

NASA-CR-157456

NASA-CR-157456
19790022077

PERSONAL COPY

The Deep Space Network Progress Report 42-52

May and June 1979

August 15, 1979

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The Deep Space Network Progress Report 42-52

May and June 1979

August 15, 1979

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

79N30248*

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under NASA Contract No. NAS7-100.

Preface

Beginning with Volume XX, the Deep Space Network Progress Report changed from the Technical Report 32- series to the Progress Report 42- series. The volume number continues the sequence of the preceding issues. Thus, Progress Report 42-20 is the twentieth volume of the Deep Space Network series, and is an uninterrupted follow-on to Technical Report 32-1526, Volume XIX.

This report presents DSN progress in flight project support, tracking and data acquisition (TDA) research and technology, network engineering, hardware and software implementation, and operations. Each issue presents material in some, but not all, of the following categories in the order indicated.

Description of the DSN

Mission Support

- Ongoing Planetary/Interplanetary Flight Projects
- Advanced Flight Projects

Radio Astronomy

Special Projects

Supporting Research and Technology

- Tracking and Ground-Based Navigation
- Communications—Spacecraft/Ground
- Station Control and Operations Technology
- Network Control and Data Processing

Network and Facility Engineering and Implementation

- Network
- Network Operations Control Center
- Ground Communications
- Deep Space Stations
- Quality Assurance

Operations

- Network Operations
- Network Operations Control Center
- Ground Communications
- Deep Space Stations

Program Planning

- TDA Planning

In each issue, the part entitled “Description of the DSN” describes the functions and facilities of the DSN and may report the current configuration of one of the seven DSN systems (Tracking, Telemetry, Command, Monitor & Control, Test & Training, Radio Science, and Very Long Baseline Interferometry).

The work described in this report series is either performed or managed by the Tracking and Data Acquisition organization of JPL for NASA.

Contents

DESCRIPTION OF THE DSN

Network Functions and Facilities1

N. A. Renzetti

MISSION SUPPORT

Ongoing Planetary/Interplanetary Flight Projects

Pioneer 11 Saturn Encounter Mission Support 4

R. B. Miller

NASA Code 311-03-22-60

SUPPORTING RESEARCH AND TECHNOLOGY

Tracking and Ground-Based Navigation

The Gyrotron: A High-Frequency Microwave Amplifier 8

A. Kupiszewski

NASA Code 310-10-64-10

**Connection and Validation of Narrow-Band
ΔVLBI Phase Observations** 13

S. C. Wu

NASA Code 310-10-60-10

**An Improved Storage Bulb Mount for DSN
Hydrogen Masers** 21

P. R. Dachel, D. P. Russell, T. K. Tucker, and L. B. Stratman

NASA Code 310-10-62-12

**A Generalization of Binary Minimum Shift Keying and Staggered
Quadrature Shift Keying Modulation** 26

J. L. Massey

NASA Code 310-10-60-12

**Internal Noise of a Phase-Locked Receiver
With a Loop-Controlled Synthesizer** 41

C. A. Greenhall

NASA Code 310-10-60-12

Communications – Spacecraft/Ground

**A Prototype DSN X-S Band Feed: DSS 13
Application Status (Third Report)** 51

W. Williams, D. Nixon, H. Reilly, J. Withington,
and D. Bathker

NASA Code 310-20-65-14

Station Control and Operations Technology

Coding for Optical Channels With Photon-Counting	61
---	-----------

R. J. McEliece and L. R. Welch
NASA Code 310-30-70-14

NETWORK AND FACILITY ENGINEERING AND IMPLEMENTATION

Network

A Fast Technique for Computing Syndromes of BCH and RS Codes	67
---	-----------

I. S. Reed, T. K. Truong, and R. L. Miller
NASA Code 311-03-43-20

Deep Space Stations

New X-Band Antenna Feeds for the DSN 64-Meter Stations	71
---	-----------

R. Hartop
NASA Code 311-03-42-55

Multi-Mission Receiver (MMR)	75
---	-----------

H. Donnelly and H. Nishimura
NASA Code 311-03-42-15

Preliminary Report on DSN System Performance Under Local Weather Effects	82
---	-----------

J. M. Urech
NASA Code 311-03-43-10

Comparison of Two Total Energy Systems for a Diesel Power Generation Plant	89
---	-----------

V. W. Chai
NASA Code 311-03-41-08

Minimizing the RMS Surface Distortions from Gravity Loadings of the 34-m HA-DEC Antenna for Deep Space Missions	94
--	-----------

M. S. Katow
NASA Code 311-03-42-59

Interim Radio Spectrum Surveillance Station	99
--	-----------

D. R. Hersey
NASA Code 311-03-42-90

X-Band Atmospheric Noise Temperature Data and Statistics at Goldstone, DSS 13, 1977-1978	108
---	------------

S. D. Slobin, K.B. Wallace, M. M. Franco, and E. M. Andres
NASA Code 311-03-42-54

Network Functions and Facilities

N. A. Renzetti

Office of Tracking and Data Acquisition

The objectives, functions, and organization of the Deep Space Network are summarized; deep space station, ground communication, and network operations control capabilities are described.

