisfying their minimum requirements). The nine-antenna network has adequate average capacity (414 passes per month) to support the requirements; the eight-antenna network does not (368 passes per month).

D. The Mix of Antenna Sizes Required for Mission Support

All passes do not require the same ground antenna capability as assumed in the simple approach above; some support requires the largest antennas (64 m); other support is satisfied, and less expensively, by the 34-m antennas. There is the question of how many large vs how many smaller antennas should be provided.

The following considerations are involved. A deep space S/C is designed to provide maximum telemetry capability at its distant prime mission target. Also, at least to the distance when the S/C is near the target, both telemetry and command links must be supportable through the S/C low gain antenna that provides communications during S/C maneuvers or non-standard S/C behavior. At those times, in consonance with the mission telecommunications system design, the S/C will require support from the DSN's most capable antennas, transmitters, and receivers. At most other times and places, the S/C telemetry and command links have considerable excess of capability. Data rates required are typically much lower during S/C cruise periods, and at least during transit to the encounter target, the distance is less.

Also, the relative sizes of the large vs small antennas affect the ratio of the number of large vs small antennas that should be provided. For example, the Pioneer 11 S/C, which is departing from the solar system at the rate of about 2.5 A.U. per year, can be supported by a 34-m antenna until late 1986. After that time its support will be shifted to a 64-m antenna. If the smaller antenna were 26 m rather than 34 m, the shift to the 64-m antenna would have to occur by early 1984.

As another example, consider support of typical planetary orbiter missions. The distance between the Earth and a planet varies regularly. That produces a range of variation in com-

munication link capability as shown below for the nearby planets:

Mercury	7 db
Venus	16 db
Mars	14 db
Jupiter	3 db
Saturn	2 db

If the S/C-ground communication system is designed to perform adequately at the time of maximum Earth-to-planet distance, it will have excess capability at other times. For example, if the link uses a 64-m antenna for support to maximum Earth-to-planet distance, an approximately 5 dB less capable 34-m antenna could support Venus and Mars orbiters more than half the time; a Mercury orbiter, for a brief time around inferior conjunction (7 - 5 = 2 dB margin at minimum distance); but Jupiter and Saturn orbiters, not at all.

By various means a quantitative estimate can be made of the relative amounts of support required from the large and small DSN antennas for typical deep space missions. One simple way is to examine the ensemble of technically laundered (i.e., S/C supported by the minimum satisfactory DSN capability) requirements established by the projects. We have done that for a sample of one-month periods – the month of August, for the years 1985 through 1993. The results show that, on average, one-third of the total passes requires support from the larger DSN antennas.

This concludes the discussion of the general planning approach to estimating the numbers of antennas and mix of sizes needed to meet projected support loading requirements. The next section addresses key design vs cost considerations that the DSN has encountered, and anticipates, in the evolution of its antenna facilities.

IV. Antenna Facility Design Approaches

A. The Beginning

The first DSN antennas for deep space mission support were 26 m in diameter with prime focus feeds. They were adapted from existing radio astronomy designs. They were used at L- and S-band frequencies to support low-yield encounter missions to Venus and Mars in the early 60s.

B. Development of the 64-m Antennas

Beginning in 1962, the DSN studied antenna designs to enable significant imaging yields from Mars. The studies concluded that the projected telemetry, command, and radio metric data requirements could best be met by apertures in the 50- to 80-m diameter range.

For significantly smaller aperture elements, used in an arrayed configuration, the costs rose steeply, dominated by electronics and O&M costs and their uncertainties; for significantly larger aperture elements the available design and construction technology was not considered adequate to support implementation with acceptable cost and performance risk. At that time the 64-m Parkes, Australia, radio astronomy antenna was the largest microwave antenna design that had been successfully executed.

The favored configuration from the many considered was a fully-steerable, symmetrical cassegrainian-fed, paraboloid operable at very low receiving system noise temperature. The prime operating frequency was S-band. Also, the studies argued that investment in DSN ground aperture provided cost effective communication increase compared with S/C capability increase. These arguments supported NASA's decision to design and implement the DSN 64-m subnet. It was started in 1963, finished in 1973.

C. More on Ground Antenna Diameter for Deep Space Mission Support

The subject of optimum antenna diameter vs cost addressed in the early studies has since been revisited several times by the DSN; it will come up several times in the present discussion. What follows will distill the proximate results of the studies.

The cost-optimum diameter depends on the technical requirements to be met by the antenna facilities. That is shown in Fig. 1, a plot of cost-optimum antenna element diameters to obtain an arbitrary increment of X- and S-band receiving, or transmitting or both, effective aperture. In Fig. 1, the horizontal axis is the approximate cost of the electronics and other equipment and facilities required with each aperture element to support the technical functions provided by the total aperture. The technical functions supported are also indicated on the horizontal axis; they are those listed in Section II of this paper. The vertical axis is the optimum diameter of the elements comprising the total aperture. Smaller elements are arrayed to make up the total aperture increment.

There is one conceptual uncertainty in Fig. 1. That involves arraying for transmission. Transmitter function costs are assumed the same for each element of an array, independent of the number, n, of elements. Actually, individual element transmitter costs decrease with n, while cost of radio frequency phase coordination of the individual elements increases with n. But the actual relationship is not known because the technology to phase relatively remote apertures for transmission is not yet in practice. The necessary practical technology can be, probably will be, available in five to ten years; meanwhile, the assumption used here is expected to be pretty close

to the final outcome. At any rate, it is clear that for the DSN's purposes the only practical way in the next many years to achieve extremely high power transmission is with a single large aperture, so the issue is somewhat peripheral to the present discussion.

Return to Fig. 1, which, as stated, displays cost-optimum array element diameter vs the technical functions (or electronics cost) associated with each element of the array. The optimum diameter to provide all of the required DSN technical functions of Section II is in the range of 60 to 70 m, consistent with the early studies supporting the 64-m subnet implementation.

Another point should be made: achieving a cost optimum diameter to realize an aperture capability is important, but, within limits, it is neither critical nor totally unforgiving. Figure 2 displays the character of the cost vs diameter function – it gets steep (but nowadays predictably so) if the elements are much too small; and if the element size gets too large to be supported by mature design and construction technology, it gets simply unpredictable.

D. Consideration of Very Large Apertures

In the early 70s the DSN studied options to further increase communications capability for projected high-yield missions to more distant planets. By that time it was considered that antenna design and construction technology could support antennas for DSN use in the 100- to 130-m class with acceptable risk.

In the studies, it was assumed that such expensive antenna facilities, to be justifiable, would have to be operated continuously, either stand-alone as independent stations or as an array. Thus, operating costs for multiple antenna arrays increased relative to the model in the earlier study. Also, the complement of electronics being used on the 64-m DSN antennas was more extensive and costly than projected in the earlier study. As a result, cost-optimum element diameter increased to 120 to 130 m. However, the facilities were indeed very expensive.

At the time, it was believed by NASA that projections of the future level of activity in the planetary exploration program would not realize the need for continuous use of such a large capability. Therefore, the pursuit of 100- to 130-m class antennas and their justification rationale was put aside.

E. Consideration of Use of Higher Microwave Frequencies

Also, during the early 70s, the potential of X-band frequencies for providing increased telemetry capability using the existing DSN antenna facilities was studied. An experimental X-band planetary radar was first implemented on the 26-m Venus site antenna and later on the DSS 14 64-m antenna. The use of X-band produced the increased capability sought by the radar experimenters, but more importantly to the DSN, demonstrated the technical performance improvement and the operational viability of X-band for deep space mission support. X-band reception was initially implemented on the 64-m subnet to support Mariner Venus-Mercury 1973 and Viking Mars 1975 technology demonstration and radio science experiments.

It became clear that the increment of telemetry capability needed to support effective imaging data return from the middle planets, Jupiter and Saturn, could be achieved by use of X-band with relatively modest costs for changes to the S/C hardware and the DSN facilities.

F. Upgrade of 26-m S-Band to 34-m S/X-Band Subnet

To meet the commitment to Voyager for prime telemetry support at X-band, it was necessary to implement X-band on one subnet of smaller antennas. The combination of S- and X-band technical requirements could be met by an approximate 34-m diameter antenna. The approach selected was to increase the diameter and upgrade the performance of the DSS 12, 42, and 61 26-m subnet antennas. Although the upgrade to 34-m X-band capability of the original 26-m S-band design involved major modifications, it was still significantly less expensive than implementing new antennas:

Approach	Cost per Antenna (FY81 $M)
S/X 26-34-m Conversion	4.8
Implement New 34-m	5.6

The X-band capability was used with excellent results as the prime telemetry link by Voyager for Jupiter and Saturn encounter imaging and general science data. It will be used for Galileo, ISPM, and VRM missions.

G. Initial Study of 64-m Upgrade

In the mid 70s, a broad study was made of X-band telemetry performance upgrade options for the 64-m antennas. Of the menu, which included the currently proposed diameter increase to 70 m and high efficiency shaped surfaces, only the most economical and cost effective items were selected for implementation. These were improved X-band masers and microwave feed system components. Those upgrades provided a needed increment of performance for Voyager at Jupiter. The other options were set aside as not needed or affordable in the near term.

H. Early Array Applications

Technology for arraying the 64-m with the 34-m antennas was under development in the late 70s. Arraying was used experimentally at Goldstone to enhance the Voyager Jupiter encounter and as a committed capability at all three complexes for the Voyager Saturn encounter. Arraying of the antennas during the encounters provided a very useful additional 1 dB of capability, and because it used available DSN antenna facilities, it was relatively very inexpensive to implement.

I. Initial Considerations for Outer Planet Imaging Mission Support

1. Background. In the late 70s, the DSN considered the projected support requirements for the middle and late 80s. In principle the DSN had, or was developing, adequate technical performance to support all requirements of all seriously proposed inner and middle planet missions (i.e., Galileo, ISPM, VOIR, Halley's Comet). However, for outer planet high-yield imaging missions, it had no visible approach to providing the downlink telemetry capability that would be required.

Specifically, the extension of the Voyager 2 mission to encounter Uranus in 1986 was becoming a recognized possibility, but the DSN was without a seriously proposed means for supporting the encounter telemetry. The short fall was not 1 or 2 dB; more like 4 to 6 dB was needed to support the minimum imaging data rate at Uranus (6 dB is the increment of signal loss between Saturn and Uranus).

2. The large advanced antenna station proposal. The basic task of the design study team, which became known as the Large Advanced Antenna Station (LAAS) Project, was to identify the most cost effective solution to the problem of providing an X-band telemetry reception capability at least 5 dB greater than that of a DSN 64-m antenna. The LAAS Project considered economical antenna designs in the diameter range from 25 to 100 m and arrived at an array configuration of 34- to 40-m elements as the proper solution to the problem (cf. the region of Fig. 1 for single frequency telemetry reception, approximately $2M/element of electronics).

Three considerations developed in the early 80s that, taken together, are considered to have caused the LAAS pursuit to be set aside.

First, the planetary program was at least temporarily divested of high yield outer planet missions in the 90s that would have required the singular capability of the LAAS; the Voyager 2 at Uranus and possibly Neptune remained as the only likely mission requiring the capability for the next 20 years.

The second factor was emergence of the feasibility of implementing effective image data compression on board the Voyager 2 S/C to support Uranus encounter. The net effect of the data compression is an approximate 4 dB increase in imaging yield.

The third factor was the emergence and acceptance in principle, by the DSN and the Voyager Project, of the concept of using non-NASA-owned antenna facilities to augment the DSN's capability for encounter support.

The combination of S/C on-board data compression, arrayed use of the Parkes, Australia, Radio Astronomy Observatory 64-m antenna, and the addition of two arrayed DSN 34-m X-band high efficiency antennas can provide a satisfactory science data yield for Voyager at Uranus. That is the planned support configuration.

3. The 34-m high efficiency antennas. What resulted in the addition of the two new 34-m X-band antennas, just cited, started as a proposal for six by the MK IVA Project. The principal justification for the antennas was Voyager Uranus encounter support. This followed withdrawal of the more ambitious LAAS initiative, but the S/C data compression and non-NASA facility support paths were still perceived as quite uncertain.

The initial approach considered in-place modification to 34 m of existing NASA 26-m antennas, as was previously done in the S-X upgrade project. Configuration architecture studies by the MK IVA Project showed that it was clearly cost effective and technically acceptable to co-locate all antennas of a complex at a single site area. The added cost of relocating the antennas is quickly recovered. The MK IVA Project selected that configuration.

Cost and performance factors were compared for moving and upgrading the existing 26-m antennas vs implementing new 34-m antennas. The cost comparison is:

Approach	Cost per Antenna (FY81 $M)
Relocate and do S/X 26 to 34m Conversion	6.3
Implement New 34 m	5.6

Performance factors favored the new antennas. The MK IVA Project decided to implement the new antennas. Budgetary trade-offs coupled with growing faith that data compression and Parkes arraying would make it for Voyager Uranus resulted in reducing the number of antennas from six to two.

The above discussion has summarized the evolution to the present of DSN large antenna capability for deep space mission support. The next section presents some future developments that are currently being seriously studied and proposed for mission support in the late 80s and early 90s.

V. Proposed Evolution for The Voyager-Neptune Era

A. Background

The Voyager signals from Neptune will be 3.5 dB (more than a factor of 2) weaker than from Uranus. To achieve a comparable data yield, that deficit must be made up. There is no further potential improvement evident from the S/C; in fact, with the aging S/C, the opposite is true. Increasing the ground aperture is the only way evident to reduce the deficit.

Except for the Voyager 2 Neptune encounter, the DSN does not now perceive any S/C mission requirement for truly exceptional telemetry support throughout the mid 90s. Therefore, it is not considered reasonable for NASA to undertake making up the entire deficit by the necessary increase in DSN aperture. It would require more than doubling the present DSN aperture, an approximate $200M enterprise.

B. The Concept for 3+ dB X-Band Telemetry Ground Aperture Increase

The current DSN proposal to support Voyager 2 at Neptune is by limited augmentation of DSN antenna facilities combined with arraying of several major non-NASA antenna facilities at near-encounter. The approach is characterized below.

For Voyager Uranus encounter support, the DSN will apply three 64-m and five 34-m antennas providing 4.45 Aperture Units (Ap.U.)[1]; that will be augmented by 0.8 Ap.U. from the Parkes 64-m antenna for a total of 5.25 Ap.U.

For the Voyager Neptune encounter, the DSN proposes to increase the NASA-owned aperture to 6.45 Ap.U. by upgrading the existing 64-m antennas (adds 1.65 Ap.U.) and adding a new 34-m high efficiency antenna at Madrid (adds 0.35 Ap.U). Also, for three weeks around encounter, the DSN will be augmented by 4.47 Ap.U. provided by major non-NASA antenna facilities: one-half of the U.S. NRAO VLA, the Parkes 64-m and the Japanese 64-m antennas (the Bonn 100-m radiotelescope is also being considered).

[1] Aperture Unit = current X-band effective aperture of a DSN 64-m antenna @25 K system temperature.

For the close encounter period, the total aperture applied will be 6.45 Ap.U. (NASA/DSN) + 4.47 Ap.U. (non-NASA) for a total of 10.92 Ap.U. Thus, in the proposed approach, the ratio of aperture applied at Neptune vs Uranus encounters is 10.92/5.25 = 2.1 = +3.2 dB, which is close to the 3.5 dB deficit created by the greater distance to Neptune (30 AU) vs Uranus (20 AU). Just adding up apertures, as above, neglects the fact that southern aperture is more effective than northern for Voyager's southern declination, etc., but it pretty well characterizes the matter.

The proposed approach to Voyager Neptune support is currently considered to be viable and cost effective. There is minimal NASA investment in facilities and equipment that do not have evident application for NASA mission support after the Voyager use.

Use of the non-NASA facilities, which is a critical and fundamental part of the Neptune support proposal, is in the planning stages. No serious obstacles to the approach have been encountered thus far.

C. Applications of Proposed DSN Antenna Capability Increase to Other Missions

In addition to the application to the Voyager Neptune encounter, the 64-m upgrades and the Madrid new 34-m antenna are vital to providing the quantity of support passes projected to be required by high priority missions of the late 80s and early 90s.

At that time period, the increment of performance provided by the 64-m upgrade is particularly useful. It relieves projected network overload by reducing the need for multi-antenna arrayed support of individual high priority spacecraft and provides much improved support of the three very distant S/C (VGRs 1,2; PIO 10), which are cruising out of the solar system.

D. Considerations of Use of Higher Frequencies

The DSN is evaluating the potential benefits to communications and precision radio metric data from the use of Ka-band frequencies. There are frequency allocations at Ka-band for deep space research support.

The potential benefits of use of that band are attractive. In addition to providing more protected radio spectrum space, they include: reduced spacecraft communication system cost; radio metric data instrumentation finally free of propagation media charged particle induced errors; a significant increment

in telemetry capability for future (circa year 2000) high-yield very distant missions.

To support development and demonstration of Ka-band system technology for deep space communications use, a new antenna and planetary radar capability are proposed for the Goldstone Venus Station. It is expected that it will be practical in the future to adapt the new 34-m high efficiency antennas and the upgraded 64-m antennas to use at Ka-band frequencies.

The history and projections of DSN antenna development described in this report are summarized in the time line of Fig. 3.

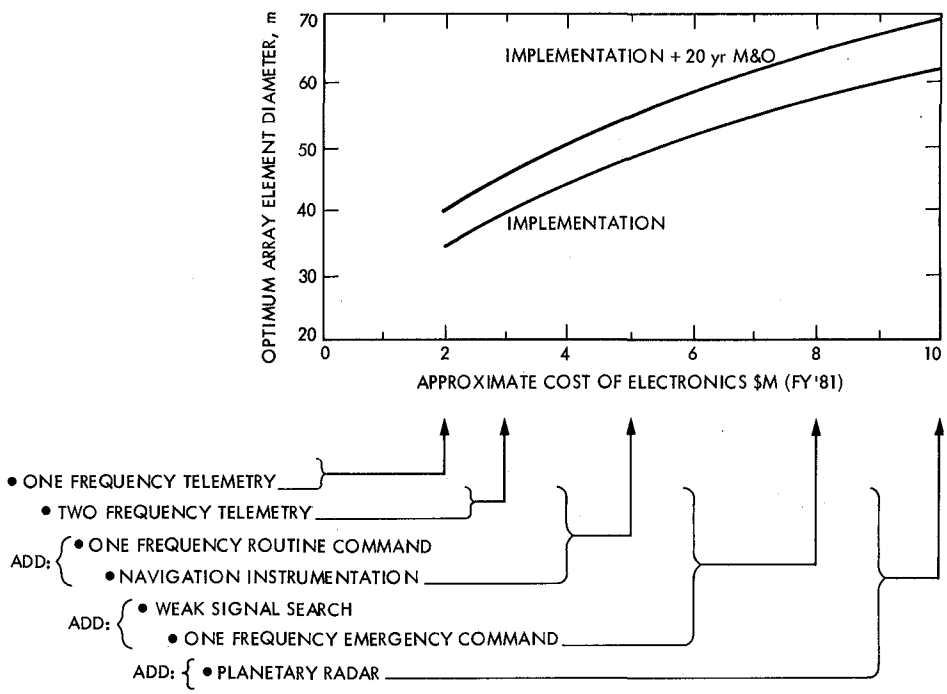


Fig. 1. Optimum element diameter vs technical functional capability of station

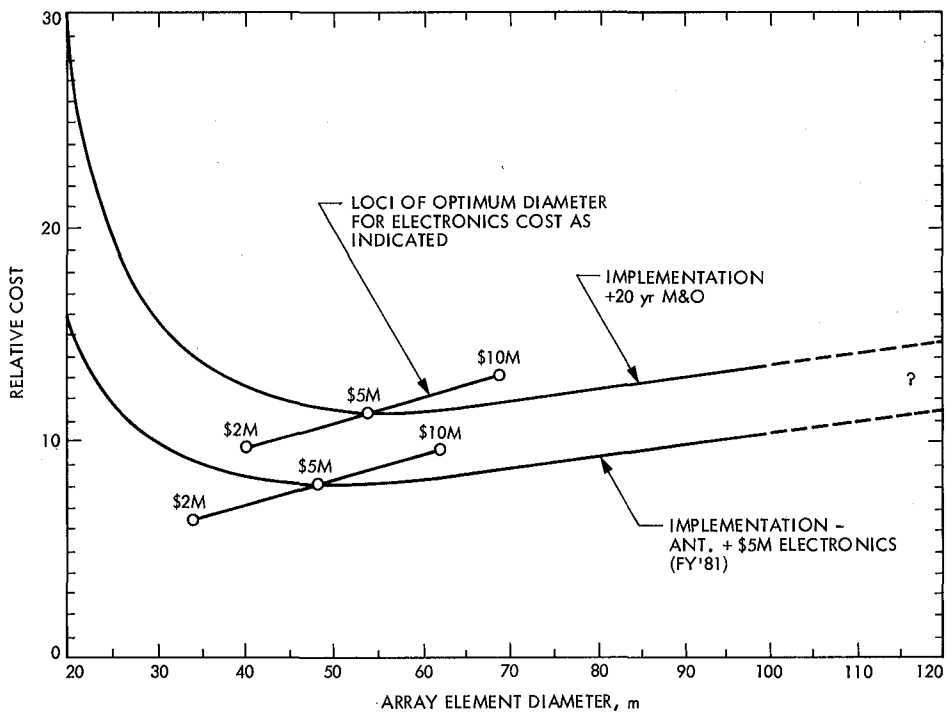


Fig. 2. Sample curve of cost vs diameter to add performance capability to DSN

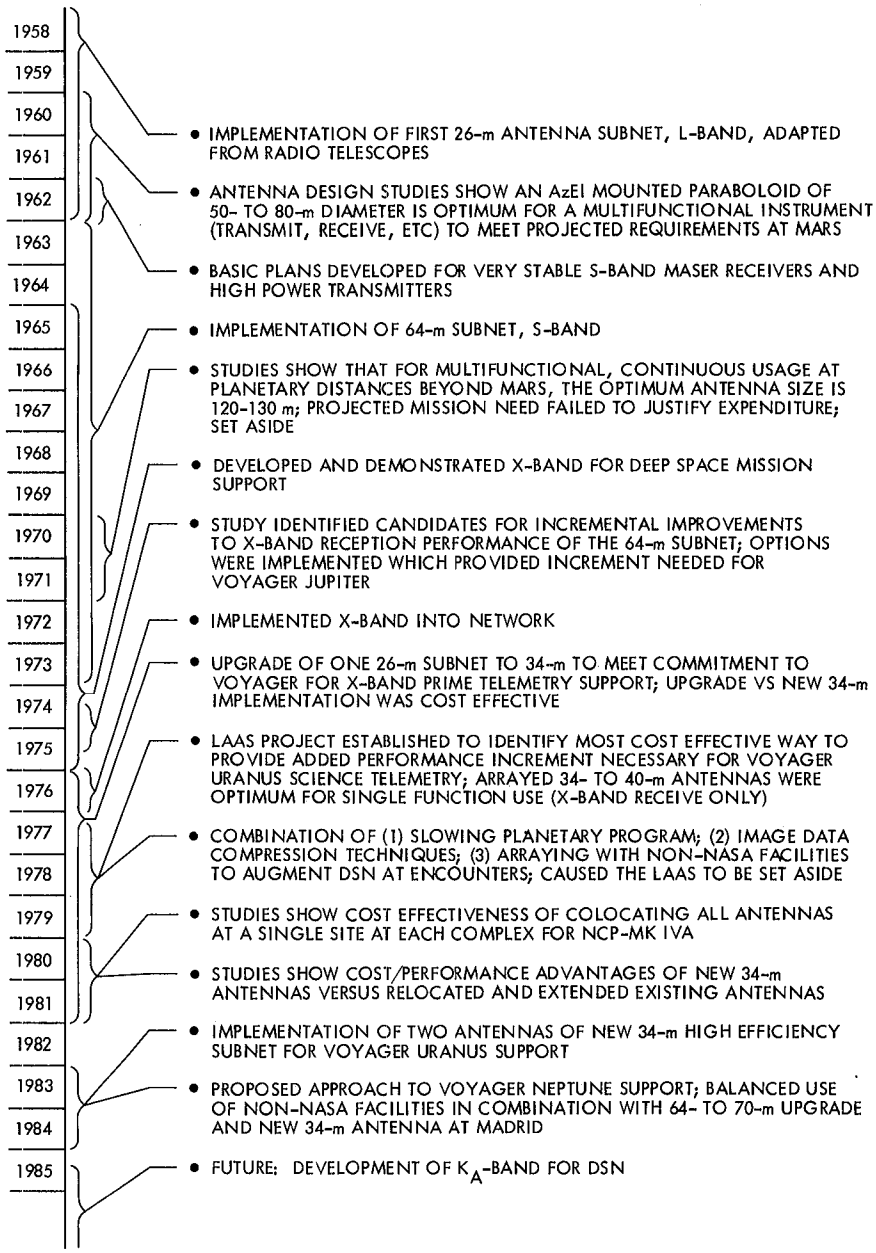


Fig. 3. Summary time line of DSN antenna implementations

Enhanced Radio Frequency Carrier Margin Improvement for an Array of Receiving Systems with Unequal Predetection Signal-to-Noise Ratios

M. H. Brockman
Telecommunications Science and Engineering Division

Enhanced radio frequency carrier margin improvement for arrayed receiving systems for coherent reception provides a significant carrier sensitivity improvement with a resultant decrease in telemetry radio loss due to reduction in rms phase noise at low carrier margins. In addition, a significant increase in doppler rate capability is realized relative to that obtained by switching to a narrower tracking loop bandwidth to obtain the same carrier sensitivity improvement. This report examines this situation for arrayed receiving systems with unequal apertures and statistically independent predetection noise.