The Deep Space Network was established by the National Aeronautics and Space Administration (NASA) Office of Space Tracking and Data Systems and is under the system management and technical direction of the Jet Propulsion Laboratory (JPL). The network is designed for two-way communications with unmanned spacecraft traveling approximately 16,000 km (10,000 miles) from Earth to the farthest planets and to the edge of our solar system. It has provided tracking and data acquisition support for the following NASA deep space exploration projects: Ranger, Surveyor, Mariner Venus 1962, Mariner Mars 1964, Mariner Venus 1967, Mariner Mars 1969, Mariner Mars 1971, and Mariner Venus-Mercury 1973, for which JPL has been responsible for the project management, the development of the spacecraft, and the conduct of mission operations; Lunar Orbiter, for which the Langley Research Center carried out the project management, spacecraft development, and conduct of mission operations; Pioneer, for which Ames Research Center carried out the project management, spacecraft development, and conduct of mission operations; and Apollo, for which the Lyndon B. Johnson Space Center was the project center and the Deep Space Network supplemented the Manned Space Flight Network, which was managed by the Goddard Space Flight Center. The network is currently providing tracking and data acquisition support for Helios, a joint U.S./West German project; Viking, for which Langley Research Center provided the Lander spacecraft and project management until May,

1978, at which time project management and mission operations were transferred to JPL, and for which JPL provided the Orbiter spacecraft; Voyager, for which JPL provides project management, spacecraft development, and is conducting mission operations; and Pioneers, for which the Ames Research Center provides project management, spacecraft development, and conduct of mission operations. The network is adding new capability to meet the requirements of the Galileo mission to Jupiter, for which JPL is providing the Orbiter spacecraft, and the Ames Research Center the probe. In addition, JPL will carry out the project management and the conduct of mission operations.

The Deep Space Network (DSN) is one of two NASA networks. The other, the Spaceflight Tracking and Data Network (STDN), is under the system management and technical direction of the Goddard Space Flight Center (GSFC). Its function is to support manned and unmanned Earth-orbiting satellites. The Deep Space Network supports lunar, planetary, and interplanetary flight projects.

From its inception, NASA has had the objective of conducting scientific investigations throughout the solar system. It was recognized that in order to meet this objective, significant supporting research and advanced technology development must be conducted in order to provide deep space telecommunications for science data return in a cost effective

manner. Therefore, the Network is continually evolved to keep pace with the state of the art of telecommunications and data handling. It was also recognized early that close coordination would be needed between the requirements of the flight projects for data return and the capabilities needed in the Network. This close collaboration was effected by the appointment of a Tracking and Data Systems Manager as part of the flight project team from the initiation of the project to the end of the mission. By this process, requirements were identified early enough to provide funding and implementation in time for use by the flight project in its flight phase.

As of July 1972, NASA undertook a change in the interface between the Network and the flight projects. Prior to that time, since 1 January 1964, in addition to consisting of the Deep Space Stations and the Ground Communications Facility, the Network had also included the mission control and computing facilities and provided the equipment in the mission support areas for the conduct of mission operations. The latter facilities were housed in a building at JPL known as the Space Flight Operations Facility (SFOF). The interface change was to accommodate a hardware interface between the support of the network operations control functions and those of the mission control and computing functions. This resulted in the flight projects assuming the cognizance of the large general-purpose digital computers which were used for both network processing and mission data processing. They also assumed cognizance of all of the equipment in the flight operations facility for display and communications necessary for the conduct of mission operations. The Network then undertook the development of hardware and computer software necessary to do its network operations control and monitor functions in separate computers. A characteristic of the new interface is that the Network provides direct data flow to and from the stations; namely, metric data, science and engineering telemetry, and such network monitor data as are useful to the flight project. This is done via appropriate ground communication equipment to mission operations centers, wherever they may be.