I. Introduction

This report presents a technique for providing enhanced radio frequency carrier margin improvement for coherent carrier reception and demodulation for an array of receiving systems with unequal antenna apertures. A later report will present enhanced RF carrier margin improvement for an array with equal antenna apertures. In this initial report, the various components of operating system noise temperature are treated as statistically independent among the receiving systems of the array. Later reports will examine the effect of partial coherence in system noise temperature on enhanced carrier margin improvement for an array of receiving systems with unequal antenna apertures and for an array with equal antenna apertures. Performance presented herein for enhanced RF carrier margin improvement is representative of 34-m diameter antenna receiving systems arrayed with a 64-m diameter antenna receiving system which also has transmit capability.

II. Receiver Configuration

Figure 1 illustrates a configuration considered herein which is the same as that shown in Fig. 1 (Ref. 1) except that the noise bandwidth of predetection IF filters F_{A_2} through F_{A_N} in receiving systems 2 through N are k times the noise bandwidth of predetection IF filter F_{A_1} in receiving system 1 where k is larger than 1. Figure 1 and a modification of Fig. 1 (so that much larger antenna separations for the array can be handled conveniently) were presented in Ref. 2 with a discussion of predetection noise resulting from operating equivalent system noise temperature (T_{o_p}). Figure 2 illustrates a second configuration that provides additional filtering of receiving system 2 (through N) local oscillator phase noise which couples into receiving system 1 via the RF carrier summing junction shown. As a consequence, Fig. 2 generally provides an increase in enhanced carrier margin improvement relative to that obtained utilizing Fig. 1.

III. Predetection Signal-to-Noise Ratio and RF Carrier Tracking Loop Phase Noise

For the situation considered herein, the improvement in predetection signal-to-noise ratio that was developed in Ref. 1, (Eq. 3) for an array of two receiving systems and (Eq. 5) for an array of N receiving systems, represents the improvement in predection carrier power to noise spectral density ratio for the RF carrier phase tracking loop in receiving system 1. These expressions in Ref. 1 apply for the case where a portion of the predetection noise is coherent (ϵ_1 and ϵ_2 in receiving systems 1 and 2, respectively). In this initial report, however, the various components of system noise are treated as statistically independent so that ϵ_1 and ϵ_2 are zero and $1 - \epsilon_1 = 1 - \epsilon_2 = 1$. The improvement in predetection carrier power to noise spectral density ratio in receiving system 1 for two receiving systems arrayed (n_2) then becomes

$$n_2 = \frac{(1 + \beta_2 \gamma_2)^2}{\left[1 + \frac{N_{o2}}{N_{o1}} \beta_2^2\right]} \tag{1}$$

where β_2 is the voltage coupling of receiving system 2 relative to receiving system 1 at the summing junction and γ_2^2 is the carrier power-to-noise spectral density ratio of receiving system 2 relative to receiving system 1. The term N_{o1} is the one-sided noise spectral density of receiving system 1 related to T_{op_1} and N_{o2} is the noise spectral density of receiving system 2 related to T_{op_2}. For N receiving systems arrayed, the improvement in predetection carrier power to noise spectral density ratio (n_N) in receiving system 1 then becomes

$$n_N = \frac{(1 + \beta_2 \gamma_2 + \cdots \beta_N \gamma_N)^2}{\left[1 + \frac{N_{o2}}{N_{o1}} \beta_2^2 + \cdots + \frac{N_{oN}}{N_{o1}} \beta_N^2\right]} \tag{2}$$

Reference 1 provides detailed performance information on improvement in predetection signal-to-noise ratio.

Consider the RF carrier phase tracking loop in receiving system 1 for the situation where the predetection IF filters F_{A_2} through F_{A_N} in receiving systems 2 through N have k times the noise bandwidth of predetection IF filter F_{A_1} (Fig. 1). The carrier tracking loop is a second-order phase tracking loop which includes a bandpass limiter and a sinusoidal phase detector. With receiving system 1 only connected to the summing junction, the RF carrier signal-to-noise power ratio at the input to the bandpass limiter is (Ref. 2, Eq. 1)

$$\frac{P_{c1}}{P_{n1}} = \frac{(A_1 \cos m_{pD})^2}{NBW_{F_{A1}} \cdot N_{o1}} \text{ or } \frac{P_{c1}}{NBW_{F_{A1}} \cdot N_{o1}} \tag{3}$$

where m_{pD} is the peak phase modulation index and $NBW_{F_{A1}}$ represents the noise bandwidth of predetection IF filter F_{A1}. The resultant rms phase noise (σ_{ϕ_n}) at the output of the RF carrier tracking loop (i.e., on the first local oscillator) can be expressed as

$$\sigma_{\phi_{n1}} = \underbrace{\left(\frac{\frac{N_{o1}}{2} \cdot 2B_{L1}}{P_{c1}}\right)^{1/2}}_{\text{(a)}} \cdot \underbrace{\left[\frac{1 + \frac{P_{c1}}{NBW_{F_{A1}} N_{o1}}}{0.862 + \frac{P_{c1}}{NBW_{F_{A1}} N_{o1}}}\right]^{1/2}}_{\text{(b)}}$$

$$\cdot \underbrace{\left[\frac{N_{o1} B_{L1}}{P_{c1}} \cdot \frac{\exp\left(\frac{N_{o1} B_{L1}}{P_{c1}}\right)}{\sinh\left(\frac{N_{o1} B_{L1}}{P_{c1}}\right)}\right]^{1/2}}_{\text{(c)}} \text{ rad, rms} \tag{4a}$$

for $\sigma_{\phi_n} \leqslant 1$ rad rms. Collecting terms, Eq. (4a) becomes (Ref. 2, Eq. 2)

$$\sigma_{\phi_{n1}} = \frac{\frac{N_{o1}}{2} \cdot 2B_{L1}}{P_{c1}} \cdot \left[\frac{1 + \frac{P_{c1}}{NBW_{F_{A1}} \cdot N_{o1}}}{0.862 + \frac{P_{c1}}{NBW_{F_{A1}} \cdot N_{o1}}} \cdot \frac{\exp\left(\frac{N_{o1} B_{L1}}{P_{c1}}\right)}{\sinh\left(\frac{N_{o1} B_{L1}}{P_{c1}}\right)}\right]^{1/2} \text{ rad, rms} \tag{4b}$$

The two-sided closed-loop noise bandwidth can be expressed as

$$2B_{L1} = \frac{2B_{Lo1}}{r_o + 1}\left(1 + r_o \frac{\alpha_1}{\alpha_{o1}}\right) \tag{5}$$

where $r_o = 2$ by design at design point (0.707 damping) and $2B_{Lo1}$ is the design point (threshold) two-sided closed-loop

noise bandwidth in receiving system 1. The term α_1 is the limiter suppression factor resulting from the noise-to-carrier power ratio due to $NBW_{F_{A1}}$ at the input to the bandpass limiter. The suppression factor α_1 has a value of α_{o1} at design point (threshold). At threshold, the predetection carrier-to-noise power ratio in a noise bandwidth equal to $2B_{L_{o1}}$ is unity (i.e., $P_c/(2B_{L_o} \cdot N_o) = 1$). Note that the term designated (b) in Eq. (4a) is determined by the carrier-to-noise power ratio at the input to the bandpass limiter.

With receiving systems 1 and 2 connected to the summing junction, the RF carrier signal-to-noise power ratio at the input to the bandpass limiter is (Ref. 1, Eq. 1 with ϵ_1 and $\epsilon_2 = 0$)

$$\frac{P_{c1_{\Sigma 1,2}}}{P_{n1_{\Sigma 1,2}}} = \frac{(A_1 \cos m_{pd} + \beta_2 A_2 \cos m_{pd})^2}{\left[NBW_{F_{A1}} \cdot N_{o1} + \beta_2^2 MBW_{F_{A2}} \cdot N_{o2}\right]} \tag{6}$$

where $NBW_{F_{A2}} = k_2 \cdot NBW_{F_{A1}}$. Equation (6) can be rewritten as

$$\frac{P_{c1_{\Sigma 1,2}}}{P_{n1_{\Sigma 1,2}}} = \frac{P_{c1}}{NBW_{F_{A1}} \cdot N_{o1}} \cdot \frac{(1 + \beta_2 \gamma_2)^2}{\left[1 + \frac{N_{o2}}{N_{o1}} \beta_2^2 k_2\right]} \tag{7}$$

The change in RF carrier signal-to-noise power ratio at the input to the bandpass limiter in receiving system 1 is then (comparison of Eqs. [7] and [3]):

$$\Delta_2 = \frac{(1 + \beta_2 \gamma_2)^2}{\left[1 + \frac{N_{o2}}{N_{o1}} \beta_2^2 k_2\right]} \tag{8}$$

The limiter suppression factor due to the change in noise-to-carrier power ratio becomes $\alpha_{1_{\Delta 2}}$ which provides a two-sided closed-loop noise bandwidth:

$$2B_{L1_{\Delta 2}} = \frac{2B_{L_{o1}}}{r_o + 1}\left(1 + r_o \frac{\alpha_{1_{\Delta 2}}}{\alpha_{o1}}\right) \tag{9}$$

The resultant rms phase noise at the output of the RF carrier tracking loop (i.e., on the first local oscillator) in receiving system 1 becomes:

$$\sigma_{\phi_{n1\Sigma 1,2}} = \frac{\frac{N_{o1}}{2} \cdot 2B_{L1_{\Delta 2}}}{P_{c1}} \cdot \frac{1}{n_2} \cdot \left[\frac{1 + \frac{P_{c1} \cdot \Delta_2}{NBW_{F_{A1}} \cdot N_{o1}}}{0.862 + \frac{P_{c1} \cdot \Delta_2}{NBW_{F_{A1}} \cdot N_{o1}}} \cdot \frac{\exp\left(\frac{N_{o1} \cdot B_{L1_{\Delta 2}}}{P_{c1} \cdot n_2}\right)}{\sinh\left(\frac{N_{o1} \cdot B_{L1_{\Delta 2}}}{P_{c1} \cdot n_2}\right)}\right]^{1/2} \tag{10}$$

For N receiving systems connected to the summing junction, the RF carrier signal-to-noise power ratio at the input to the bandpass limiter becomes

$$\frac{P_{c1\Sigma 1,\ldots,N}}{P_{n1\Sigma 1,\ldots,N}} = \frac{P_{c1}}{NBW_{F_{A1}} \cdot N_{o1}} \cdot \frac{(1 + \beta_2 \gamma_2 + \cdots + \beta_N \gamma_N)^2}{\left[1 + \frac{N_{o2}}{N_{o1}} \beta_2^2 k_2 + \cdots + \frac{N_{oN}}{N_{o1}} \beta_N^2 k_N\right]} \tag{11}$$

The change in RF carrier signal-to-noise power ratio at the input to the bandpass limiter in receiving system 1 is (comparison of Eqs. [11] and [3]):

$$\Delta_N = \frac{(1 + \beta_2 \gamma_2 + \cdots + \beta_N \gamma_N)^2}{\left[1 + \frac{N_{o2}}{N_{o1}} \beta_2^2 k_2 + \cdots + \frac{N_{oN}}{N_{o1}} \beta_N^2 k_N\right]} \tag{12}$$

The limiter suppression factor becomes $\alpha_{1_{\Delta N}}$ which provides a two-sided closed-loop noise bandwidth

$$2B_{L1_{\Delta N}} = \frac{2B_{L_{o1}}}{r_o + 1}\left(1 + r_o \frac{\alpha_{1_{\Delta N}}}{\alpha_{o1}}\right) \tag{13}$$

For N receiving systems arrayed, the resultant rms phase noise becomes:

$$\sigma_{\phi_{n1\Sigma 1,\ldots,N}} = \frac{\frac{N_{o_N}}{2} \cdot 2B_{L1_{\Delta N}}}{P_{c1}} \cdot \frac{1}{n_N}$$

$$\cdot \left[\frac{1 + \frac{P_{c1} \cdot \Delta_N}{NBW_{F_{A1}} \cdot N_{o1}}}{0.862 + \frac{P_{c1} \cdot \Delta_N}{NBW_{F_{A1}} \cdot N_{o1}}} \cdot \frac{\exp\left(\frac{N_{o1} \cdot B_{L1_{\Delta N}}}{P_{c1} \cdot n_N}\right)}{\sinh\left(\frac{N_{o1} \cdot B_{L1_{\Delta N}}}{P_{c1} \cdot n_N}\right)} \right]^{1/2} \text{rad, rms} \quad (14)$$

Note that the total rms phase noise at the output of the RF carrier tracking loop (i.e., on the first local oscillator) in receiving system 1 for Fig. 1 is

$$\left[\sigma^2_{\phi_{n1\Sigma 1,2}} + \left(\frac{\beta_2 \sigma_{\phi_{n2}}}{1 + \beta_2} \right)^2 \right]^{1/2} \quad (15)$$

for two receiving systems arrayed (Ref. 2). For N receiving systems arrayed, the total rms phase noise for Fig. 1 is

$$\left[\sigma^2_{\phi_{n1\Sigma 1,\ldots,N}} + \left(\frac{\beta_2 \sigma_{\phi_{n2}}}{1 + \beta_2} \right)^2 + \cdots + \left(\frac{\beta_N \sigma_{\phi_{nN}}}{1 + \beta_N} \right)^2 \right]^{1/2} \quad (16)$$

Expressions for rms phase noise in the carrier phase tracking in receiving systems 2 through N were developed in Ref. 2.

In Fig. 2, additional filtering of the output rms phase noise $\sigma_{\phi_{n2}}$ is provided by the local oscillator tracking loop in receiving system 2. Designate the additionally filtered rms phase noise as $\sigma'_{\phi_{n2}}$ which is less than $\sigma_{\phi_{n2}}$ by the square root of the ratio of local oscillator tracking loop noise bandwidth to $2B_{L2}$. Consequently for Fig. 2, $\sigma'_{\phi_{n2}}$ and $\sigma'_{\phi_{nN}}$ are substituted into Eqs. (15) and (16) in place of $\sigma_{\phi_{n2}}$ and $\sigma_{\phi_{nN}}$.

The rms phase noise ($\sigma_{\phi n1}$) for receiving system 1 by itself is determined, using Eq. (4), from threshold to stronger RF carrier levels for a given set of parameters. A given RF carrier level above threshold (carrier margin) results in a specific $\sigma_{\phi_{n1}}$ at the output of RF carrier phase tracking loop in receiving system 1. With receiving system 1 and 2 connected to the summing junction, total rms phase noise at the output of the RF carrier tracking loop in receiving system 1 is determined from Eqs. (10) and (15). The rms phase noise, Eqs. (15) and (10), represents a different RF carrier margin when compared to $\sigma_{\phi_{n1}}$ for receiving system 1 alone. The change in carrier margin represents the enhanced carrier margin improvement in receiving system 1 for two receiving systems arrayed. With N receiving systems arrayed, Eqs. (16) and (14) represent the change in RF carrier margin relative to receiving system 1 alone to provide the enhanced carrier margin improvement in receiving system 1 for N receiving systems arrayed. Note that receiving systems 2 through N have essentially the same RF carrier characteristics and sensitivity as receiving system 1.

IV. Performance

The enhanced RF carrier margin improvement that can be obtained by arraying receiving systems can now be determined for representative sets of design parameters using the preceding development. The following sets of design parameters apply for the performance presented in this report. The sets of design parameters for receiving system 1 are:

Threshold Two-Sided Noise Bandwidth				
$2B_{L_{o1}}$	10.0	30.0	100.0	300.0 Hz
Predetection IF Filter Noise Bandwidth				
$NBW_{F_{A1}}$	2000.0	2000.0	20000.0	20000.0 Hz

while the corresponding sets of parameters for receiving system 2 through N are:

Threshold Two-Sided Noise Bandwidth				
$2B_{L_{o2,\ldots,N}}$	0.3	1.0	3.0	10.0 Hz
Predetection IF Filter Noise Bandwidth				
$NBW_{FA2,\ldots,N}$	$k \cdot NBW_{F_{A1}}$	$k \cdot NBW_{F_{A1}}$	$k \cdot NBW_{F_{A1}}$	$k \cdot NBW_{F_{A1}}$
Local Oscillator Tracking Loop Two-Sided Noise Bandwidth (Fig. 2)				
	1.0	1.0	1.0	1.0 Hz

Using the bandwidth parameters above and the array parameters described below, the enhanced RF carrier margin improvement for receiving system 1 is presented in the following material which is representative of 34-m diameter antenna receiving system(s) arrayed with a 64-m diameter antenna receiving system (system 1).

Figure 3 shows enhanced RF carrier margin improvement for receiving system 1 for array configurations representative

of Figs. 1 and 2 as a function of k (where k is the ratio NBW_{FA2}/NBW_{FA1}) with the summing junction voltage coupling β_2 equal to one and $2B_{L_{o1}}$ = 10.0 Hz. Figure 3 represents an array of two receiving systems with a 64-m diameter antenna (system 1) and a 34-m diameter antenna-listen only (system 2) with γ_2 = 0.61 (-4.3 dB). As described earlier in this report, γ_2^2 is the carrier power-to-noise spectral density ratio of receiving system 2 relative to receiving system 1. For Fig. 3, receiving system 1 by itself prior to arraying has an RF carrier level 10 dB above design point threshold (10 dB RF carrier margin), and a two-sided closed loop noise bandwidth of 24.5 Hz for the RF carrier phase tracking loop at this carrier level. The ratio of noise spectral densities (N_{o2}/N_{o1}) (of receiving system 2 relative to receiving system 1) is 0.925 (18.5 K/20.0 K) for noncoherent predetection noise ($(1 - \epsilon_1) = 1$).

Note that in Fig. 3, the enhanced RF carrier margin improvement for receiving system 1 increases to 7.8 dB (when k becomes 27) for array configurations representative of Figs. 1 or 2. Since γ_2 = 0.61 for system 2, the corresponding enhanced RF carrier margin improvement for receiving system 2 is 12.1 dB (for k = 27) relative to receiving system 2 by itself operating with a $2B_{L_o}$ of 10 Hz. At k = 27 (Fig. 3), the two-sided closed-loop noise bandwidth of the RF carrier tracking loop in receiving system 1 is 10 Hz (design point noise bandwidth).

The information shown in Fig. 3 is rearranged and expanded in Fig. 4 to show enhanced RF carrier margin improvement for receiving system 1 as a function of the RF carrier level in dB above threshold (carrier margin) of receiving system 1 by itself prior to arraying for selected values of k. Enhanced RF carrier margin improvement is shown for values of k equal to 3.5, 9, 16, 27, and 70 for the two receiving systems arrayed. For k = 9 (Fig. 4), the initial RF carrier level for receiving system 1 by itself prior to arraying must be at least 5.5 dB above threshold. At this 5.5 dB level and with the two receiving systems arrayed, the two-sided closed loop noise bandwidth of the RF carrier tracking loop in receiving system 1 is 10 Hz (design point noise bandwidth). Consequently, the doppler and doppler rate characteristics vs. received carrier level of the RF carrier tracking loop for the array (k = 9) are transferred (moved) 5.5 dB to the right relative to receiving system 1 operation by itself. At this 5.5 dB carrier level, the enhanced RF carrier margin improvement (for k = 9) is 5.3 dB for array configurations representative of Figs. 1 and 2. For k = 16, 27, and 70, the initial RF carrier level of receiving system 1 by itself prior to arraying must be at least 8 dB, 10 dB, and 14 dB above threshold respectively. The doppler and doppler rate characteristics vs. received carrier level of the RF carrier tracking loop for the array are transferred (moved) 8 dB, 10 dB and 14 dB respectively to the right for k = 16, 27, and 70 as described above.

Figures 5 and 7 show performance characteristics (which are similar to those in Figs. 3 and 4) for an array of two receiving systems utilizing a 64-m diameter antenna and a 34-m diameter antenna-listen only. For this array, receiving system 1 has a $2B_{L_{o1}}$ of 30 Hz for the RF carrier phase tracking loop. For Fig. 5, receiving system 1 by itself prior to arraying has a 10 dB RF carrier margin and a two-sided closed loop noise bandwidth of 70.2 Hz for the RF carrier phase tracking loop at this carrier level. At k = 36 (Fig. 5), the two-sided closed loop noise bandwidth of the RF carrier tracking loop in receiving system 1 is 30 Hz (design point noise bandwidth) and the enhanced RF carrier margin improvement is 7.1 and 7.2 dB for array configurations representative of Figs. 1 and 2, respectively. Figure 6 shows the effect of varying the voltage coupling β_2 of receiving system 2 relative to receving system 1 at the summing junction on enhanced RF carrier margin improvement for k = 18.