The principal deliverables to the users of the Network are carried out by data system configurations as follows:

- (1) The DSN Tracking System generates radio metric data, i.e., angles, one- and two-way doppler and range, and transmits raw data to Mission Control.
- (2) The DSN Telemetry System receives, decodes, records, and retransmits engineering and scientific data generated in the spacecraft to Mission Control.
- (3) The DSN Command System accepts spacecraft commands from Mission Control and transmits the commands via the Ground Communication Facility to a

Deep Space Station. The commands are then radiated to the spacecraft in order to initiate spacecraft functions in flight.

- (4) The DSN Radio Science System generates radio science data, i.e., the frequency and amplitude of spacecraft transmitted signals affected by passage through media such as the solar corona, planetary atmospheres, and planetary rings, and transmits this data to Mission Control.
- (5) The DSN Very Long Baseline Interferometry System generates time and frequency data to synchronize the clocks among the three Deep Space Communications complexes. It will generate universal time and polar motion and relative Deep Space Station locations as by-products of the primary data delivery function.

The data system configurations supporting testing, training, and network operations control functions are as follows:

- (1) The DSN Monitor and Control System instruments, transmits, records, and displays those parameters of the DSN necessary to verify configuration and validate the Network. It provides the tools necessary for Network Operations personnel to control and monitor the Network and interface with flight project mission control personnel.
- (2) The DSN Test and Training System generates and controls simulated data to support development, test, training and fault isolation within the DSN. It participates in mission simulation with flight projects.

The capabilities needed to carry out the above functions have evolved in the following technical areas:

- (1) The Deep Space Stations, which are distributed around Earth and which, prior to 1964, formed part of the Deep Space Instrumentation Facility. The technology involved in equipping these stations is strongly related to the state of the art of telecommunications and flight-ground design considerations, and is almost completely multimission in character.
- (2) The Ground Communications Facility provides the capability required for the transmission, reception, and monitoring of Earth-based, point-to-point communications between the stations and the Network Operations Control Center at JPL, Pasadena, and to the JPL Mission Operations Centers. Four communications disciplines are provided: teletype, voice, high-speed, and wideband. The Ground Communications Facility uses the capabilities provided by common carriers throughout the world, engineered into an integrated system by Goddard Space Flight Center, and controlled from the

communications Center located in the Space Flight Operations Facility (Building 230) at JPL.

The Network Operations Control Center is the functional entity for centralized operational control of the Network and interfaces with the users. It has two separable functional elements; namely, Network Operations Control and Network Data Processing. The functions of the Network Operations Control are:

- (1) Control and coordination of Network support to meet commitments to Network users.
- (2) Utilization of the Network data processing computing capability to generate all standards and limits required for Network operations.
- (3) Utilization of Network data processing computing capability to analyze and validate the performance of all Network systems.

The personnel who carry out the above functions are located in the Space Flight Operations Facility, where mission opera-

tions functions are carried out by certain flight projects. Network personnel are directed by an Operations Control Chief. The functions of the Network Data Processing are:

- (1) Processing of data used by Network Operations Control for control and analysis of the Network.
- (2) Display in the Network Operations Control Area of data processed in the Network Data Processing Area.
- (3) Interface with communications circuits for input to and output from the Network Data Processing Area.
- (4) Data logging and production of the intermediate data records.

The personnel who carry out these functions are located approximately 200 meters from the Space Flight Operations Facility. The equipment consists of minicomputers for real-time data system monitoring, two XDS Sigma 5s, display, magnetic tape recorders, and appropriate interface equipment with the ground data communications.

Pioneer 11 Saturn Encounter Mission Support

R. B. Miller
Mission Support Office

The Pioneer 11 spacecraft will fly by the Planet Saturn in August and September 1979. The Pioneer 11 Saturn encounter activities are described, followed by the DSN plans for supporting the event.

I. Pioneer 11 Saturn Encounter Description

The Pioneer 11 spacecraft was launched on April 6, 1973, and flew by the planet Jupiter, with the closest approach on December 3, 1974. After nearly 6-1/2 years of flight, the Pioneer 11 spacecraft will have a closest approach to the planet Saturn on September 1, 1979. At that time, the spacecraft will be over 10 astronomical units from Earth, and the radio transmission will take 2 hours and 53 minutes from Earth out to the spacecraft and back to Earth again.