In Fig. 7 ($2B_{L_{o1}}$ = 30 Hz), enhanced RF carrier margin improvement for receiving system 1 is shown for values of k equal to 3.5, 9, 18, 36, and 70. In this case, the initial RF carrier level for receiving system 1 prior to arraying must be at least 5.5, 8, 10, and 14 dB above threshold for k equal to 9, 18, 36, and 70, respectively. As before, the doppler and doppler rate characteristics vs. received carrier level for the array are transferred (moved) a corresponding amount to the right in RF carrier level relative to receiving system 1 by itself prior to arraying.

Figures 8 and 9 show enhanced RF carrier margin improvement characteristics (similar to Figs. 3 and 4) for two receiving systems arrayed (64-m diameter antenna and 34-m diameter-listen only antenna) with receiving system 1 operating with a $2B_{L_{o1}}$ of 100 Hz. Figures 10 and 11 show performance characteristics (similar to Figs. 5 and 7) for the case where $2B_{L_{o1}}$ = 300 Hz. It should be noted that for the sets of design parameters considered in this report, enhanced RF carrier margin improvement for a $2B_{L_{o1}}$ for 100 Hz and 300 Hz are the same as for a $2B_{L_{o1}}$ of 10 Hz and 30 Hz respectively for a configuration representative of Fig. 1.

Consider the case for an array of three receiving systems consisting of a 64-m diameter antenna (system 1), a 34-m diameter antenna-listen only (system 2) and a 34-m diameter antenna with transmit/receive capability (system 3). As described above, γ_2 is 0.61 (-4.3 dB) and N_{o2}/N_{o1} is 0.925 for receiving system 2 relative to receiving system 1. For receiving system 3, γ_3 is 0.53 (-5.5 dB) and N_{o3}/N_{o1} is 1.075 (21.5 K/20.0 K) relative to receiving system 1 for noncoherent predetection noise ($(1 - \epsilon_1) = 1$). Enhanced RF carrier margin

improvement is presented in Figs. 12 and 13 for a $2B_{L_{o1}}$ = 10 Hz; Figs. 14 and 15 for a $2B_{L_{o1}}$ = 30 Hz; Figs. 16 and 17 for a $2B_{L_{o1}}$ = 100 Hz; and Figs. 18 and 19 for a $2B_{L_{o1}}$ = 300 Hz. The performance characteristics shown in these figures for the three-aperture array are similar to those discussed above for the two-aperture array for configurations representative of Figs. 1 and 2.

Initial measurements of enhanced RF carrier margin improvement have been made in the laboratory for two and three receiving systems arrayed. These measurements were made with a predetection filter noise bandwidth of 2200 Hz with a $2B_{L_{o1}}$ of 152 Hz for system 1 and with a $2B_{L_o}$ of 1 Hz with $k = 3.5$, $\gamma = 0.61$ for systems 2 and 3 (and $N_{o2}/N_{o1} = N_{o3}/N_{o1} = 1$). Measured enhanced RF carrier margin improvement was 1.3 dB with $\beta_2 = 0.56$ and $1 - \epsilon_1 = 0.73$ for two receiving systems arrayed (systems 1 and 2). This measurement was made with an initial RF carrier margin of 14.6 dB for receiving system 1 (prior to arraying). Predicted performance is 1.2 dB for $1 - \epsilon_1 = 0.73$ and 1.65 dB for $1 - \epsilon_1 = 1.0$ at this RF carrier margin for two systems arrayed. Measured enhanced RF carrier margin improvement was 2.2 dB with $\beta_2 = \beta_3 = 0.69$ and $1 - \epsilon_1 = 0.73$ for three receiving systems arrayed. This measurement was made with an initial RF carrier margin of 14 dB for receiving system 1 (prior to arraying). Predicted performance is 2.1 dB for $1 - \epsilon_1 = 0.73$ and 2.7 dB for $1 - \epsilon_1 = 1.0$ for three systems arrayed. Description of the laboratory measurement technique is presented in Ref. 1 (Section III).

The enhanced radio frequency carrier margin improvement for an array of receiving systems presented in this report results from a significant reduction in rms phase noise on the first local oscillator (the output of the RF carrier tracking loop) at a given input carrier level. This raises the point of switching operation to the next narrower bandwidth in the RF carrier tracking loop to accomplish essentially the same reduction in rms phase noise as with enhanced RF carrier margin improvement. Consider the following situation.

Receiving system 1 (with a 64-m diameter antenna) operating with a threshold two-sided noise bandwidth ($2B_{L_{o1}}$) of 30 Hz has an initial carrier margin (carrier level above design-point threshold) of 10 dB by itself prior to arraying. When arrayed with a second receiving system (with a 34-m diameter antenna) with $k = 9$, the improvement in 64-m receiver carrier margin is 4.8 dB for an array configuration representative of Fig. 1 (see Fig. 7). The resultant receiver array enhanced carrier margin is 10 + 4.8 or 14.8 dB. Receiving system 1 operating with a $2B_{L_o}$ of 10 Hz would have a carrier margin of 14.77 dB since the ratio of 30/10 represents 4.77 dB. Comparison of the doppler rate capability for a given phase error (e.g., 10 deg) due to doppler rate shows that the enhanced carrier margin array ($2B_{L_{o1}}$ = 30 Hz) has 2.9 times the doppler rate capability of receiving system 1 with $2B_{L_{o1}}$ = 10 Hz. This ratio increases to 5.4 at strong signals. A similar comparison for $k = 9$ and $2B_{L_{o1}}$ equal to 100 Hz (see Fig. 9) for the enhanced carrier margin array as compared to receiving system 1 operating with a $2B_{L_{o1}}$ of 30 Hz provides a 3.9 times doppler rate capability. In this case, the ratio increases to 18.6 at strong signals. This last doppler rate comparison also applies for $2B_{L_{o1}}$ equal to 10 Hz for the enhanced carrier margin array ($K = 9$) as compared to receiving system 1 operating with a $2B_{L_{o1}}$ of 3 Hz.

V. Discussion

Enhanced RF carrier margin improvement for arrayed receiving systems is shown for values of k up to 70 in Section IV of this report. It should be noted that application of the parameters described herein to receiving equipment presently at the 64-m stations requires some limits on the maximum value of k. A value of k up to 70 can be accommodated for threshold two-sided noise bandwidths ($2B_{L_{o1}}$) of 10 and 30 Hz. However, for $2B_{L_{o1}}$ equal to 100 and 300 Hz, the maximum value of k is limited to 8 or 9.

The additional RF carrier sensitivity realized by enhanced carrier margin improvement provides a performance capability (illustrated below) that should be investigated and verified by further laboratory measurements. As an illustration, the performance in Fig. 4 shows the improvement in the 64-m receiving system carrier margin for an array of two receiving system utilizing a 64-m diameter antenna and a 34-m diameter antenna-listen only. For $k = 9$, this array should acquire the RF carrier when the initial 64-m receiver carrier margin is as low as 5.5 dB. After RF carrier acquisition at this (5.5 dB) level, the effective RF carrier margin of the 64-m receiver (arrayed) would be 10.8 dB (5.5 + 5.3 dB) as determined by rms phase (σ_{ϕ_n}) in the RF carrier phase tracking loop.

The following situation is presented to illustrate an application of enhanced RF carrier margin improvement during the fly-by encounter of a spacecraft with a planet in the solar system. When operating with a modulation index of 72 deg peak, the RF carrier margin for the 64-m diameter antenna receiving system at an elevation angle of 31 deg is 17.7 dB when operation with a $2B_{L_{o1}}$ = 30 Hz (clear dry weather). This is representative for the 64-m antenna at the Goldstone complex with Voyager 2 at Neptune.[1] Increasing the peak

[1] Voyager Telecom Data Base, Design Control Table, $P_T = -138.1$ dBm at Uranus, $T_{op} = 26.7$ K at DSS-14, 31 deg elevation angle, clear dry weather. Signal level is 3.5 dB less at Neptune (private communication with B. D. Madsen, JPL).

modulation index to 80 deg decreases the RF carrier margin by 5 dB and increases the received telemetry signal-to-noise ratio by 0.30 dB. The RF carrier margin is reduced 2 dB at an antenna elevation angle of approximately 17 deg relative to a 31 deg elevation angle.

Table 1 shows the RF carrier margin for the 64-m diameter antenna receiving system by itself ($2B_{L_{o1}}$ = 30 Hz) at antenna elevation angles of approximately 31 and 17 deg for peak modulation indices of 72 and 80 deg. Radio loss is also shown in Table 1 for the above antenna elevation angles and modulation indices for short constraint length (7), rate 1/2 convolutional coded telemetry (Ref. 3, Figs. 14 and 3) for an array consisting of a 64-m diameter antenna and a 34-m diameter-listen only antenna. Radio loss is shown for baseband arraying and for an array with both enhanced RF carrier margin improvement and baseband arraying. Enhanced RF carrier margin improvement is obtained from Fig. 7 (Fig. 1 configuration). Because of the 64-m enhanced carrier margin, the radio loss is less for this case.

The preceeding discussion has assumed that the doppler rate (due to range acceleration), as seen at the DSN complex, produces a small phase error in the RF carrier phase tracking loops of the 64-m and 34-m receiving systems. This is true (small phase error) except during the time period when range acceleration peaks. For a few hours during fly-by of the planet (Neptune), the two-way X-band doppler rate increases to a peak of about 170 Hz/s and then decreases to less than 0.1 this value (A. B. Sergeysky, D. M. Wegener, and R. J. Cesarone, private communication). Consider the baseband array with a 17.7 dB carrier margin for the 64-m antenna receiving system and 17.7 - 4.3 = 13.4 dB carrier margin for the 34-m antenna-listen only receiving system ($2B_{L_{o1}}$ = 30 Hz, 72 deg modulation index and antenna elevation angle approximately 31 deg). The resultant phase errors due to a doppler rate of 170 Hz/s are 13 and 19 deg in the RF carrier phase tracking loop for the 64-m and 34-m antenna receiving systems, respectively. At an antenna elevation angle of approximately 17 deg, the RF carrier tracking loop phase errors become 16 and 23 deg, respectively. Consider next, the array with both enhanced RF carrier margin improvement and base-band arraying operating with a $2B_{L_{o1}}$ = 100 Hz and k = 9. At an antenna elevation angle of approximately 31 deg, the phase error in the RF carrier phase tracking loop is 3 deg (due to 170 Hz/s) and the enhanced RF carrier margin of the 64-m antenna receiver is 17.6 dB (12.5 + 5.1 dB) from Fig. 9 (Fig. 1 configuration). At an antenna elevation angle of approximately 17 deg the phase error is 4 deg and the enhanced RF carrier margin of the 64-m antenna receiver is 17.2 dB (10.5 + 6.7 dB).

References

1. Brockman, M. H., The Effect of Partial Coherence in Receiving System Noise Temperature on Array Gain for Telemetry and Radio Frequency Carrier Reception for Receiving Systems with Unequal Predetection Signal-to-Noise Ratios, *TDA Progress Report 42-72*, pp. 95-117, Jet Propulsion Laboratory, Pasadena, California, February 15, 1983.

2. Brockman, M. H., The Effect of Partial Coherence in Receiving System Noise Temperature on Array Gain for Telemetry and Radio Frequency Carrier Reception for Similar Receiving Systems, *TDA Progress Report 42-66*, pp. 219-235, Jet Propulsion Laboratory, Pasadena, California, December 15, 1981.

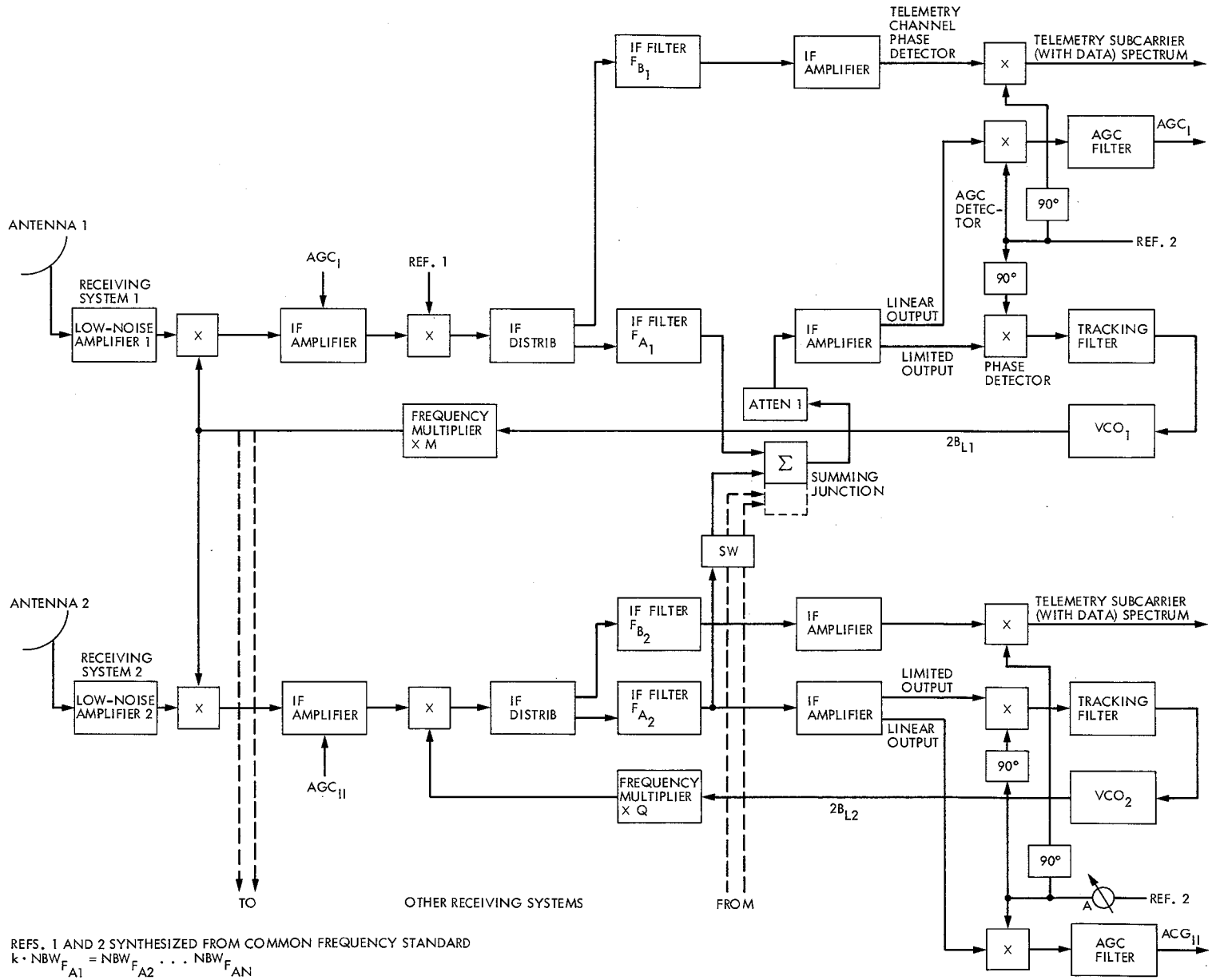
3. Divsalar, D., and Yuen, J. H., Improved Carrier Tracking Performance with Coupled Phase-Locked Loops, *TDA Progress Report 42-66*, pp. 148-171, Jet Propulsion Laboratory, Pasadena, California, December 15, 1981.

Table 1. 64-m and 34-m (listen only) antenna receiver array ($2B_{L_o}$ = 30 Hz)

Antenna Elev. Angle, deg	Modulation Index, deg. peak	Single 64-m Carrier Margin, dB	Baseband Array Radio Loss, dB		Enhanced Carrier Margin and Baseband Array			
			BER 5×10^{-3}	BER 5×10^{-5}	k	64-m Enhanced Carrier Margin, dB	Radio Loss, dB BER 5×10^{-3}	Radio Loss, dB BER 5×10^{-5}
~31	72	17.7	~0.25	~0.5	9	21.	~0.05	~0.1
					18	22.	~0.05	~0.1
	80	12.7	~1.3	~3.5	9	17.1	~0.25	~0.4
					18	18.4	~0.2	~0.25
~17	72	15.7	~0.45	~0.9	9	19.5	~0.1	~0.15
					18	20.6	<0.1	~0.1
	80	10.7	*	*	9	15.4	~0.4	~1.0
					18	16.8	~0.3	~0.4

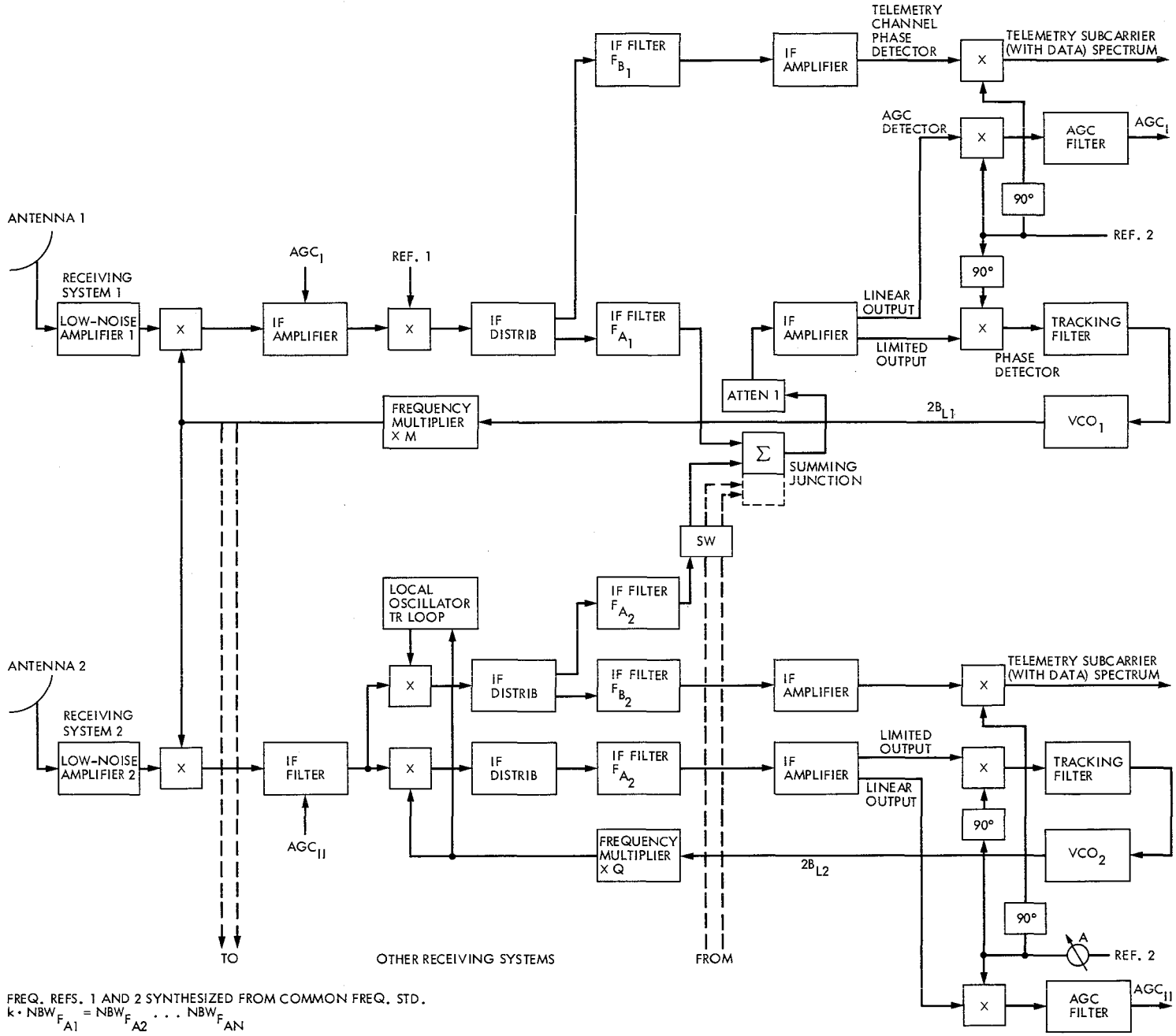
*34-m receiver cycle slipping

Convolutional coded telemetry
Short constraint length (7)
Rate 1/2
$Q = 3$



REFS. 1 AND 2 SYNTHESIZED FROM COMMON FREQUENCY STANDARD
$k \cdot NBW_{F_{A1}} = NBW_{F_{A2}} \ldots NBW_{F_{AN}}$

Fig. 1. Enhanced radio frequency carrier arraying



FREQ. REFS. 1 AND 2 SYNTHESIZED FROM COMMON FREQ. STD.
$k \cdot NBW_{F_{A1}} = NBW_{F_{A2}} \ldots NBW_{F_{AN}}$

Fig. 2. Enhanced radio frequency carrier arraying with additional local oscillator filtering

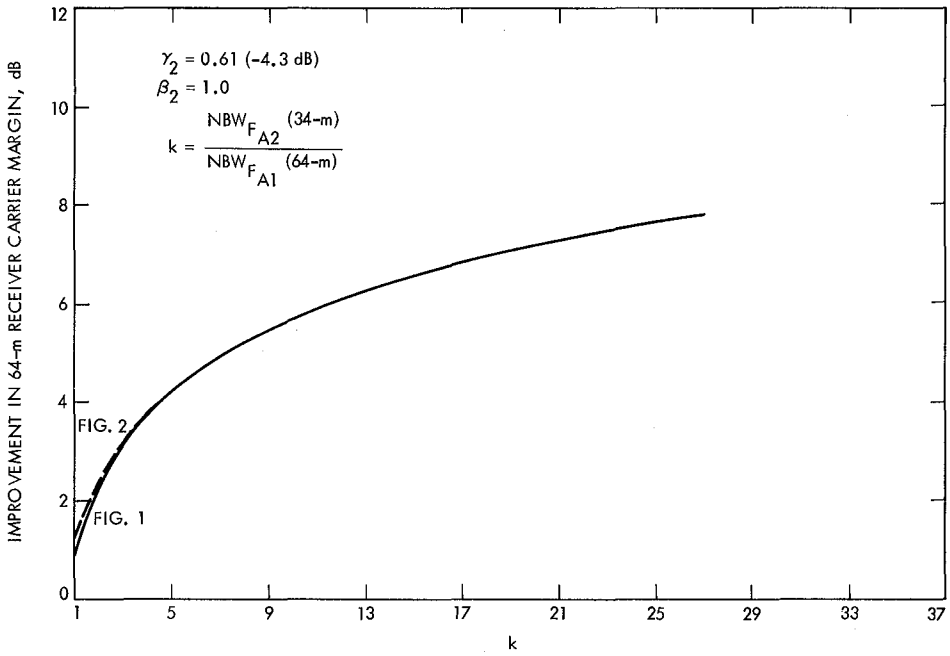


Fig. 3. Enhanced RF carrier margin improvement vs k ($2B_{L/o1}$ = 10. Hz). Two receiving systems are arrayed, the 64-m and 34-m (listen-only) diameter antennas. The 64-m was initially at 10-dB carrier margin.