The near-encounter trajectory is pictured in Figs. 1 through 3 (which were provided by the JPL Pioneer Navigation Team). Figure 1 is a view of Saturn from Earth during the closest approach with the spacecraft entering from the upper left and penetrating the ring plane at the point marked "inbound descending node," which is at a distance of 2.87 Saturn radii (R_s) from the center of Saturn. The ring plane penetration is symmetrical and takes place at closest approach ± 2 hours at the same distance from the center of Saturn. The 2.87- R_s distance was selected because this is the distance that Voyager 2 will have to penetrate the ring plane if it is to go on to the planet Uranus. Therefore, Pioneer 11 will provide two data points to determine whether that is a safe distance to penetrate the ring plane. Periapsis occurs at 1.36 R_s , just one-third of a radius of Saturn above the cloud tops, which is

81,426 km from the center of the planet. Periapsis passage will occur at approximately 1800 GMT, ground observed time. Figure 2 shows the trajectory in a polar view and includes the orbit of the inner moon Mimas, and the visible ring. The line marked "occultation zones" shows the location of the spacecraft as it passes into the shadow of Saturn; this location is essentially the same as the point where Earth is occulted by the planet. Notice that periapsis occurs just before enter occultation, right under the inner visible ring (referred to as the C-ring). Figure 3 is a scale drawing (except for ring thickness), of Saturn, with rings edge on, which rotates with the trajectory to show the spacecraft path with respect to the ring plane and the disk of Saturn. At the inbound and outbound crossing points, labeled "a" and "c", the spacecraft is at the distance that duplicates Voyager 2's Uranus trajectory. This will also be approximately the periapsis distance for the Voyager 2 mission. When the spacecraft passes under the outer edge of the A-ring, it will be only 2,000 km from the ring plane, while at the closest approach point, labeled "b", the spacecraft will be approximately 9,000 km below the ring plane. The ring plane penetration times of closest approach ± 2 hours also corresponds to the closest distance for which it is expected that Pioneer 11 will be able to successfully image the planet. This is because the relative motion of the target body will be too great for the spin scan imaging system. However, while under the ring plane, several imaging single-

spin scans of the ring plane will be performed to gain, hopefully, information about other smaller-scale features of the rings, such as smaller gaps than can be detected by Earth-based observation.

The encounter time period is defined as August 3 to October 2, 1979. In this time period, approximately 15,000 commands will be transmitted to the spacecraft, all of which will be acted upon at the time of receipt. The majority of these commands are required for the operation of the Imaging Photopolarimeter. To understand the operation of this instrument and why it requires such an extensive amount of commanding, see Ref. 1. The Earth occultation will last from closest approach to E + 79 minutes.

The imaging will exceed Earth-based resolution (which is approximately 1,200 km) at ± 6 days around closest approach. The best resolution of Saturn is expected to be on the order of 80 to 100 km achieved at $E \pm 2$ hours. The rings will fill the field of view at 39 hours before closest approach, and the disk of the planet will fill the field of view 33 hours before closest approach. The trajectory allows for good imaging after closest approach. Approximately 50 images better than Earth-based are expected, with 20 of them being before and 30 after closest approach. Several of the satellites will be observed, both in imaging and polarimetry modes, with the best target being Titan. Closest approach to Titan will be 356,000 km at 25-1/2 hours after closest approach. At this distance, the resolution is expected to be about 180 km.

II. Voyager's Expectations for the Pioneer 11 Saturn Encounter

There are several areas in which the scientists involved in the sequence planning activities of Voyager hope to benefit from the data returned by Pioneer 11. First and foremost is to determine if $2.87 R_s$ is a safe distance to penetrate the ring plane. If Pioneer 11 is not able to safely pass through the ring plane at this distance, it is quite likely that the Uranus option of Voyager 2 will be deleted.

There are several areas in which Voyager hopes to benefit in the general category of characterizing the fields and particle environment. Earth-based observation indicates that the fields and particle environment around Saturn is one-tenth or less as intense as it is for Jupiter; however, there is great uncertainty in the Earth-based observations. It is hoped that the Pioneer 11 data will enable the determination of the magnitude and orientation of the magnetic field around Saturn. In particular, Voyager is interested in the orientation of the magnetic equator with respect to the planet's equator in order to properly time Voyager's fields and particle roll maneuvers. These are maneuvers where the entire Voyager spacecraft is

rolled in order to act as a "scan platform" for the fields and particle instruments at key times in passing through Saturn's magnetosphere. It is desired to have these roll maneuvers symmetrically on either side of the magnetic equator and one right on the magnetic equator. Another item is to know whether the corotating magnetosphere of Saturn extends as far as the moon Titan. If the corotating magnetosphere extends as far as Titan, there will be a requirement to observe the magnetospheric wake behind Titan that will prevent the spacecraft from executing a roll maneuver in order to look back at Titan after closest approach to that moon. Characterizing the intensity and energy levels of the fields and particle environment will help Voyager determine what special precautions may be necessary to keep the spacecraft and its instruments safe, as well as help in determining optimum gain settings for the fields and particle instruments.