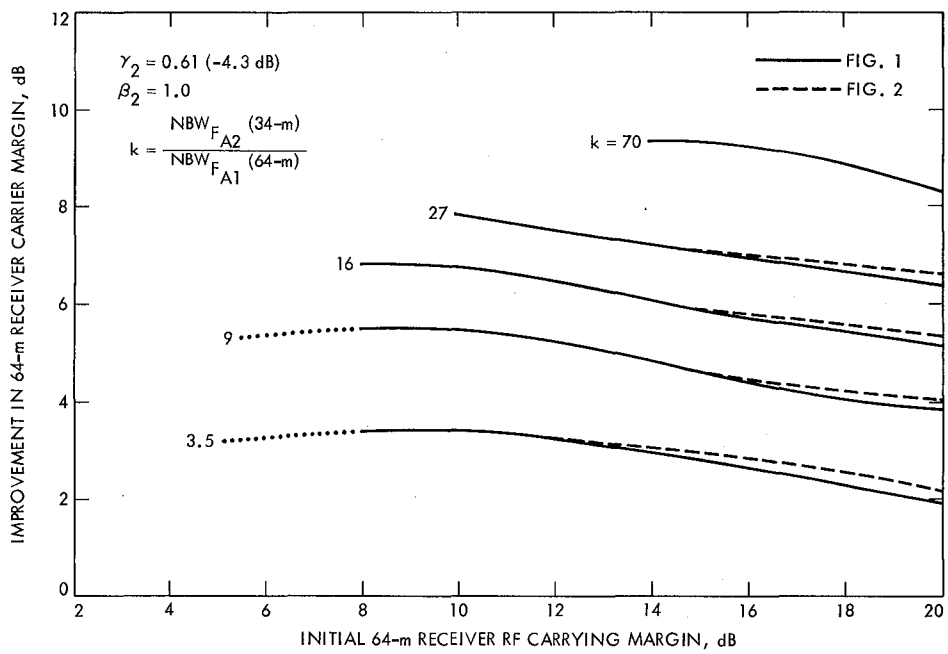


Fig. 4. Enhanced RF carrier margin improvement vs initial 64-m receiver carrier margin ($2B_{L/o1}$ = 10. Hz). Two receiving systems are arrayed, the 64-m and 34-m (listen-only) diameter antennas.

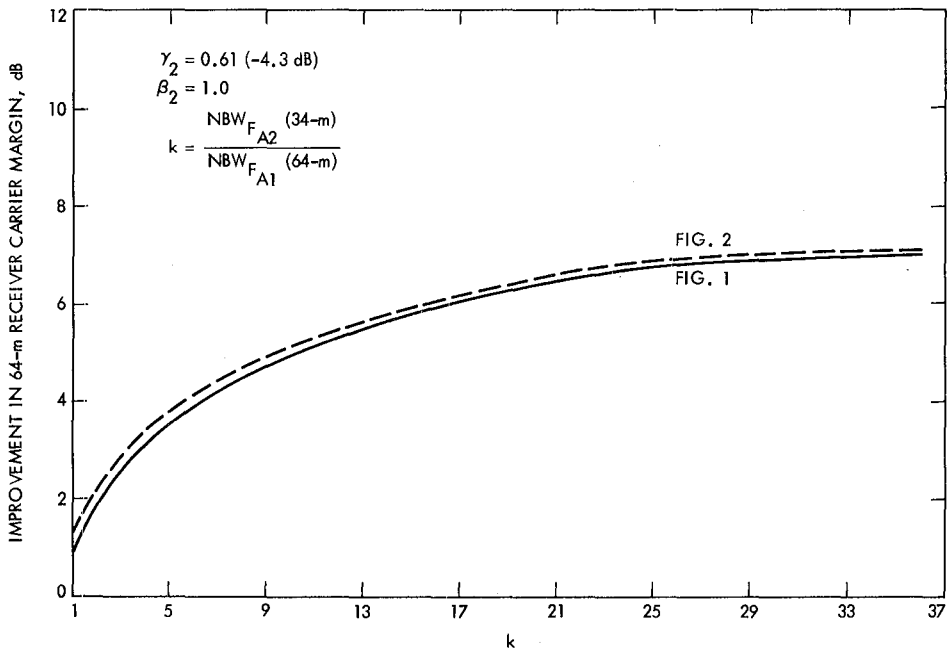


Fig. 5. Enhanced RF carrier margin improvement vs k ($2B_{L/o1}$ = 30. Hz). Two receiving systems are arrayed, the 64-m and 34-m (listen-only) diameter antennas. The 64-m was initially at 10-dB carrier margin.

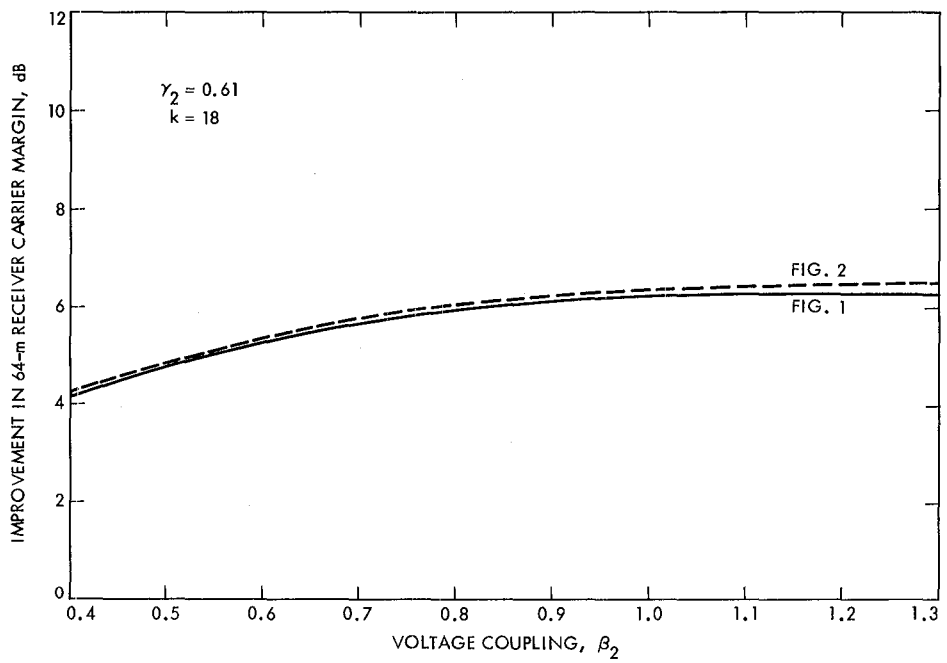


Fig. 6. Effect of summing junction voltage coupling on enhanced RF carrier margin improvement. Two receiving systems are arrayed, the 64-m and 34-m (listen-only) diameter antennas. The 64-m was initially at 10-dB carrier margin ($2B_{L/o1}$ = 30. Hz).

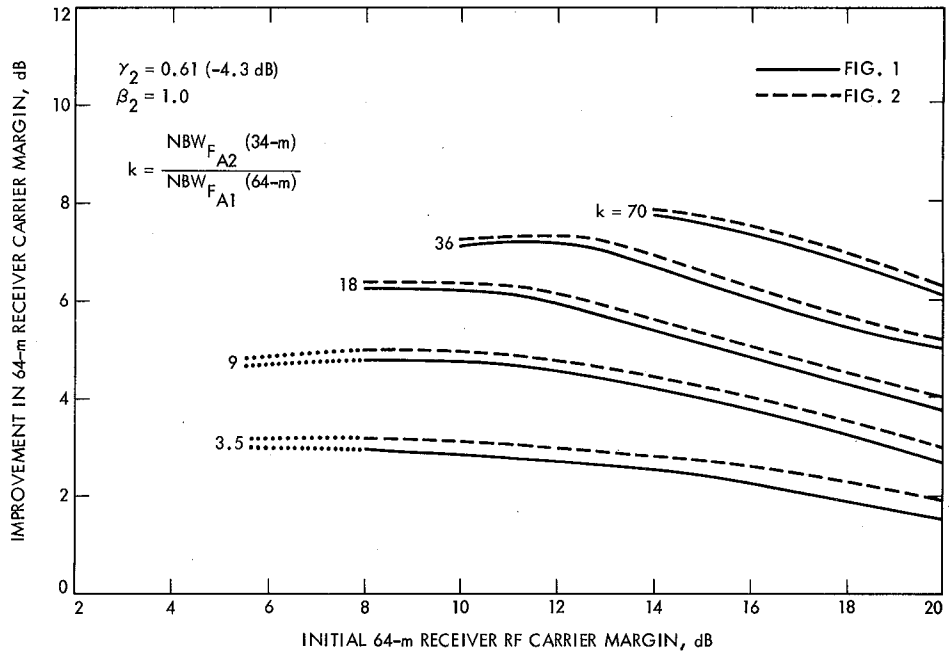


Fig. 7. Enhanced RF carrier margin improvement vs initial 64-m receiver RF carrier margin ($2B_{L/o1}$ = 30. Hz). Two receiving systems are arrayed, the 64-m and 34-m (listen-only) diameter antennas.

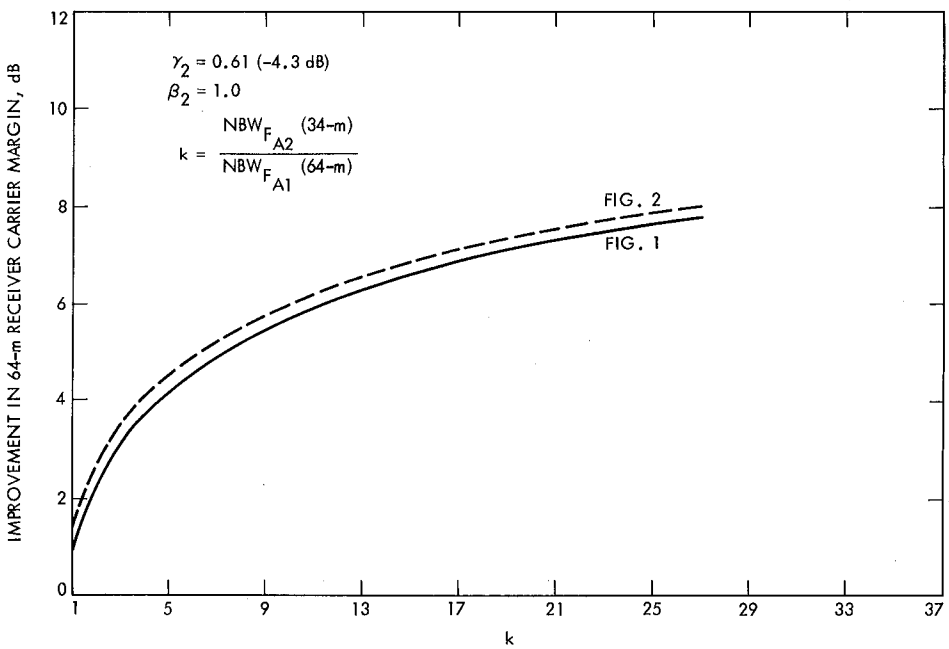


Fig. 8. Enhanced RF carrier margin improvement vs *k* ($2B_{L/o1}$ = 100. Hz). Two receiving systems are arrayed, the 64-m and 34-m (listen-only) diameter antennas. The 64-m was initially at 10-dB carrier margin.

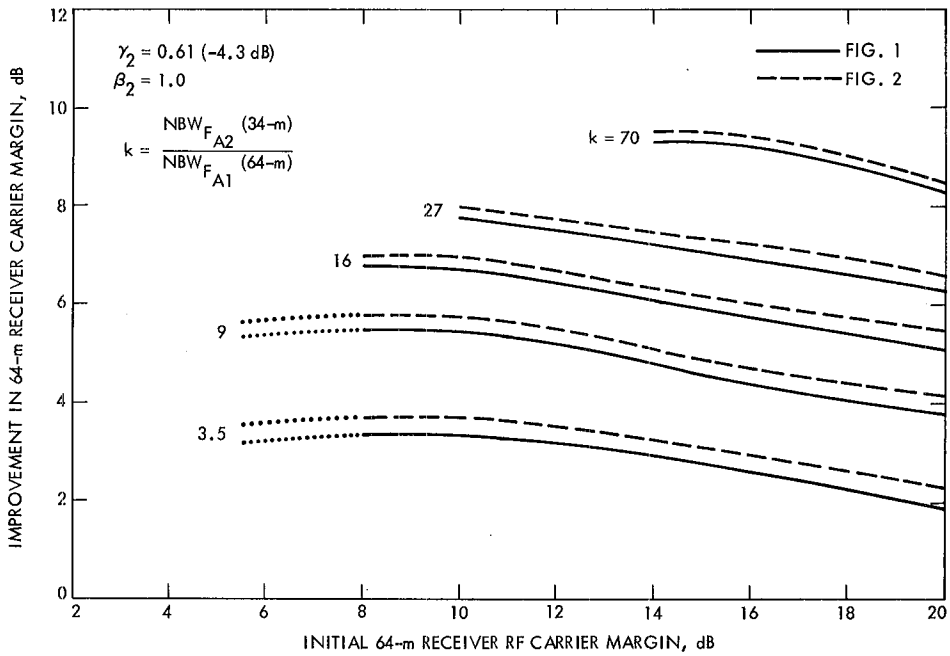


Fig. 9. Enhanced RF carrier margin improvement vs initial 64-m receiver RF carrier margin ($2B_{L/o1}$ = 100. Hz). Two receiving systems are arrayed, the 64-m and 34-m (listen-only) diameter antennas.

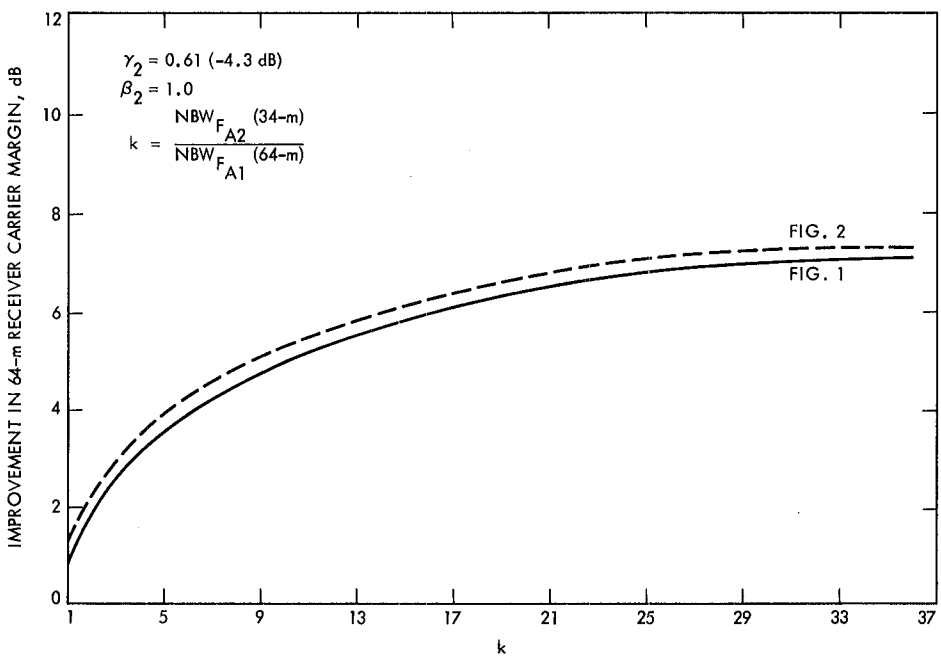


Fig. 10. Enhanced RF carrier margin improvement vs k ($2B_{L/o1}$ = 300. Hz). Two receiving systems are arrayed, the 64-m and 34-m (listen-only) diameter antennas. The 64-m was initially at 10-dB carrier margin.

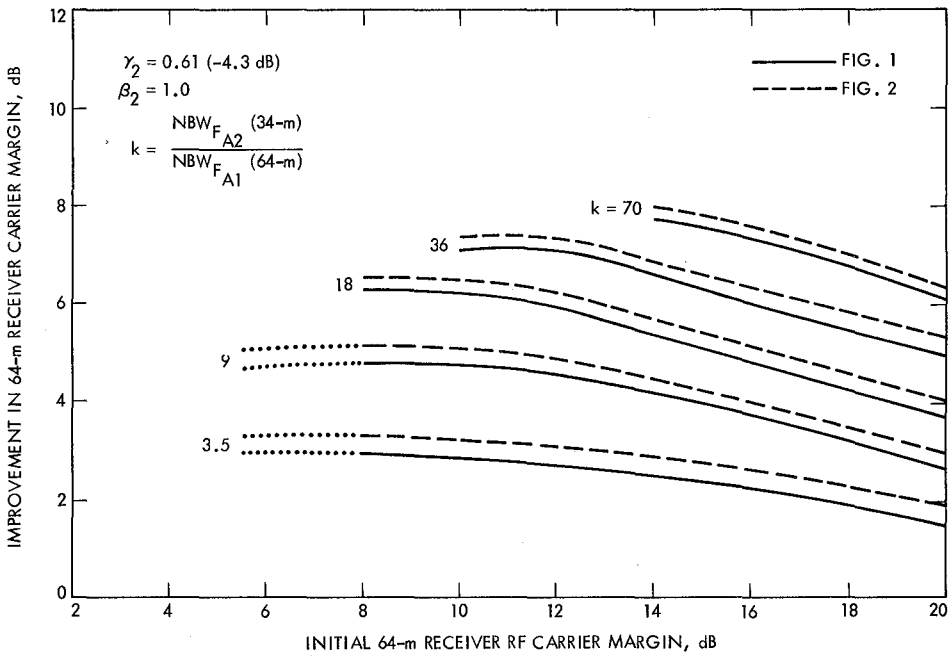


Fig. 11. Enhanced RF carrier margin improvement vs initial 64-m receiver RF carrier margin ($2B_{L/o1}$ = 300. Hz). Two receiving systems are arrayed, the 64-m and 34-m (listen-only) diameter antennas.

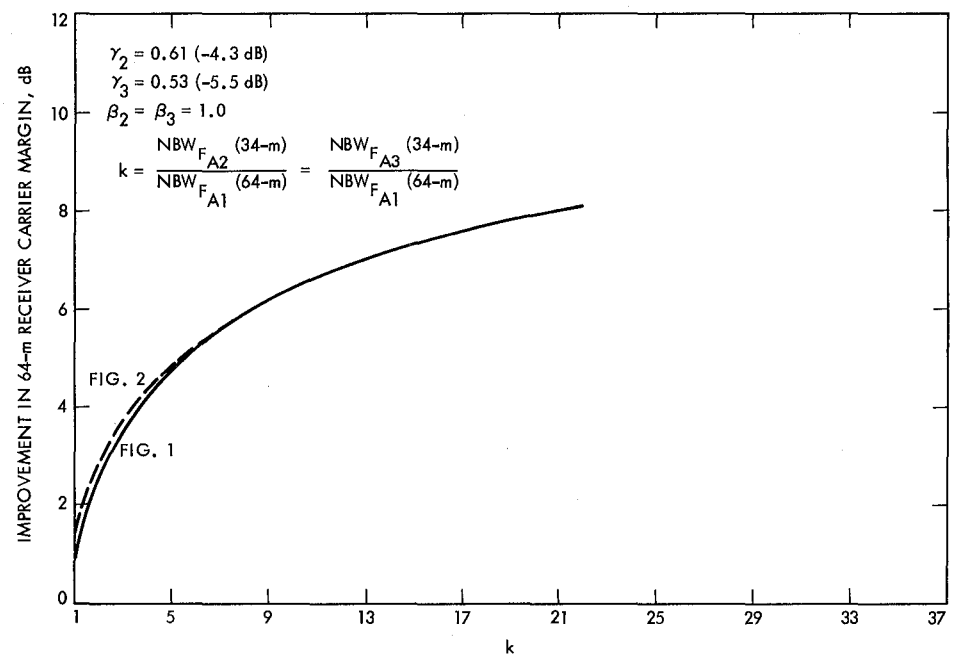


Fig. 12. Enhanced RF carrier margin improvement vs k ($2B_{L/o1}$ = 10. Hz). Three receiving systems are arrayed, the 64-m, 34-m (listen-only), and 34-m (transmit/receive) diameter antennas. The 64-m was initially at 10-dB carrier margin.

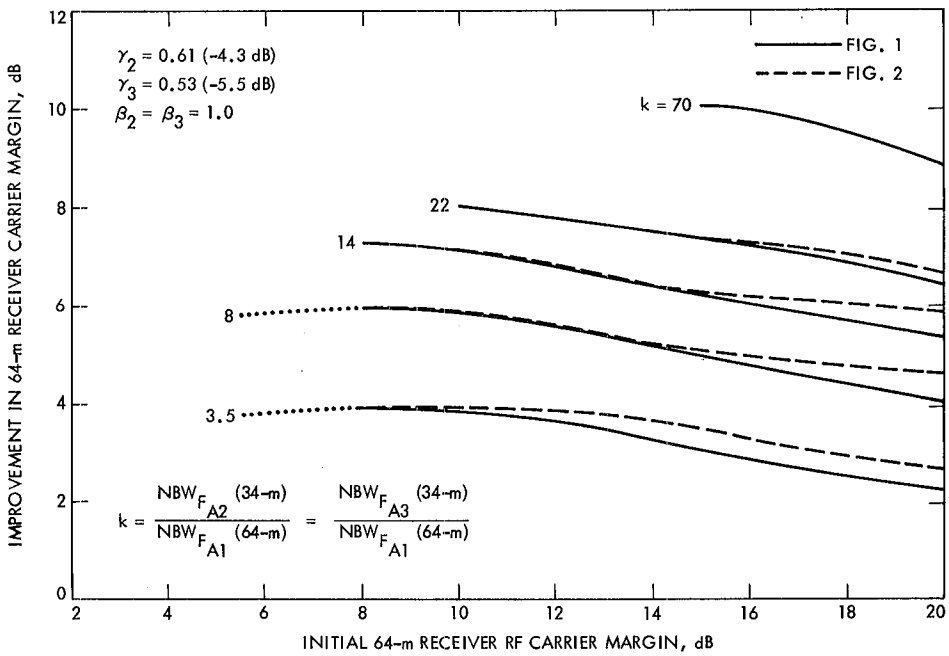


Fig. 13. Enhanced RF carrier margin improvement vs initial 64-m receiver RF carrier margin ($2B_{L/o1}$ = 10. Hz). Three receiving systems are arrayed, the 64-m, 34-m (listen-only), and 34-m (transmit/receive) diameter antennas.

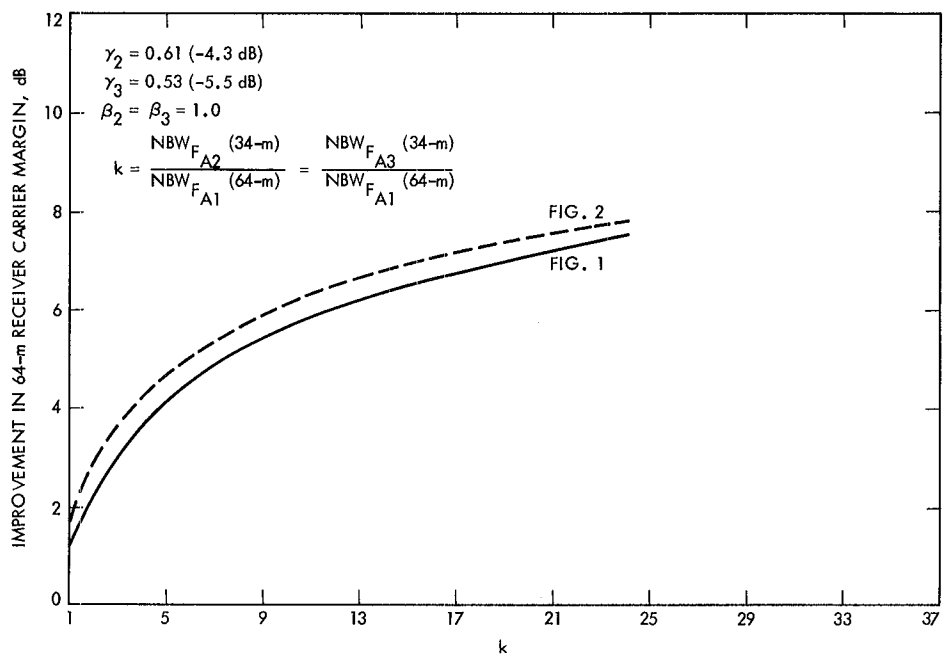


Fig. 14. Enhanced RF carrier margin improvement vs *k* ($2B_{L/o1}$ = 30. Hz). Three receiving systems are arrayed, the 64-m, 34-m (listen-only), and 34-m (transmit/receive) diameter antennas. The 64-m was initially at 10-dB carrier margin.

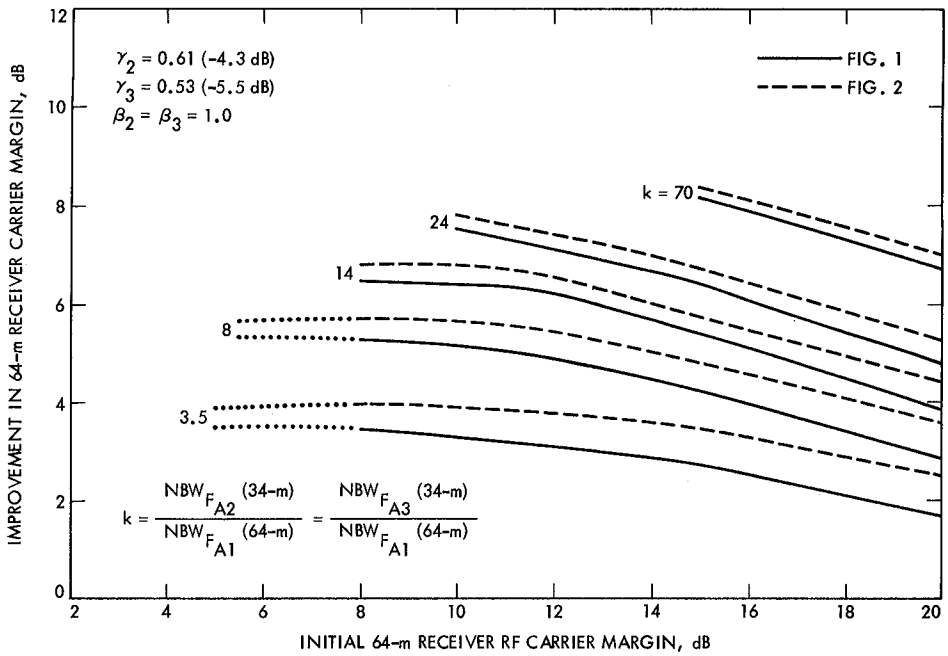


Fig. 15. Enhanced RF carrier margin improvement vs initial 64-m receiver RF carrier margin ($2B_{L/o1}$ = 30. Hz). Three receiving systems are arrayed, the 64-m, 34-m (listen-only), and 34-m (transmit/receive) diameter antennas.

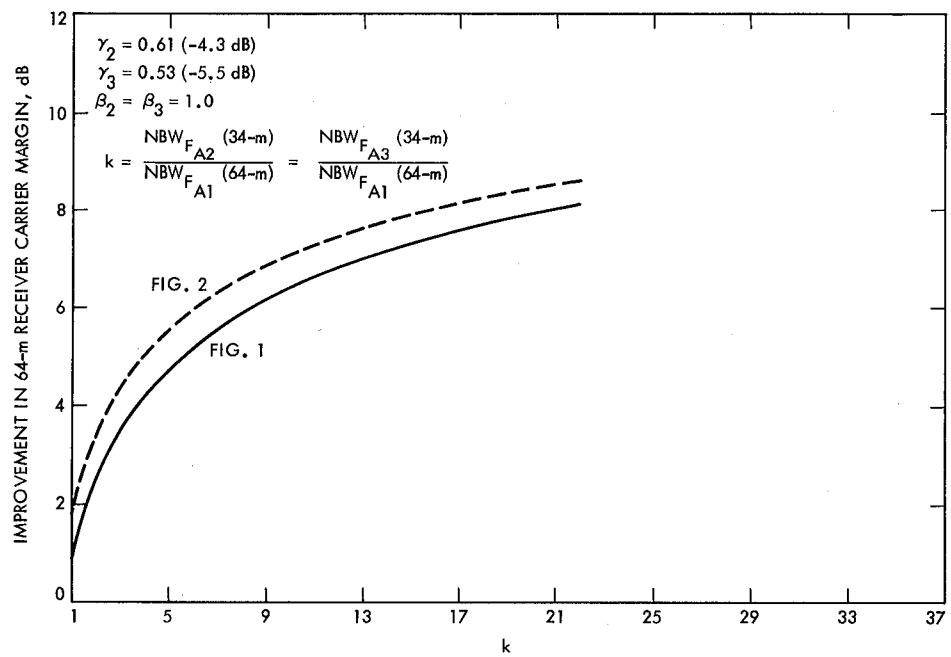


Fig. 16. Enhanced RF carrier margin improvement vs *k* ($2B_{L/o1}$ = 100. Hz). Three receiving systems are arrayed, the 64-m, 34-m (listen-only), and 34-m (transmit/receive) diameter antennas. The 64-m was initially at 10-dB carrier margin.

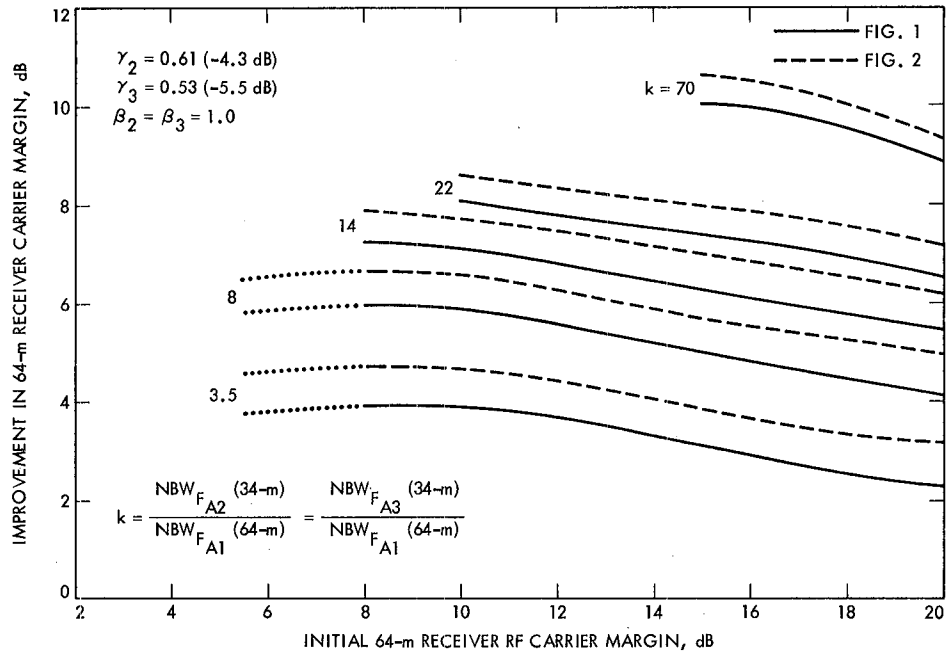


Fig. 17. Enhanced RF carrier margin improvement vs initial 64-m receiver RF carrier margin ($2B_{L/o1}$ = 100. Hz). Three receiving systems are arrayed, the 64-m, 34-m (listen-only), and 34-m (transmit/receive) diameter antennas.

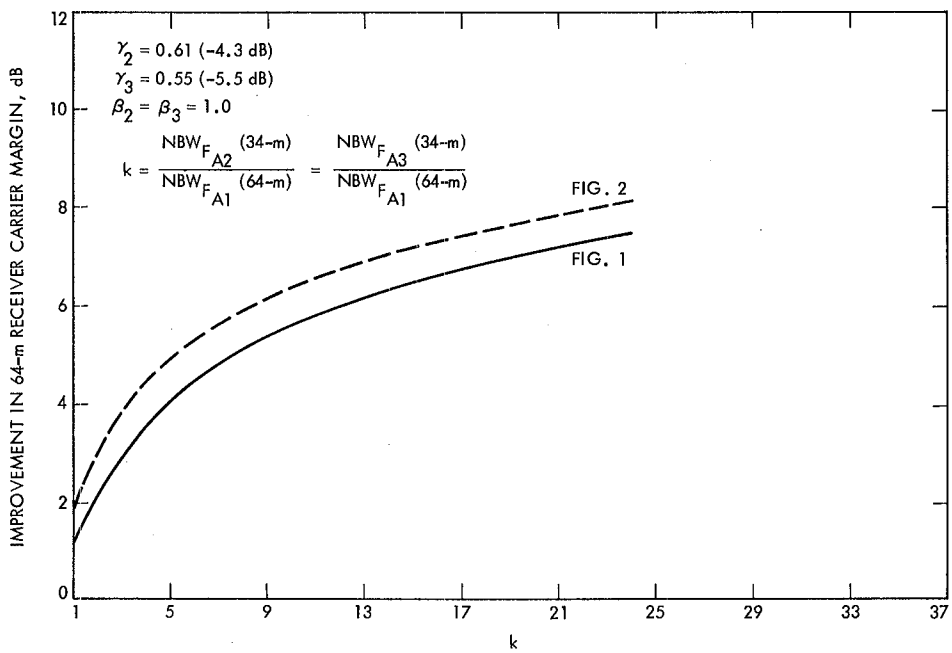


Fig. 18. Enhanced RF carrier margin improvement vs *k* ($2B_{L/o1}$ = 300. Hz). Three receiving systems are arrayed, the 64-m, 34-m (listen-only), and 34-m (transmit/receive) diameter antennas. The 64-m was initially at 10-dB carrier margin.

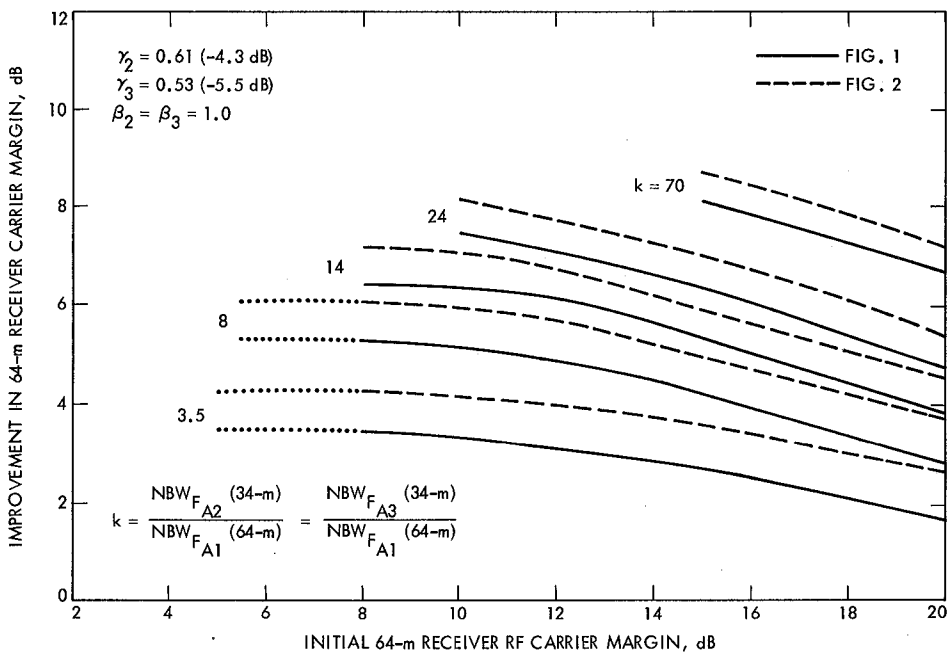


Fig. 19. Enhanced RF carrier margin improvement vs initial 64-m receiver RF carrier margin ($2B_{L/o1}$ = 300. Hz). Three receiving systems are arrayed, the 64-m, 34-m (listen-only), and 34-m (transmit/receive) diameter antennas.

Optimum Cyclic Redundancy Codes for Noisier Channels

P. Merkey
California Institute of Technology, Student

E. C. Posner
Telecommunications and Data Acquisition Office

This article considers binary cyclic redundancy codes for feedback communication over noisy digital links. The standard 16 bit ADCCP (American Data and Computer Communication Protocol) polynomial is designed for digital links which already have a low input bit error probability. For file transfer between personal computers over telephone circuits, the quality of the resulting digital circuit may be much lower. Thus we are led to consider 3 byte (24 bit) and 4 byte (32 bit) polynomials. We find generator polynomials of a certain class which have minimum weight and yet achieve the bound on minimum distance for arbitrary codes. Particular choices for 24 bit and 32 bit redundancies are exhibited: of weight and distance 6 in the 24-bit case; and weight 10 and distance 8 in the 32-bit case. This could be useful as a NASA Standard.

I. Introduction

The ADCCP 16-bit cyclic redundancy check (CRC) polynomial (Ref. 1, Sec. 4.3.1, p. 187) $p(x) = x^{16} + x^{12} + x^5 + 1$ has been specified for error detection and retransmission on digital links using the *X*.25 packet communication standard and for frame error detection in the Packet Telemetry Recommendation of the Consultative Committee on Space Data Systems. This produces a code of minimum distance 4, so it can detect three errors in a block. For this, a block may have any length (conceptually) from 17 (when one information bit is transmitted) to $2^{15} - 1 = 32{,}767$, which is one less than the period of the polynomial. Typical lengths are often in the 500 to 2500 range. Encoding of CRC polynomials is very simple algebraically, for we mrely append 16 check bits to the information bits in the unique way which makes the resulting polynomial divisible by $p(x)$ (Ref. 1, Sec. 4.1.4, p. 163). Error detections is even simpler, for, in the absence of errors, the received polynomial must be a multiple of $p(x)$.

The problem with this is that many file transfers are over voice circuits, not digital links. Thus, the input bit error probability can be much higher than for truly digital links. Furthermore, for some uses such as electronic banking, we may want extremely low output bit error probability, say 10^{-12}. For these reasons, distance-4 codes may not be powerful enough. This will mean that we must add more redundancy, for, as we show below, the largest minimum distance of an $(n, n - 16)$ code, linear or not, is 4 if $362 \leqslant n \leqslant 2^{15} - 1$.

A self-imposed constraint is to make the redundancy a multiple of 8 bits. This is because computers operate on 8-bit bytes. We may want encoding and decoding to be done in the

computers which store and use the data rather than in a special purpose outboard device. Thus we are led to look for good binary CRC codes where the redundancy, instead of being 16, is 24 or 32. This means that we want 24 and 32 degree polynomials with good distance properties. The rest of the paper finds these and shows how close they come to the largest minimum distance for any code of the same parameters, cyclic redundancy or not. Our construction will be based on known results for BCH codes or, rather, for generalized BCH codes. We recommend particular polynomials of degrees 24 and 32 which have as few nonzero terms as possible given some algebraic restrictions. This may make the programming of the encoding and decoding easier or quicker, depending on language and instruction set. Factorizations and other algebraic details of the polynomials are given as well.

II. Largest Minimum Distances

We will want to know how good our codes are, even compared with codes that don't satisfy our self-imposed constraints. By "good" we mean that our codes have close to the largest minimum distance that any code with the same redundancy can have, for a range of code lengths of interest. The following theorem allows us to find this largest minimum distance.

Theorem 1: Suppose we are given an integer $r > 0$. Let t be a non-negative integer such that $2t + 1 \leqslant r$. Let $n_b = n_b(t)$ be the smallest integer such that

$$\sum_{j=0}^{t+1} \binom{n_b}{j} > 2^r$$

Set

$$n_c(0) = \infty$$

and, for $t > 0$, let

$$n_c = n_c(t) = 2^{\lfloor (r-1)/t \rfloor} - 1$$

where $\lfloor x \rfloor$ means the greatest integer at most x. If $n_b < n_c$, then for all n such that $n_b \leqslant n \leqslant n_c$, the largest minimum distance for an $(n, n - r)$ binary code is $2t + 2$.

Remark: If $r < 2t + 1$, there is no hope of getting a minimum distance greater than or equal to $2t + 2$. Merely let exactly one of the $n - r$ information bits be 1. There are only r check bits, so if d is the minimum distance, then $d \leqslant 1 + r < 2t + 2$.

Proof: We first dispose of the case $t = 0$. Note that for $t = 0$, $n_b(0) = 2^r$. We have $n_c(0) = \infty$. The theorem holds since the Hamming Bound (Ref. 2, Chap. 1, Para. 5, p. 19) forces the minimum distance to be less than or equal to 2 for $n > 2^r$. For, if the minimum distance were 3, we would have

$$(1 + n)2^{n-r} \leqslant 2^n$$

$$1 + n \leqslant 2^r$$

$$n < 2^r$$

So the minimum distance for $n \geqslant 2^r$ is 2. We can achieve distance 2 with the code consisting of n-tuples of even parity, an $(n, n - 1)$ code. Because $r \geqslant 1$, we can also achieve distance 2 for an $(n, n - r)$ code. This completes the proof for the case $t = 0$.

Now let t be greater than 0. The Hamming Bound here too implies that the minimum distance d satisfies $d \leqslant t + 2$ if $n \geqslant n_b(t)$. For otherwise, if $d \geqslant 2t + 3$, we would have an $(n, n - r)$ $(t + 1)$-error correcting code. From this, we would find

$$\left[\sum_{j=0}^{t+1} \binom{n}{j}\right] 2^{n-r} \leqslant 2^n$$

$$\sum_{j=0}^{t+1} \binom{n}{j} \leqslant 2^r$$

But the above sum *exceeds* 2^r for $n = n_b$, and all the more for $n \geqslant n_b$. So we certainly can't do better than $d = 2t + 2$ for $m \geqslant n_b$. Can we achieve $d = 2t + 2$ if $n_b \leqslant n \leqslant n_c$? We can, as the following argument shows.

If $n \leqslant n_c(t)$ we will construct a code which has the required $d = 2t + 2$ by taking a subcode of a BCH code (Ref. 2, Chap. 7, Para. 6, p. 201). Let

$$m = \left\lfloor \frac{r - 1}{t} \right\rfloor$$

so that $m \geqslant 2$ and $n_c(t) = 2^m \geqslant 4$. Let α be a primitive element in $GF(2^m)$, and let

$$M^{(1)}(x), M^{(3)}(x), M^{(5)}(x), \ldots, M^{(2t-1)}(x)$$

be the t minimum polynomials for α, α^3, α^5, . . . , α^{2t-1}, respectively. Each of these $M^{(i)}$ has degree $\leqslant m$. From Ref. 2 (Chap. 7, Para. 6, p. 201), we know that the product

$$G_{\mathrm{BCH}}(x) = M^{(1)}\, M^{(3)}, \ldots, M^{(2t-1)}$$

with repeated factors deleted is the generator polynomial for a minimum distance $\geqslant 2t+1$ t-error correcting BCH code. We note that deg $[G_{\mathrm{BCH}}(x)] \leqslant tm \leqslant t(r-1)/t = r-1$.

Now we define a code generator polynomial $g(x)$ of degree $r' \leqslant r$:

$$g(x) = (x+1)\, G_{\mathrm{BCH}}(x)$$

The code generated by $g(x)$ has "natural" length $2^m - 1$. Since it is a subcode of the even-weight codewords in a t-error correcting BCH code, it has minimum distance at least $2t+2$, as desired. Here $r' \leqslant r$, so we can pad in $r' - r$ check bits of 0 if we really want an $(n, n-r)$ code.

We can get a code of any other length less than 2^m by simply setting the appropriate number of high-order terms equal to zero before encoding, and then padding with $r' - r$ zeros as before. As long as we do not shorten the code to a length less than $n_b(t)$, the minimum distance $2t+2$ will be as large as possible by our earlier result. ■

We now look more closely at r, t and the set of integers $n_b(t) \leqslant n \leqslant n_c(t)$. For all we know, $n_c < n_b$ and the theorem is vacuous.

If n is such that

$$2^r < \frac{(n-t)^{t+1}}{(t+1)!}$$

then

$$\sum_{j=0}^{t+1} \binom{n}{j} \geqslant \binom{n}{t+1} \geqslant \frac{(n-t)^{t+1}}{(t+1)!} > 2^r$$

This means that $n_b(t)$ is at most such an n, or

$$n_b(t) < t + 2^{r/(t+1)}[(t+1)!)]^{1/(t+1)}$$

We note that when t is small relative to n, the usual case, this serves as a good approximation to n_b.

Now let us get a *lower* bound on $n_c(t)$. We have

$$n_c(t) = 2^{\lfloor (r-1)/t \rfloor} - 1 > 2^{[(r-1)/t]-1} - 1$$

$$= 2^{(r-t-1)/t} - 1; \quad n_c(t) > 2^{(r-t-1)/t} - 1$$

Together, the upper bound on n_b and lower bound on n_c imply the following:

$$2^{(r-t-1)/t} - 1 - t - 2^{r/(t+1)}\,((t+1)!)^{1/(t+1)} < n_c(t) - n_b(t)$$

Define $f(r, t)$ as the left-hand side of this inequality. Then

$$f(r, t) = 2^{r/(t+1)}\,(2^{[r-(t+1)^2]/[t(t+1)]} - [(t+1)!]^{1/[t+1]}) - (t+1)$$

For t fixed, then, we see that

$$f(r, t) \to \infty \quad \text{as} \quad r \to \infty$$

Specifically, since for $t \geqslant 0$

$$(t+1)! \leqslant (t+1)^{t+1}$$

we will have

$$f(r, t) \geqslant 2^{r/(t+1)}\,[2^{[r-(t+1)^2/[t(t+1)]} - (t+1)] - (t+1)$$

So if

$$r > (t+1)^2 + (t^2+t)\log_2(t+2)$$

then $f(r, t) > 0$ and $n_b(t) < n_c(t)$. This shows that the inequality of the statement of the theorem is satisfied for a non-trivial set of r and t.

We have exactly computed the interval $n_b(t) \leqslant n \leqslant n_c(t)$ for $r = 16$, 24 and 32 in Table 1. We see from this that for codes with dimensions near 576 the minimum distances will be 4, 6, 8 for redundances 16, 24, 32, respectively. We single out $n = 576$ (72 bytes) because some file-transfer protocols for personal computers use blocks with 68 information bytes, which implies codewords of 72 bytes if $r = 32$.

III. Minimum-Weight Generator for $r = 32$

To actually find a CRC code rather than just any code presents a minor annoyance, because using the relation $r' \leqslant r$ in the preceding section, we may have to pad with zeros. We will handle this in a somewhat ad hoc fashion.

We saw from Table 1 that for $568 \leqslant n \leqslant 1023$; we can generate a distance $d = 8$ ($t = 3$) optimum code using the generator polynomial

$$g(x) = (x + 1)\, G_{\mathrm{BCH}}(x)$$

(here the degree of $G_{\mathrm{BCH}}(x)$ is 30). This is because of the following reasoning. We find that

$$m = \left\lfloor \frac{32 - 1}{3} \right\rfloor = 10$$

If α is a primitive 2^{10}th root of unity over $GF(2)$, then we seek the degrees of the minimal polynomials of α (this degree is of course 10), α^3, and α^5. Now 1023 is not prime, but it turns out that α^3 and α^5 still have all 10 distinct conjugates. Thus, the three minimal polynomials $M^{(1)}(x)$, $M^{(3)}(x)$, and $M^{(5)}(x)$ are each of degree 10, and $g(x)$ is of degree 31.

To get a generator polynomial giving a CRC of degree 32, we can multiply $g(x)$ by x or by $x + 1$ to obtain a $g_1(x)$. In the former case, we will have the rightmost position always 0. In the latter case, the period of the degree-32 generator is not 1024 but 2048. Both of these "deficiencies" are irrelevant for our application, so we shall indeed work with just such $g_1(x)$'s.

Using the primitive trinomial

$$x^{10} + x^3 + 1$$

from Peterson's tables (Ref. 3, App. C, pp. 472-492), we obtain a primitive element α of $GF(2^{10})$. Now the construction of BCH codes generalizes from that used in the previous section. Specifically, let $0 \leqslant b \leqslant 2^m - 2$ and let s be prime to 2^{m-1} (=1023 in our case), where α is a primitive element in $GF(2^m)$. Consider the $2t$ elements

$$\alpha^b, \alpha^{b+s}, \alpha^{b+2s}, \ldots, \alpha^{b+(2t-2)s}, \alpha^{b+(2t-1)s}$$

Let these elements be contained in exactly t cyclotomic cosets, say

$$C_i, \quad i = 1, 3, 5, \ldots, 2t - 1$$

Let

$$M^{(i)}(x), \quad i = 1, 3, 5, \ldots, 2t - 1$$

be the t minimal polynomials of these cosets. Then the product

$$G_{\mathrm{BCH}}(x) = M^{(1)}\, M^{(3)}\, M^{(5)}, \ldots, M^{(2t-1)}$$

generates t-error-correcting BCH code.

We are interested in $m = 10$, s prime to $1023 = 3 \cdot 11 \cdot 31$, $0 \leqslant b \leqslant 1022$, $t = 3$. There are 1023 b's to check and

$$\phi(1023) = 1023 \left(1 - \frac{1}{3}\right)\left(1 - \frac{1}{11}\right)\left(1 - \frac{1}{31}\right)$$

$$= 2 \cdot 10 \cdot 30 = 600$$

values of s to check. By "check" we mean to find the $M^{(i)}$ and thus the weight (number of 1's) of $(x + 1)G_{\mathrm{BCH}}(x)$, or rather the minimum of the weights of $x(x + 1)G_{\mathrm{BCH}}(x)$ and $(x + 1)^2 G_{\mathrm{BCH}}(x)$, so that we get a polynomial of the desired degree 32.

The degree is 32, because in this case each $M^{(i)}$ has degree exactly 10. For, if one ever had degree less than 10, it would have degree at most 5, being a divisor of 10. The resulting $G_{\mathrm{BCH}}(x)$ would generate a code of length 1023, of distance $\geqslant 8$, with an $r \leqslant 26$. But calculations based on Theorem 1 imply that a code with $r \leqslant 26$ and of length 1023 cannot have minimum distance $\geqslant 8$.

Why is the code generated by $(x + 1)G_{\mathrm{BCH}}(x)$ of length 1023? If the 6 elements

$$\alpha^b, \alpha^{b+s}, \ldots, \alpha^{b+5s}$$

were all in a smaller field, then the element

$$\alpha^s = \frac{\alpha^{b+s}}{\alpha^b}$$

would be in that field as well. But s is prime to $n = 1023$, so α^s generates all of $GF(2^{10})$ and is not in a smaller subfield.

It looks as if we have $600 \cdot 1023 = 613{,}800$ polynomials (times 2 because of the x or $x + 1$ multiplier choice) of $g(x)$ to

check for weight. However, if (using the normal abuse of notation)

$$b,\ b+s,\ b+2s,\ b+3s,\ b+4s,\ b+5s$$

are in C_1, C_3, C_5, then so are

$$2b,\ 2b+2s,\ 2b+4s,\ 2b+6s,\ 2b+8s,\ 2b+10s$$

And since $(n, s) = 1$ implies $(n, 2s) = 1$ for n odd, we need only check the cases where b is a coset representative.

Further, if $(n, s) = 1$, then $(n, n - s) = 1$. So if

$$b,\ b+s,\ b+2s,\ b+3s,\ b+4s,\ b+5s$$

are in C_1, C_3, C_5, then

$$-b,\ -b-s,\ -b-2s,\ -b-3s,\ -b-4s,\ -b-5s$$

are in $\tilde{C}_1$, $\tilde{C}_3$, $\tilde{C}_5$, where $\tilde{C}_1$, $\tilde{C}_3$, $\tilde{C}_5$ contain the inverses of the elements in the cosets C_1, C_3, C_5. Thus the $G_{\mathrm{BCH}}(x)$ resulting from the choices $-b$ and $-s$ will be the reciprocal of the $G_{\mathrm{BCH}}(x)$ resulting from the choices b and s. When these reciprocals are multiplied by $(x + 1)$, the results are again reciprocals. Then, since multiplying by x doesn't change the weight and multiplying by $(x + 1)$ again preserves reciprocals, the polynomials resulting from b and s will have the same weights as those resulting from $-b$ and $-s$.

These considerations reduce our list of b's to only 55 that need to be checked. The polynomials in this list were enumerated and the lowest weight polynomials found were of weight 10. They are:

$$\left.\begin{aligned} g(x) &= x^{32}+x^{30}+x^{22}+x^{15}+x^{11}+x^7+x^6+x^5+x \\ &= x(x+1)(x^{10}+x^8+x^5+x^4+1)(x^{10}+x^7+x^6+x^3+1) \\ &\quad \times (x^{10}+x^9+x^8+x^6+x^2+x+1) \end{aligned}\right\} \text{(i)}$$

$$\left.\begin{aligned} g(x) &= x^{32}+x^{27}+x^{23}+x^{18}+x^{15}+x^{14}+x^{11}+x^8+x^4+x \\ &= x(x+1)(x^{10}+x^5+x^3+x^2+1)(x^{10}+x^9+x^5+x^4+1) \\ &\quad \times (x^{10}+x^6+x^5+x^3+x^2+x+1) \end{aligned}\right\} \text{(ii)}$$

We defined α as a root of $x^{10} + x^3 + 1$. Using this,

(1) α^{29} is a root of $x^{10}+x^8+x^5+x^4+1 = M^{(1)}(x)$

α^{87} is a root of $x^{10}+x^7+x^6+x^3+1 = M^{(3)}(x)$

α^{145} is a root of $x^{10}+x^9+x^8+x^6+x^2+x+1 = M^{(5)}(x)$

(2) α^{101} is a root of $x^{10}+x^5+x^3+x^2+1 = M^{(1)}(x)$

α^{303} is a root of $x^{10}+x^9+x^5+x^4+1 = M^{(3)}(x)$

α^{505} is a root of $x^{10}+x^6+x^5+x^3+x^2+x+1 = M^{(5)}(x)$

We recommend that one of these be adopted, specifically the second, because the gaps are more uniform. This may make hardware or software easier.

Of course, we had hoped to find a polynomial of weight 8; we do not know whether or not such a polynomial of any form exists. This seems hard to rule out.

IV. The Case r= 24

A similar search was carried out using irreducible polynomials of degree 11. We found

$$\left.\begin{aligned} g(x) &= x^{24}+x^{21}+x^{20}+x^{17}+x^{13}+x^{12}+x^3+1 \\ &= (x+1)^2(x^{11}+x^{10}+x^9+x^8+x^7+x^6+x^5+x^2+1) \\ &\quad \times (x^{11}+x^{10}+x^9+x^7+x^6+x^5+x^4+x^3+1) \end{aligned}\right\} \text{(i)}$$

$$\left.\begin{aligned} g(x) &= x^{24}+x^{22}+x^{12}+x^{10}+x^9+x^2+x+1 \\ &= (x+1)^2(x^{11}+x^9+1)(x^{11}+x^9+x^7+x^5+x^3+x+1) \end{aligned}\right\} \text{(ii)}$$

Both have weight 8 and give the largest minimum distance 6 for lengths n between 446 and 2047. We will get a weight-6 below, but only for n up to 1023. Here, α is a primitive element of $GF(2^{11})$, a root of the polynomial

$$x^{11}+x^2+1$$

Then,

(1) α^{163} is a root of $x^{11}+x^{10}+x^9+x^8+x^7+x^6+x^5+x^2+1 = M^{(1)}(x)$

α^{489} is a root of $x^{11}+x^{10}+x^9+x^7+x^6+x^5+x^4+x^3+1 = M^{(3)}(x)$

(2) α^{341} is a root of $x^{11}+x^9+x^7+x^5+x^3+x+1 = M^{(1)}(x)$

α^{1023} is a root of $x^{11}+x^9+1 = M^{(3)}(x)$

As an alternative approach, using irreducibles of degree 10, we found

$$g(x) = x^{24} + x^{23} + x^{14} + x^{12} + x^8 + 1$$

$$= (x^3 + x^2 + 1)(x + 1)(x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x + 1)$$

$$\times (x^{10} + x^9 + x^6 + x^4 + 1)$$

which (with its reciprocal) was the only polynomial of this type of weight 6. Again it gives a distance 6 code but it can be used only up to a length of 1023. It does, however, have weight equal to distance, 6 instead of 8, which is satisfying. Note that this time we pad the degree to 24 with $x^3 + x^2 + 1$, which complicates the periodicity or natural length. But the argument that the distance is 6 out to length 1023 and not beyond still works. We recommend this for $r = 24$ check bits if codewords of length around 600 are desired.

As before, α is a root of $x^{10} + x^3 + 1$, which is primitive. We have

(1) α^{19} is a root of $x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x + 1 = M^{(1)}(x)$

(2) α^{57} is a root of $x^{10} + x^9 + x^6 + x^4 + 1 = M^{(3)}(x)$

This completes our description of the optimal polynomials for $r = 32$ and $r = 24$.

We close this article by noting that for $r = 16$, the ADCCP polynomial is the best choice. From Table 1, distance 4 is the largest attainable. The polynomial is $(x + 1)$ times a primitive, so its distance is 4. And the weight is 4. It was a good choice as a standard.

References

1. Inose, Hiroshi, 1979, *An Introduction of Digital Integrated Communications Systems*, Univ. of Tokyo Press, Tokyo, Japan.
2. MacWilliams, F. J., and Sloane, N. J. A., 1977, *The Theory of Error-Correcting Codes*, North-Holland, Amsterdam.
3. Peterson, W. Wesley, and Weldon, Jr., E. J., 1972, *Error-Correcting Codes*, Second Edition, MIT Press, Cambridge, MA.

Table 1. The interval $n_b \leq n \leq n_c$ for various r and t

Largest Minimum Distance	t	Interval of Code Lengths n		
		$r = 16$	$r = 24$	$r = 32$
$d = 4$	$t = 1$	$362 \leqslant n \leqslant 2^{15}-1$	$5793 \leqslant n \leqslant 2^{23}-1$	$92682 \leqslant n \leqslant 2^{31}-1$
$d = 6$	$t = 2$	$74 \leqslant n \leqslant 2^{7}-1$	$466 \leqslant n \leqslant 2^{10}-1$	$2954 \leqslant n \leqslant 2^{15}-1$
$d = 8$	$t = 3$	no n	no n	$568 \leqslant n \leqslant 2^{10}-1$

SETI Investigations at Jodrell Bank, England: September Through November 1983

G. S. Downs
Communications Systems Research Section

S. Gulkis
Atmospheric Sciences Section

The RFI environment at the Nuffield Radio Astronomy Laboratories was examined extensively in the frequency band 1404 to 1444 MHz using the DSN Advanced Systems' Radio Frequency Interference Surveillance System. An uncooled FET amplifier (62 K system temperature) was placed on the newly acquired 12.8-m antenna, and unlimited observing time was made available to the SETI project. One night of observing at 1667 MHz was made available on the 76-m antenna. Preliminary results of four investigations are reported here: (1) full scans of the horizon, cataloging RFI events between 1404 and 1444 MHz; (2) lunar reflections of terrestrial RFI signals between 1424 and 1444 MHz; (3) noise background (spectral baseline) distortions caused by galactic neutral hydrogen emission at 1420.4 MHz; (4) low sensitivity search for spectral features of F, G, and K stars.

I. Introduction

Several groups of scientists and engineers in the United States are cooperatively devising plans to carry out a major, systematic search for radio signals of extraterrestrial origin. This search for extraterrestrial intelligence (SETI) is sponsored by the Office of Life Science within the NASA, with most of the effort concentrated at the Ames Research Center (ARC) of NASA and at JPL. The intent is to search the 1- to 10-GHz microwave band using a special purpose digital spectrum analyzer at existing radio telescopes. This analyzer will provide more than 10^6 channels with a spectral resolution up to 1 Hz. A prototype spectrum analyzer is being constructed at Stanford University that will provide 65,000 (2^{16}) channels with simultaneous resolutions of 1, 32, and 1000 Hz, and total power.

Field tests have begun using an existing 2^{16}-channel spectrum analyzer (Ref. 1 and Morris and Wilck, private communication), built at JPL originally as a radio frequency interference (RFI) surveillance system (RFISS). This device is limited to one spectral resolution, though it is now variable between 1 and 300 Hz. Both the JPL and Stanford spectrum analyzers will be used over the next five years to test hardware concepts, detection algorithms, and search strategies.

Of particular interest are the presence and character of RFI at potential observing sites. These interfering signals, not catalogued in advance, will cause long hours of delay while observers study a particular signal to determine whether it is of terrestrial origin, from a regularly passing satellite, or indeed from beyond the solar system. A frequency band of particular interest to SETI observers is that between emission lines at 1420 (HI, or neutral hydrogen) and 1667 MHz (OH radical). During September through November 1983, the Nuffield Radio Astronomy Laboratories at Jodrell Bank, England, made available to the SETI project a large amount of observing time on a new, moderately sensitive radio telescope at frequencies near 1420 MHz. This opportunity came at a time when field tests were just beginning, and the required time was not available at other observatories. Furthermore, the NASA Goldstone facility, where much field testing will be performed, does not yet have a receiving capability below 2250 MHz, and will not have this for at least one more year.

Therefore, the JPL mobile spectrum analyzer was shipped to Jodrell Bank for several weeks (1) to collect data on the RFI environment at a major observatory in a frequency band of critical importance, and (2) to determine how to improve the methods and associated software used to collect RFI data. To this end, the observing plan included the following activities:

(1) Using the 12.8-m telescope, scan the entire horizon at frequencies between 1420 and 1667 MHz as allowed by the current receiver.

(2) Explore, using the 12.8-m telescope, the RFI environment reflected by the Moon, and compare this with the known local environment.

Furthermore, several activities were planned that impact SETI search strategies:

(3) Obtain HI spectra at high spectral resolution (30 to 60 Hz) in selected regions of the Galaxy. (The question of baseline problems related to the presence of HI over a wide frequency band will be explored with this data).

(4) Obtain spectra at high resolution of F, G, and K stars located within 25 parsecs of the Sun. (A preliminary "eavesdropping" activity at a relatively low sensitivity.)

II. The Measurement System

The receiving system consisted of an uncooled FET amplifier mounted on the newly arrived 12.8-m antenna, providing a system temperature T_s of approximately 62 K in left-circular polarization at 1420 MHz. This value was obtained by observing the HI source S7 (peak brightness temperature of 80 K; R. D. Davies, private communication). *Assuming* the source is unresolved by the 1.4° beam (R. Pritchard, private communication), $T_s \sim 125$ K was obtained. Assuming an aperture efficiency of 0.5, $T_s \sim 62$ K. These assumptions require verification. The effective operating bandwidth was 1370 to 1440 MHz, the lower frequency determined by the receiver passband, and the upper frequency by the presence of a strong microwave link signal at 1457.625 MHz (the Mow Cop to Risley link).

The intermediate frequency (IF) signal from the Jodrell Bank system was passed on to an IF system in the NASA van (Fig. 1). It is the unique function of the JPL IF system to provide two signal channels of 10-MHz bandwidth, each with an in-phase (or real) and a quadrature (or imaginary) component. The system is usually used such that each channel samples a full 10 MHz of the IF passband. The two channels may be overlapping in frequency but represent two different polarizations (as in some of the OH observations reported below), or they may represent two contiguous portions of the same IF band (as in the RFI studies reported below).

Digitization for subsequent Fourier transformation was performed by four A/D converters. The hardware computation of the Fourier transform was done using the fast Fourier transform (FFT) algorithm. In fact, two independent 2^{15}-sample transforms were computed simultaneously, yielding a total of 2^{16} spectral points. The computations were pipelined, so that the usual two-dimensional butterfly diagram for the FFT algorithm must here be viewed as a three-dimensional diagram, with time the third dimension.

Running at the maximum rate (10-MHz bandwidth in each channel), a complete set of spectral points was computed every 3.2 ms. Subsequently, the real (V_r) and imaginary (V_i) components of the spectrum were combined to form the power spectrum ($V_r^2 + V_i^2$). These spectra were accumulated in a separate memory, to be read later by the controlling computer (Modcomp II) for storage on magnetic disk and tapes.

III. Low-Elevation Azimuth Scans

The 12.8-m radio telescope was used to scan 360° of horizon (elevation 10.0°) in several 10-MHz frequency bands. The volume of data and lack of automation precluded analyzing all the data during the visit to Jodrell Bank; however, partial results for the 1404- to 1414-MHz and the 1414- to 1424-MHz bands were obtained and are summarized in Table 1 and Fig. 2. The analysis summarized here resulted from a visual inspection of 664 spectral plots, each with a resolution of 16 kHz. (A compression of 53 is used in displaying most of the spectra, wherein the *maximum* value in each 16-kHz bin is displayed). The frequencies listed in Table 1 should therefore be considered reliable to within 50 kHz.

The number of RFI events encountered at each azimuth in the frequency bands of 1404 to 1414 and 1414 to 1424 MHz are presented in Fig. 2. (A complete data set has been collected for the bands 1424 to 1434 and 1434 to 1444 MHz, but the analysis has been deferred until the appropriate software exists). A major difficulty in characterizing RFI is readily apparent: *time variability*. It is unlikely near 1420 MHz that this variability is due to the source being an artificial satellite, but this possibility should always be kept in mind. The spatial correlation noted in Table 1 and suggested in Fig. 2 is undoubtedly due to the colocation of several transmitters (e.g., a microwave tower). In addition, several weaker signals in a particular direction may be merely modulation sidebands of the major signal in that direction.

Most signals appeared over a range of azimuth much larger than the primary antenna beam (approximately 1.4°). Several frequencies noted in Table 1 were seen over 180° of azimuth or more. We can only conclude that strong signals, surrounding hills, and local structures conspire to make some RFI inescapable. The one signal noted in Table 1 seen at all azimuths may be a signal generated within the receiver, though we observed a varying signal level in this case.

The conclusions to date are:

(1) Several *days* of observing are required upon arrival at a new observatory to build up a catalog to RFI events.

(2) The analysis of the RFI data must be automated, with particular attention to baseline removal in the presence of RFI, and to the appropriate threshold for cataloging an event such that true extraterrestrial signals are *not* included in the catalog.

(3) The catalog will necessarily grow as the survey proceeds and the sensitivity to weak signals increases.

Examples of spectra contaminated by RFI appear in Figs. 3 through 6. A modest, typical level of RFI is apparent in Fig. 3. Note, however, the presence of a signal within 300 kHz of the broad HI feature, well within the protected radio astronomy band. A more severe but not unusual level of RFI appears in Fig. 4. (The central peaks in all of these spectra are the result of a nonzero DC component in the signals presented to the A/D converters and should be ignored.) An unusually severe case is presented in Fig. 5, where several of the signals are clearly related to the strong interference near 1408.6 MHz. That same signal is presented in Fig. 6 at 16-kHz and 300-Hz resolution. These data were collected on two distinct days, partially illustrating time variations in the signal structure. Furthermore, the complexity and breadth of the signal are clearly brought out. Throughout the preliminary examination of these spectra, many exciting and unique examples of telemetry signals were found. A larger gallery of these snowflakes of the electromagnetic spectrum is clearly feasible, but such an endeavor is left to the final report.

To complete these observations, the frequency band 1404 to 1444 MHz was repeatedly monitored in 20 MHz blocks in long azimuth scans, each at an elevation of 10.0°, so that each azimuth was examined *several* times at each frequency. The scans proceeded slowly (0.2°/min), requiring 7 min to scan one beamwidth (1.4°). Thus 21 hours were required to fill the disk storage with 180 spectra while covering 250° in azimuth. (In fact, the scans proceeded at a much slower rate because of hardware difficulties and the considerable software development required to handle the data tapes properly.)

IV. Lunar Reflections of RFI

It may be possible to use the Moon to observe RFI generated around the globe. This technique is subject, however, to very severe restrictions. Let us compare a signal observed locally with a similar signal reflected from the Moon. The locally or directly observed signal will yield a receiver power P_d of

$$P_d = \frac{P_T G_T(\theta_d) A}{4 \pi D^2}$$

where

P_T = transmitter power

$G_T(\theta_d)$ = transmitter antenna gain at angle θ_d relative to the beam axis

A = effective antenna aperture

and

D = distance to the transmitter

However, the power P_r received via lunar reflections becomes

$$P_r = \frac{P_T G_T(\theta_r) \sigma A}{(4\pi)^2 R^4}$$

where

σ = σ (reflecting area, roughness, dielectric constant)

= radar cross section

and

R = the Earth-Moon distance

(We assume here that since angles of incidence and reflection are ~1°, the quasi-specular component of backscatter is in fact observed.) The ratio of these two received powers becomes

$$\frac{P_r}{P_d} = \frac{G_T(\theta_r)}{G_T(\theta_d)} \frac{\sigma D^2}{4\pi R^4}$$

Taking $\sigma \sim 0.07 \times$ (projected lunar area) $\sim 6.7 \times 10^5$ km^2, $D \sim 50$ km, and $R^4 \sim 2.1 \times 10^{22}$ km^4, we obtain

$$\frac{P_r}{P_d} \sim 6.4 \times 10^{-15} \frac{G_T(\theta_r)}{G_T(\theta_d)}$$

representing a nominal loss of −142 dB. If we observe P_d in a sidelobe where $G_T(\theta_d) \sim 10^{-4}\ G_T(0)$, and if $\theta_r \sim 0$ (direct transmission to the Moon), then P_r is only about 100 dB down from the signal observed locally. By various combinations of θ_r and θ_d we can see that lunar reflections will be 100 to 180 dB lower than those observed locally.

Nevertheless, estimates of the expected signal-to-noise (S/N) ratio expected for large communications antennas are at first surprisingly encouraging. We expect

$$S/N = \frac{P_r}{kT_s/\sqrt{\tau \Delta f}}$$

where T_s = the system temperature, Δf = the channel bandwidth, and τ = the integration time. Taking $G_T(0) \sim 55$ dB (a 26-m antenna at 21-cm wavelength), $P_T \sim 1$ kW, A $\sim 3.2 \times 10^{-6}$ km^2 (*assuming* an aperture efficiency of 0.5), and letting $g_T(\phi) = G_T(\phi)/G_T(0)$, then

$$\text{S/N} = 3 \times 10^7\ g_T(\phi)\ \tau^{1/2}$$

Naturally, the probability of a signal being directed at maximum gain towards the Moon is extremely low. Viewing leakage through sidelobes is considerably more probable. The expected S/N ratio for a range of sidelobe responses for both P_T = 1 kW and 100 W are summarized in Table 2. We have considered the transmitter power to be concentrated in one narrow channel in constructing this table. Clearly, the power would be divided among several channels, decreasing the probability of detection correspondingly.

Because of lunar libration, the reflected signals can be Doppler broadened as much as 20 Hz at 1420 MHz. Thus, a finer resolution (~30 Hz) is appropriate. However, these early observations were performed at the coarsest (300-Hz) resolution since our lack of experience in this technique does not justify confining the bandwidth to obtain only a modest improvement in sensitivity. All of the spectra examined in this preliminary study are confined to the 1424- to 1444-MHz frequency band. Few contained signals. Nevertheless, some signals were observed at high elevations (~50°), and an example is shown in Fig. 7. Two possible reflections are expanded to 300-Hz resolution in Fig. 8. A stronger statement on the source of these signals must await the analysis of the azimuth scans in this (1424- to 1444-MHz) frequency band. (While a preliminary analysis is available of the 1404- to 1424-MHz band, time permitted only cursory checks of the data quality in the higher frequency band.)

V. Galactic Spectra

Narrow HI features (width ~1 kHz) are not known to exist. The sharpness of these features relative to the spectral undulations imposed by receiver filters can be bothersome, however, if not anticipated in baseline-removal schemes. Therefore, some higher resolution (64-Hz) data in the 2-MHz band surrounding the HI line (1420.4 MHz) was collected at several galactic longitudes. An example is shown in Fig. 9(a), where 5 minutes of integration at (ℓ, b) = (0.0°, 0.0°) yielded a relatively complicated structure. A higher resolution view in Fig. 9(b) of the central region of one HI clump reveals no sharp surprises.

VI. F, G, and K Stars

The F, G, and K class stars bracket the Sun in size and temperature, and are therefore prime candidates for life-containing planets. Certain K stars are flare stars, creating harsh circumstellar conditions, and should not be included in the list (B. Lovell, private communication). The investigations at Jodrell Bank were limited to those previously examined at OH frequencies at Arecibo and whose spectra suggested sharp features may be present (J. Tarter, private communication). A night of Mark IV (76-m diameter) telescope observing was provided by R. J. Cohen to study four objects (star list #9100 and #9220B, Orion and Venus) in both left- and right-circular polarizations with a spectral resolution of 300 Hz. No sharp features were observed.

(The spectral studies by R. J. Cohen of OH emission in certain regions were supplemented using the RFISS with 20-MHz spectra at 300-Hz resolution and with dual-polarization 2-MHz spectra at 60-Hz resolution.)

Hydrogen emission was observed with the 12.8-m telescope in two objects, RGO 672 and Vega, with a resolution of 64 Hz. The velocity of the emission in RGO 672 was similar to that previously published, and no sharp feature was seen at the frequency suggested by the Arecibo observations. Two clumps of HI emission were observed in Vega at −15 and −25 km/s (relative to the local standard of rest), and no sharp features were observed.

Acknowledgments

Many members of the staff at Jodrell Bank contributed to this series of measurements, and to them the SETI project is deeply indebted. Participants at Jodrell Bank from elsewhere were G. Downs, R. Emerson, M. Grimm, S. Gulkis, and G. Stevens of JPL, and J. Tarter of ARC and UC Berkeley.

References

1. Morris, G. A., Jr., and Wilck, H. C., "JPL 2^{20} Channel 300 MHz Bandwidth Digital Spectrum Analyzer," in *Deep Space Network Progress Report 42-46,* Jet Propulsion Laboratory, Pasadena, California, 1978, pp. 57-61.

Table 1. Discrete RFI frequencies obtained in the 1404–1424 MHz band

Frequency (1404-1414), MHz	Frequency (1414-1424), MHz
1404.108	1414.068
1404.863	1414.772
1404.909	1414.818
1405.454	1414.909
1406.107	1415.045
1406.363	1415.127[c]
1406.803	1416.272
1407.045[a]	1416.500
1407.954	1417.678
1408.090	1417.714
1408.590[a]	1418.489
1409.402[b]	1418.768
1409.500	1419.863
1409.590	1420.000
1409.909	1420.090
1409.955	1420.199
1410.000[c]	1422.090
1410.202[b]	1422.498[c]
1410.590	1422.909
1410.727	1423.068
1410.879	1423.227
1411.068	
1411.591	
1411.800	
1411.954	
1412.136	
1412.973	

[a]Peaks near azimuth 130° to 132°, and contains considerable sideband structure.
[b]These frequencies appear with the 1408.590 signal and may be more than spatially correlated with it.
[c]Occurs over an azimuth range greater than 180°.

Table 2. Required integration time (τ) S/N ratio = 20

Sidelobe response $g_T(\phi)$, dB	Integration Times	
	P_T = 1 kW	P_T = 100 W
−80	70 min	4.6 d
−70	40 s	1.1 h
−60	0.4 s	40 s
−50	4 ms	0.4 s
−40	40 μs	4 ms
−30	0.4 μs	40 μs

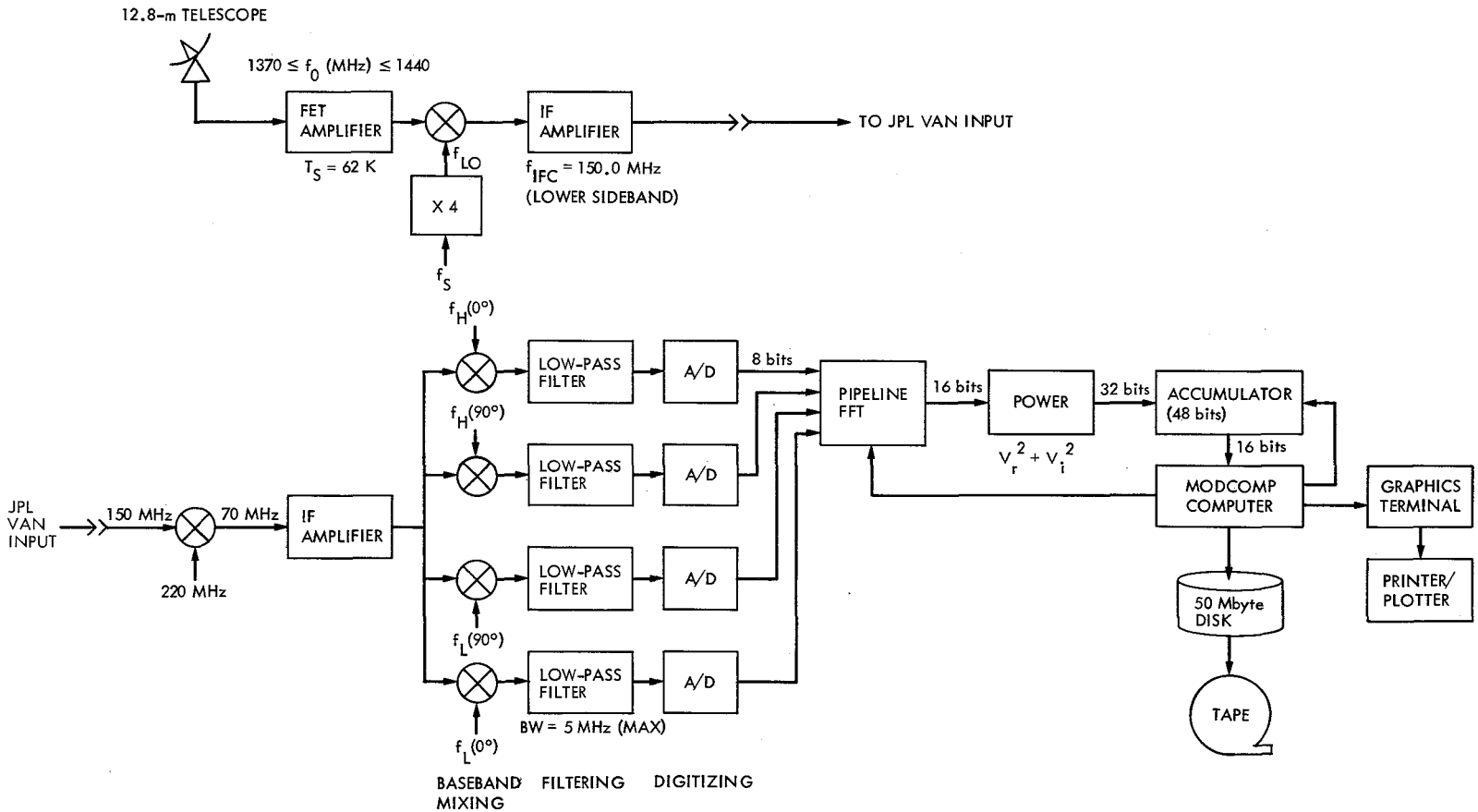


Fig. 1. The spectrometer system at Jodrell Bank

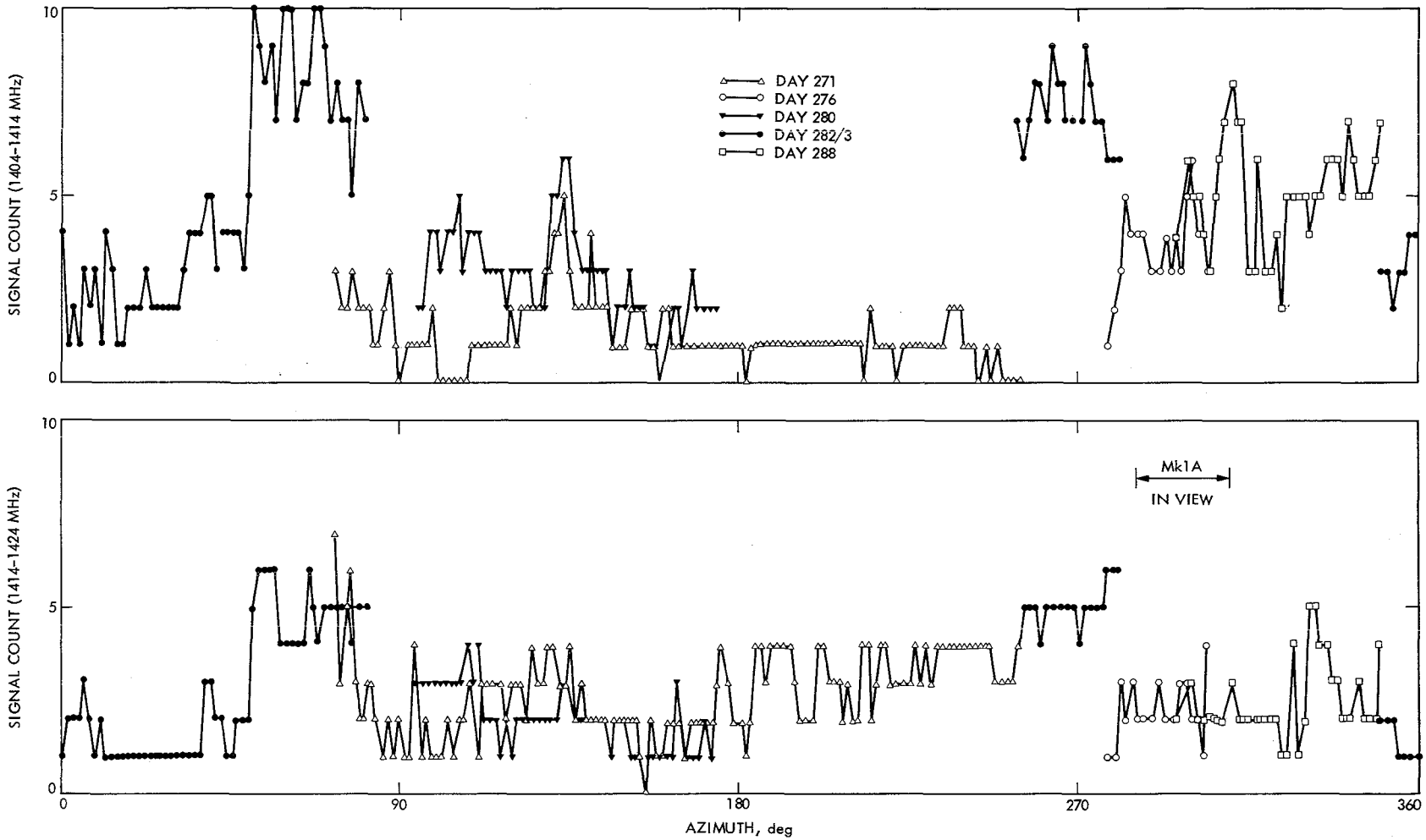


Fig. 2. Occurrences of RFI events versus azimuth in the 1404- to 1424-MHz frequency band

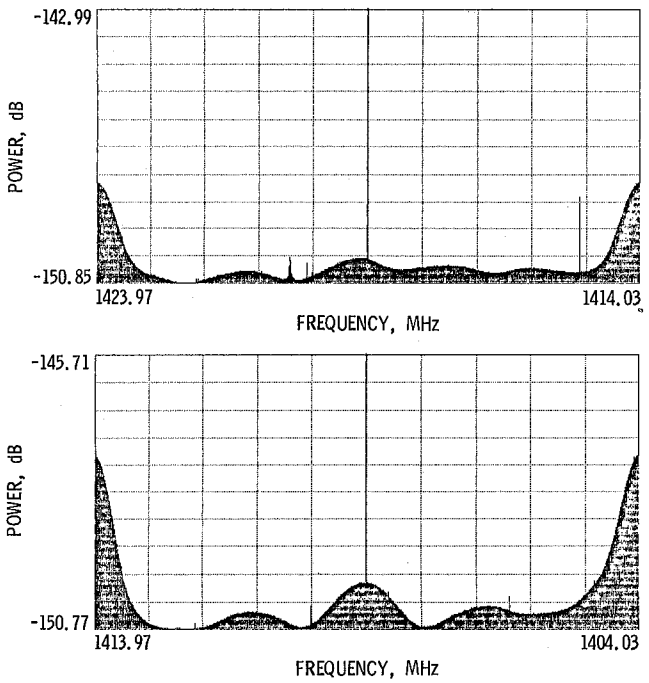


Fig. 3. Relative power (logarithmic) vs frequency near longitude 111.1° on day 288 (1983); resolution = 16 kHz

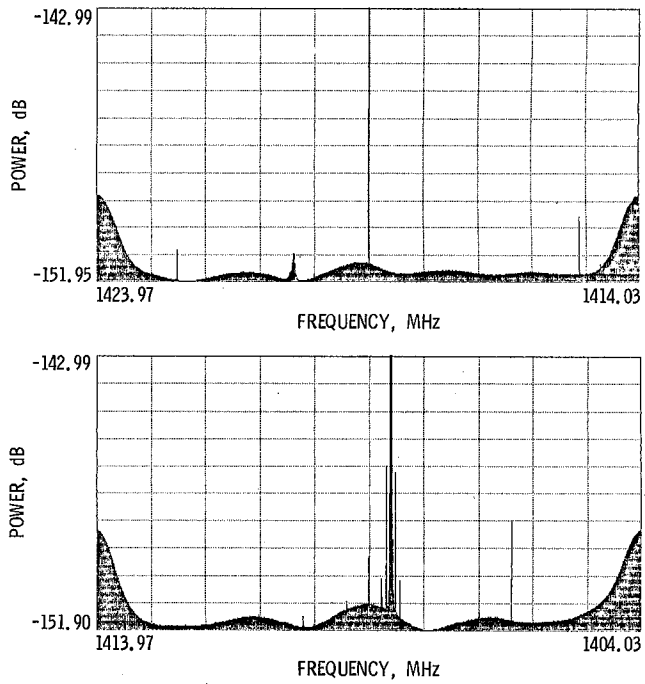


Fig. 5. Relative power (logarithmic) vs frequency near longitude 131.5° on day 288 (1983); resolution = 16 kHz

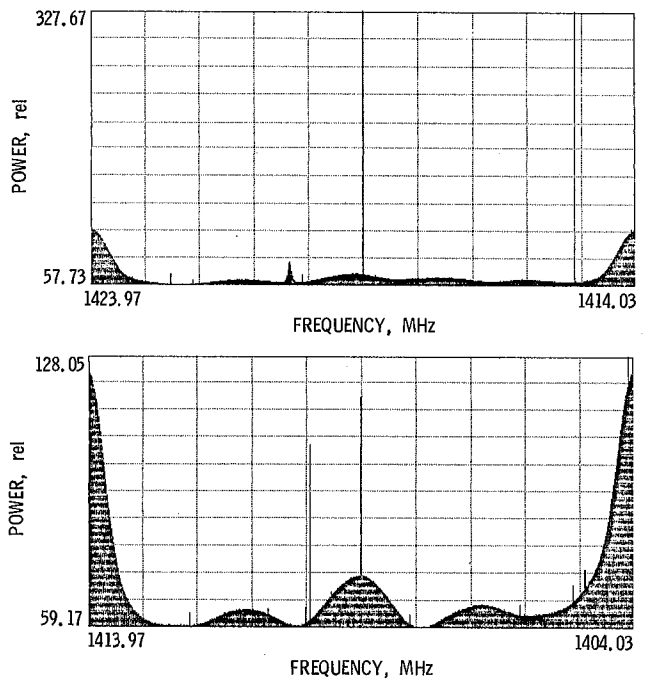


Fig. 4. Relative power (linear) vs frequency near longitude 67.1° on day 283 (1983); resolution = 16 kHz

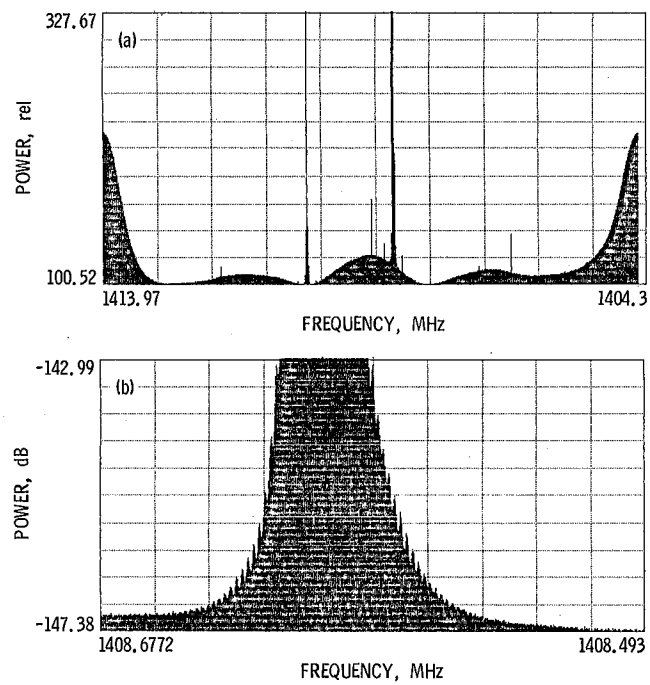


Fig. 6. Relative power (linear) vs frequency near longitude 130.1° on day 271 (1983): (a) resolution = 16 kHz; (b) high resolution view of feature in (a) near 1408.6 MHz; resolution = 300 Hz

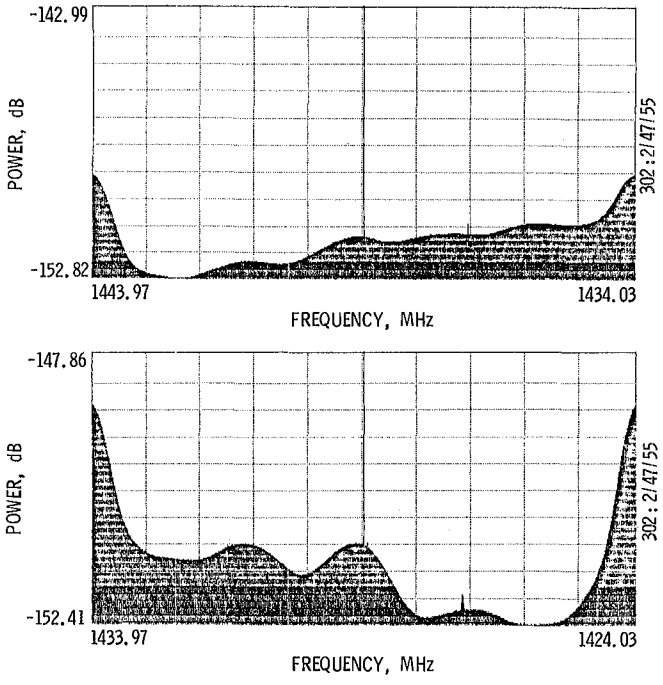


Fig. 7. Relative power (logarithmic) vs frequency while tracking the Moon on day 302 (1983)

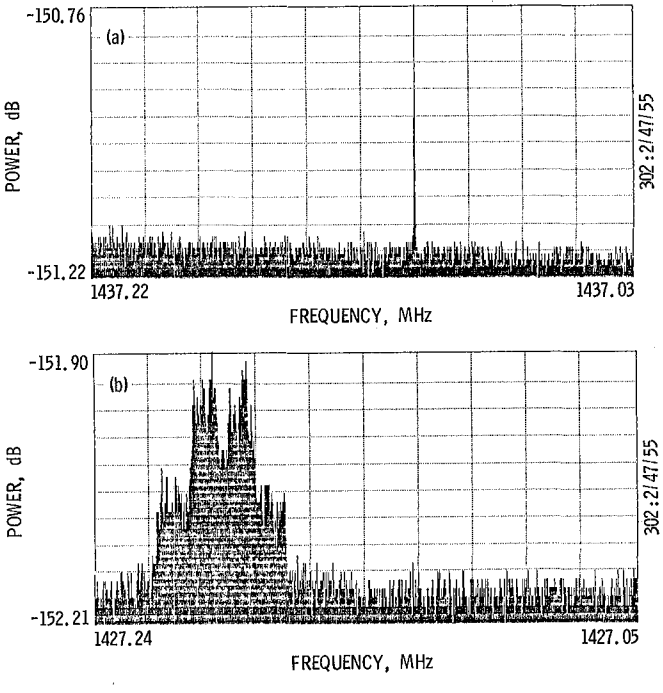


Fig. 8. Relative power (logarithmic) vs frequency of two *possible* lunar echoes on day 302 (1983), resolution = 300 Hz: (a) signal near 1437 MHz; (b) signal near 1427 MHz

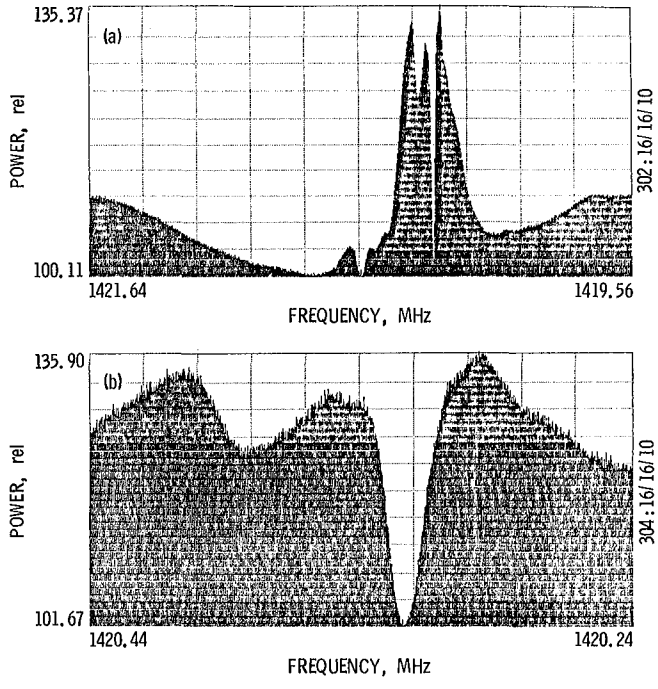


Fig. 9. Relative power (linear) vs frequency of a complicated HI feature at the galactic center: (a) resolution = 3.4 kHz; (b) view of central HI feature; resolution = 320 Hz

Author Index[1], 1983

The Telecommunications and Data Acquisition Progress Report

42-73, January–March, May 15, 1983
42-74, April–June, August 15, 1983
42-75, July–September, November 15, 1983
42-76, October–December, February 15, 1984

Abichandani, K.

42-73 Unified Analysis for Antenna Pointing and Structural Errors: Part I, Review, pp. 40-59.

Alberda, M. E.

42-76 Telemetry Simulation Assembly Implementation in the DSN, pp. 120-131.

Allen, L. H.

42-75 Superconducting Niobium Thin Film Slow-Wave Structures, pp. 39-51.

See Bautista, J. J.

Anderson, G.

42-73 Evaluation of Antenna Foundation Elastic Modulus, pp. 89-91.

See McGinness, H.

Bathker, D. A.

42-74 Interagency Array Study Report, pp. 117-148.

See Layland, J. W.

42-75 RF Performance of a Proposed L-Band Antenna System, pp. 91-97.

See Withington, J. R.

Batty, M. J.

42-74 A Southern Hemisphere VLBI Survey on a 275-km Baseline, pp. 183-187.

See Morabito, D. D.

Bautista, J. J.

42-75 Superconducting Niobium Thin Film Slow-Wave Structures, pp. 39-51.

J. J. Bautista, S. M. Petty, L. H. Allen, M. R. Beasely, and R. H. Hammond

Beasley, M. R.

42-75 Superconducting Niobium Thin Film Slow-Wave Structures, pp. 39-51.

See Bautista, J. J.

Bogan, J. R.

42-74 Planetary Radar, pp. 114-116.

J. R. Bogan, R. D. Shaffer, R. M. Goldstein, and R. F. Jurgens

Brokl, S. S.

42-75 Polynomial Driven Time Base and PN Generator, pp. 84-90.

Brockman, M. H.

42-76 Enhanced Radio Frequency Carrier Margin Improvement for an Array of Receiving Systems with Unequal Predetection Signal-to-Noise Ratios, pp. 170-188.

Brown, D. W.

42-74 Interagency Array Study Report, pp. 117-148.

See Layland, J. W.

Cha, A. G.

42-73 The JPL 1.5-Meter Clear Aperture Antenna with 84.5 Percent Efficiency, pp. 1-14.

[1] In the case of joint authorship, the reader is referred to the citation under the first author where all authors of the article are listed.

Charng, T.

42-73 The Application of the Implicit Alternating-Direction Numerical Technique to Thermal Analysis Involving Conduction and Convection, pp. 119-127.

See Hughes, R. D.

Chian, C. T.

42-74 NASTRAN Structural Model for the Large 64-Meter Antenna Pedestal, Part I, pp. 86-96.

C. T. Chian, M. S. Katow, and H. McGinness

42-75 NASTRAN Structural Model for the Large 64-Meter Antenna Pedestal, Part II – Improved Model, pp. 105-113.

Churchill, P. N.

42-76 First Results of GPS Time Transfer to Australia, pp. 141-160.

See Luck, J. McK.

Clauss, R. C.

42-74 Interagency Array Study Report, pp. 117-148.

See Layland, J. W.

Clements, P. A.

42-75 Intercontinental Time and Frequency Transfer Using a Global Positioning System Timing Receiver, pp. 98-104.

42-76 First Results of GPS Time Transfer to Australia, pp. 141-160.

See Luck, J. McK.

Cohen, E. J.

42-73 Radio Interferometric Determination of Source Positions, Intercontinental Baselines, and Earth Orientation With Deep Space Network Antennas – 1971-1980, pp. 128-155.

See Thomas, J. B.

Crowe, R. A.

42-75 Design Issues in the GCF Mark IV Development, pp. 132-139.

de Groot, N. F.

42-73 Developments Related to the Future Use of the 32-GHz Allocation for Deep Space Research, pp. 165-175.

Deutsch, L. J.

42-75 The Effects of Reed-Solomon Code Shortening on the Performance of Coded Telemetry Systems, pp. 14-20.

42-76 A Systolic VLSI Design of a Pipeline Reed-Solomon Decoder, pp. 99-113.

See Shao, H. M.

42-75 VLSI Architectures for Computing Multiplications and Inverses in GF(2^m), pp. 52-64.

See Wang, C. C.

Divsalar, D.

42-74 Symbol Stream Combining Versus Baseband Combining for Telemetry Arraying, pp. 13-28.

Downs, G. S.

42-76 SETI Investigations at Jodrell Bank, England: September Through November 1983, pp. 196-205.

G. S. Downs and S. Gulkis

Ellis, J.

42-74 Deep Space Navigation with Noncoherent Tracking Data, pp. 1-12.

42-75 Orbit Determination of Highly Elliptical Earth Orbiters Using VLBI and ΔVLBI Measurements, pp. 1-13.

See Frauenholz, R. B.

Fanselow, J. L.

42-73 Radio Interferometric Determination of Source Positions, Intercontinental Baselines, and Earth Orientation With Deep Space Network Antennas – 1971-1980, pp. 128-155.

See Thomas, J. B.

Faulkner, J.

42-74 A Southern Hemisphere VLBI Survey on a 275-km Baseline, pp. 183-187.

See Morabito, D. D.

Fowler, L.

42-75 An EPROM-Based Function Generator, pp. 79-83.

L. Fowler and J. A. McNeil

Frauenholz, R. B.

42-75 Orbit Determination of Highly Elliptical Earth Orbiters Using VLBI and ΔVLBI Measurements, pp. 1-13.

R. B. Frauenholz and J. Ellis

Freiley, A.

42-74 Antenna Microwave Subsystem Controller, pp. 63-72.

Galvez, J. L.

42-76 ISEE-3 Microwave Filter Requirements, pp. 114-119.

J. L. Galvez, H. Marlin, and P. Stanton

Goldstein, R. M.

42-74 Planetary Radar, pp. 114-116.

See Bogan, J. R.

Gulkis, S.

42-74 Gain Stability Measurements at S-Band and X-Band, pp. 159-168.

S. Gulkis and E. T. Olsen

42-76 SETI Investigations at Jodrell Bank, England: September Through November 1983, pp. 196-205.

See Downs, G. S.

Hammond, R. H.

42-75 Superconducting Niobium Thin Film Slow-Wave Structures, pp. 39-51.

See Bautista, J. J.

Harvey, B. R.

42-75 Results of the Australian Geodetic VLBI Experiment, pp. 140-146.

B. R. Harvey, A. Stolz, D. L. Jauncey, A. Niell, D. Morabito, and R. Preston

Hayes, N. V.

42-75 Energy Consumption Analysis of the Venus Deep Space Station (DSS-13), pp. 114-123.

Haynes, R. F.

42-74 A Southern Hemisphere VLBI Survey on a 275-km Baseline, pp. 183-187.

See Morabito, D. D.

Hird, E. E.

42-76 Operator Workload Measurement Validation for the Mark IV-A DSCC Monitor and Control Subsystem, pp. 132-140.

See Le May, M.

Hughes, R. D.

42-73 The Application of the Implicit Alternating-Direction Numerical Technique to Thermal Analysis Involving Conduction and Convection, pp. 119-127.

R. D. Hughes and T. Charng

42-74 Thermal Analysis of the X-Band 34-Meter Antenna Feedcone, pp. 73-85.

Jacobson, G. N.

42-74 Syntax Editing for Mark IV-A System Performance Test Software, pp. 97-111.

Jauncey, D. L.

42-74 A Southern Hemisphere VLBI Survey on a 275-km Baseline, pp. 183-187.

See Morabito, D. D.

42-75 Results of the Australian Geodetic VLBI Experiment, pp. 141-146.

See Harvey, B. R.

Johns, C. E.

42-76 X-Band Uplink Ground Systems Development, pp. 53-57.

Jurgens, R. F.

42-74 Planetary Radar, pp. 114-116.

See Bogan, J. R.

Katow, M. S.

42-74 NASTRAN Structural Model for the Large 64-Meter Antenna Pedestal, Part I, pp. 86-96.

See Chian, C. T.

42-75 Deformable Subreflector Computed by Geometric Optics, pp. 65-78.

M. S. Katow, I. Khan, and W. F. Williams

Katz, J.

42-75 Detectors for Optical Communications: A Review, pp. 21-38.

Kerridge, S. J.

42-74 Interagency Array Study Report, pp. 117-148.

See Layland, J. W.

Khan, I.

42-75 Deformable Subreflector Computed by Geometric Optics, pp. 65-78.

See Katow, M. S.

Klein, M. J.

42-74 Interagency Array Study Report, pp. 117-148.

See Layland, J. W.

Koerner, M. A.

42-76 Doppler System Phase Transfer Functions for a System With an X-Band Uplink and X-Band and S-Band Down Links, pp. 58-69.

Kohlhase, C. E.

42-74 Interagency Array Study Report, pp. 117-148.

See Layland, J. W.

Layland, J. W.

42-73 Limits to Arraying, pp. 156-164.

42-74 Interagency Array Study Report, pp. 117-148

J. W. Layland, A. M. Ruskin, D. A. Bathker, R. C. Rydgig, D. W. Brown, B. D. Madsen, R. C. Clauss, G. S. Levy, S. J. Kerridge, M. J. Klein, C. E. Kohlhase, J. I. Molinder, R. D. Shaffer, and M. R. Traxler

Lee, P. J.

42-73 Approximation to the Probability Density at the Output of the Photomultiplier Tube, pp. 36-39.

See Stokey, R. J.

42-76 Transfer Function Bounds for Partial-Unit-Memory Convolutional Codes Based on Reduced State Diagram, pp. 70-76.

Lansing, F. L.

42-74 Thermal Analysis of Antenna Structures: Part II – Panel Temperature Distribution, pp. 29-39.

See Schonfeld, D.

Le May, M.

42-76 Operator Workload Measurement Validation for the Mark IVA DSCC Monitor and Control Subsystem, pp. 132-140.

M. Le May, E. E. Hird, and B. Y. Rodriquez

Levitt, B. K.

42-74 SETI Pulse Detection Algorithm: Analysis of False-Alarm Rates, pp. 149-158.

Levy, G. S.

42-74 Interagency Array Study Report, pp. 117-148.

See Layland, J. W.

Lokshin, A.

42-74 The SETI Interpreter Program (SIP) – A Software Package for the SETI Field Tests, pp. 169-182.

See Olsen, E. T.

Lorden, G.

42-73 Node Synchronization for the Viterbi Decoder, pp. 22-35

G. Lorden, R. McEliece, and L. Swanson

Luck, Mck. J.

42-76 First Results of GPS Time Transfer to Australia, pp. 141-160.

J. Mck. Luck, J. R. Woodger, J. E. Wells, P. N. Churchill, and P. A. Clements

Lyzenga, G. A.

42-74 Two-Dimensional Finite Element Modeling for Modeling Tectonic Stress and Strain, pp. 188-202.

G. A. Lyzenga and A. Raefsky

Madsen, B. D.

42-74 Interagency Array Study Report, pp. 117-148.

See Layland, J. W.

Marlin, H.

42-76 ISEE-3 Microwave Filter Requirements, pp. 114-119.

See Galvez, J. L.

McEliece, R.

42-73 Node Synchronization for the Viterbi Decoder, pp. 22-35.

See Lorden, G.

McGinness, H.

42-73 Evaluation of Antenna Foundation Elastic Modulus, pp. 89-91.

H. McGinness and G. Anderson

42-74 NASTRAN Structural Model for the Large 64-Meter Antenna Pedestal, Part I, pp. 86-96.

See Chian, C. T.

McNeil, J. A.

42-75 An EPROM-Based Function Generator, pp. 79-83.

See Fowler, L.

Merkey, P.

42-76 Optimum Cyclic Redundancy Codes for Noisier Channels, pp. 189-195.

P. Merkey and E. C. Posner

Molinder, J. I.

42-74 Interagency Array Study Report, pp. 117-148.

See Layland, J. W.

Morabito, D. D.

42-74 A Southern Hemisphere VLBI Survey on a 275-km Baseline, pp. 183-187.

D. D. Morabito, R. A. Preston, J. Faulkner, A. E. Wehrle, D. L. Jauncey, M. J. Batty, R. F. Haynes, and A. E. Wright

42-75 Results of the Australian Geodetic VLBI Experiment, pp. 140-146.

See Harvey, B. R.

Niell, A.

42-75 Results of the Australian Geodetic VLBI Experiment, pp. 140-146.

See Harvey, B. R.

Olsen, E. T.

42-74 Gain Stability Measurements at S-Band and X-Band, pp. 159-168.

See Gulkis, S.

42-74 The SETI Interpreter Program (SIP) – A Software Package for the SETI Field Tests, pp. 169-182.

E. T. Olsen and A. Lokshin

Omura, J. K.

42-75 VLSI Architectures for Computing Multiplications and Inverses in GF(2^m), pp. 52-64.

See Wang, C. C.

Petty, S. M.

42-75 Superconducting Niobium Thin Film Slow-Wave Structures, pp. 39-51.

See Bautista, J. J.

Posner, E. A.

42-76 Optimum Cyclic Redundancy Codes for Noisier Channels, pp. 189-195.

See Merkey, P.

Preston, R. A.

42-74 A Southern Hemisphere VLBI Survey on a 275-km Baseline, pp. 183-187.

See Morabito, D. D.

42-75 Results of the Australian Geodetic VLBI Experiment, pp. 140-146.

See Harvey, B. R.

Purcell, Jr., G. H.

42-73 Radio Interferometric Determination of Source Positions, Intercontinental Baselines, and Earth Orientation With Deep Space Network Antennas – 1971-1980, pp. 128-155.

See Thomas, J. B.

Quach, C. T.

42-73 Techniques for Analysis of DSN 64-Meter Antenna Azimuth Bearing Film Height Records, pp. 92-118.

See Stevens, R.

Raefsky, A.

42-74 Two-Dimensional Finite Element Modeling for Modeling Tectonic Stress and Strain, pp. 188-202.

See Lyzenga, G. A.

Reed, I. S.

42-73 Simplified Syndrome Decoding of (n, 1) Convolutional Codes, pp. 15-21.

I. S. Reed and T. K. Truong

42-74 New Syndrome Decoding Techniques for the (n, k) Convolutional Codes, pp. 40-53.

I. S. Reed and T. K. Truong

42-75 VLSI Architectures for Computing Multiplications and Inverses in GF(2^m), pp. 52-64.

See Wang, C. C.

42-76 A Systolic VLSI Design of a Pipeline Reed-Solomon Decoder, pp. 99-113.

See Shao, H. M.

Reilly, H. F.

42-75 RF Performance of a Proposed L-Band Antenna System, pp. 91-97.

See Withington, J. R.

Resch, G. M.

42-76 Another Look at the Optimum Frequencies for a Water Vapor Radiometer, pp. 1-11.

42-76 Inversion Algorithms for Water Vapor Radiometers Operating at 20.7 and 31.4 GHz, pp. 12-26.

Rochblatt, D. J.

42-76 DSN Microwave Antenna Holography, pp. 27-42.

D. J. Rochblatt and B. L. Seidel

Rodriquez, B. Y.

42-76 Operator Workload Measurement Validation for the Mark IV-A DSCC Monitor and Control Subsystem, pp. 132-140.

See Le May, M.

Rogstad, D. H.

42-73 Radio Interferometric Determination of Source Positions, Intercontinental Baselines, and Earth Orientation With Deep Space Network Antennas – 1971-1980, pp. 128-155.

See Thomas, J. B.

Roth, M. G.

42-74 Phase Calibration for the Block I VLBI System, pp. 54-62.

M. G. Roth and T. F. Runge

Runge, T. F.

42-74 Phase Calibration for the Block I VLBI System, pp. 54-62.

See Roth, M. G.

Ruskin, A. M.

42-74 Interagency Array Study Report, pp. 117-148.

See Layland, J. W.

Rydgig, R. C.

42-74 Interagency Array Study Report, pp. 117-148.

See Layland, J. W.

Schonfeld, D.

42-74 Thermal Analysis of Antenna Structures: Part II – Panel Temperature Distribution, pp. 29-39.

D. Schonfeld and F. L. Lansing

Seidel, B. L.

42-76 DSN Microwave Antenna Holography, pp. 27-42.

See Rochblatt, D. J.

Shaffer, R. D.

42-74 Radio Astronomy, pp. 112-113.

See Wolken, P. R.

42-74 Planetary Radar, pp. 114-116.

See Bogan, J. R.

42-74 Interagency Array Study Report, pp. 117-148.

See Layland, J. W.

Shao, H. M.

42-75 VLSI Architectures for Computing Multiplications and Inverses in GF(2^m), pp. 52-64.

See Wang, C. C.

42-76 A Systolic VLSI Design of a Pipeline Reed-Solomon Decoder, pp. 99-113.

H. M. Shao, T. K. Truong, L. J. Deutsch, J. H. Yuen, and I. S. Reed

Skjerve, L. J.

42-73 Radio Interferometric Determination of Source Positions, Intercontinental Baselines, and Earth Orientation With Deep Space Network Antennas – 1971-1980, pp. 128-155.

See Thomas, J. B.

Sovers, O. J.

42-73 Radio Interferometric Determination of Source Positions, Intercontinental Baselines, and Earth Orientation With Deep Space Network Antennas – 1971-1980, pp. 128-155.

See Thomas, J. B.

Spitzmesser, D. J.

42-73 Radio Interferometric Determination of Source Positions, Intercontinental Baselines, and Earth Orientation With Deep Space Network Antennas – 1971-1980, pp. 128-155.

See Thomas, J. B.

Stanton, P.

42-76 ISEE-3 Microwave Filter Requirements, pp. 114-119.

See Galvez, J. L.

Stevens, R.

42-73 Techniques for Analysis of DSN 64-Meter Antenna Azimuth Bearing Film Height Records, pp. 92-118.

R. Stevens and C. T. Quach

42-76 Implementation of Large Antennas for Deep Space Mission Support, pp. 161-169.

Stokey, R. J.

42-73 Approximation to the Probability Density at the Output of a Photomultiplier Tube, pp. 36-39.

R. J. Stokey and P. J. Lee

Stolz, A.

42-75 Results of the Australian Geodetic VLBI Experiment, pp. 140-146.

See Harvey, B. R.

Swanson, L.

42-73 Node Synchronization for the Viterbi Decoder, pp. 22-35.

See Lorden, G.

Thomas, J. B.

42-73 Radio Interferometric Determination of Source Positions, Intercontinental Baselines, and Earth Orientation With Deep Space Network Antennas – 1971-1980, pp. 128-155.

J. B. Thomas, O. J. Sovers, J. L. Fanselow, E. J. Cohen, G. H. Purcell, Jr., D. H. Rogstad, L. J. Skjerve and D. J. Spitzmesser

Townes, S. A.

42-75 A Study of the Charged Particle Calibration Requirements for the Deep Space Network, pp. 124-131.

Traxler, M. R.

42-74 Interagency Array Study Report, pp. 117-148.

See Layland, J. W.

Truong, T. K.

42-73 Simplified Syndrome Decoding of $(n, 1)$ Convolutional Codes, pp. 15-21.

See Reed, I. S.

42-74 New Syndrome Decoding Techniques for the (n, k) Convolutional Codes, pp. 40-53.

See Reed, I. S.

42-75 VLSI Architectures for Computing Multiplications and Inverses in GF(2^m), pp. 52-64.

See Wang, C. C.

42-76 A Systolic VLSI Design of a Pipeline Reed-Solomon Decoder, pp. 99-113.

See Shao, H. M.

Vo, Q. D.

42-76 Signal-to-Noise Ratio and Combiner Weight Estimation for Symbol Stream Combining, pp. 86-98.

42-76 Simulations for Full Unit-Memory and Partial Unit-Memory Convolutional Codes With Real-Time Minimal-Byte-Error Probability Decoding Algorithm, pp. 77-85.

Wales, K. J.

42-73 The Network Information Management System (NIMS) in the Deep Space Network, pp. 85-88.

Wang, C. C.

42-75 VLSI Architectures for Computing Multiplications and Inverses in GF(2^m), pp. 52-64.

C. C. Wang, T. K. Truong, H. M. Shao, L. J. Deutsch, J. K. Omura and I. S. Reed

Wechsler, E. R.

42-73 Root Locus Algorithms for Programmable Pocket Calculators, pp. 60-64.

Wehrle, A. E.

42-74 A Southern Hemisphere VLBI Survey on a 275-km Baseline, pp. 183-187.

See Morabito, D. D.

Wells, J. E.

42-76 First Results of GPS Time Transfer to Australia, pp. 141-160.

See Luck, J. Mck.

Williams, W. F.

42-73 RF Design and Predicted Performance for a Future 34-Meter Shaped Dual-Reflector Antenna System Using the Common Aperture X-S Feedhorn, pp. 74-84.

42-75 Deformable Subreflector Computed by Geometric Optics, pp. 65-78.

See Katow, M. S.

Withington, J. R.

42-75 RF Performance of a Proposed L-Band Antenna System, pp. 91-97.

J. R. Withington, H. F. Reilly, Jr., and D. A. Bathker

Wolken, P. R.

42-74 Radio Astronomy, pp. 112-113.

P. R. Wolken and R. D. Shaffer

Woodger, J. R.

42-76 First Results of GPS Time Transfer to Australia, pp. 141-160.

See Luck, J. McK.

Wright, A. E.

42-74 A Southern Hemisphere VLBI Survey on a 275-km Baseline, pp. 183-187.

See Morabito, D. D.

Wu, S. C.

42-73 A Covariance Analysis for the Determination of Baselines Observing GPS Satellites, pp. 65-73.

Young, P. H.

42-76 Loop Gain and Circuit Parameters for Residual Carrier Tracking in the Advanced DSN Block V Receiver, pp. 43-52.

Yuen, J. H.

42-76 A Systolic VLSI Design of a Pipeline Reed-Solomon Decoder, pp. 99-113.

See Shao, H. M.

NASA—JPL—Coml., L.A., Calif.

End of Document