If the Pioneer 11 imaging is able to identify any unique visual features of Saturn, Voyager would use this information to refine the imaging plans. In particular, the Voyager team hopes that the Pioneer 11 imaging of Titan will enable a decrease in the uncertainties of the radius of that body. Current uncertainties in the radius of Titan are ± 300 km; it is hoped that the Pioneer 11 data will reduce that uncertainty to on the order of 100 km. This would enable Voyager to produce a more efficient mosaicing plan of the moon Titan, as well as to assure the return of a limb picture with one-third as many pictures dedicated to that purpose.

Lastly, Pioneer 11's polarimetry of Saturn and its satellites may be the only polarimetry into the distant future, since this function has failed on the Voyager 1 spacecraft and may not be used on the Voyager 2 spacecraft because of the fear that the rotating filter wheels might stick in an undesirable position.

III. DSN Support Plans

Even though the Pioneer 11 Saturn encounter will involve the same level of activity as the Pioneer 10 and 11 Jupiter encounters, very limited special training on the part of the DSN is necessary to prepare for the event because it will look to the DSN very similar to the configurations and level of activity required for supporting the on-going Pioneer Venus orbital operations. The training will, therefore, concentrate on the special functions that are added to support the Saturn event.

Only a single significant implementation in addition to the configuration supplied for the Pioneer 10 and 11 Jupiter encounters was necessary for the Pioneer 11 Saturn encounter. This implementation was needed to push the telecommunications performance to the absolute limit available with current

technology: necessitated by attempting a Saturn encounter with a spacecraft that was designed for the Jupiter distance. The specific implementation was the installation of low-noise S-band masers at Australia and Spain; these installations were completed on May 6, 1979. The masers give an improvement in performance of approximately 0.8 dB when operated in a diplex mode (transmit and receive simultaneously). A similar low-noise maser had already existed at the Goldstone 64-meter station. An additional implementation done on a short-term basis is the installation of a programmable receiver at Station 12 to accommodate the high doppler rates that will be experienced at the periapsis passage.

In addition, to gain additional performance for ± 7 days around closest approach (which covers the time period when Earth-based resolution is exceeded for the imaging instrument), the listen-only mode will be utilized at the 64-meter sites; this will give an additional 0.7 dB performance improvement. Because of the extensive command activity required, a second smaller station will have to be scheduled with each 64-meter site for this 14-day period. The stations that will support during these 14 days are: DSSs 12 and 14 (34- and 64-meter stations at Goldstone), DSSs 44 and 43 (Honey-suckle and Tidbinbilla in Australia), and DSSs 62 and 63 (Cebreros and Robledo near Madrid, Spain).

An R&D real-time arraying system, which has been installed at Goldstone on a demonstration basis for the Voyager 2 Jupiter encounter, will be utilized for the 14 days around closest approach. This system will receive the spacecraft signal at the 64-meter site and 34-meter site simultaneously, with a microwave transmission of the 34-meter data over to the 64-meter site where the signals are automatically phase-matched and added to achieve a performance improvement that is expected to be greater than one-half a dB.

Against a Project requirement for continuous 64-meter coverage for closest approach ± 30 days, the actual coverage

negotiated between Voyager, Pioneer Venus, and Pioneer 11 will be 10 to 19 hours a day from 30 to 15 days before closest approach, and then continuous coverage from that point in time until 30 days after closest approach.

There will be some special training of each shift at all stations to support extensive, contingency manual commanding required for the Saturn encounter. Since the imaging sequence is executed by the receipt of real-time commands, the possibility of a loss of communications circuits or a computer outage at the Ames Research Center, where the commands originate, is covered by plans that allow uninterrupted command capability for the stations by voice instruction from Ames. The level of commanding required is much greater than is usually used for manual commanding, so practice exercises have been scheduled to cover all affected shifts.

At least two arraying tests will be scheduled to measure the actual performance improvement for the Pioneer 11 bit rate and coding type. In addition, it is expected that one or two periapsis passage training exercises will be scheduled for Goldstone and Madrid because of the complexity of the operations required during that time period. Complex receiver tuning will be necessary to maintain receivers in lock because of the high doppler rates that will be experienced. This is the reason that the programmable receiver had to be installed at Station 12 so that that station could retain its receivers in lock to support the arraying. Programmable receivers already exist at the 64-meter sites. There will also be complex requirements to support radio science experiments, and to operate the occultation data assembly equipment and associated special recording functions. Finally, to guard the supercritical data, particularly at ring plane crossing and exit occultation, an activity is underway to investigate the feasibility of utilizing the precarrier detection telemetry recovery technique that was implemented for the Pioneer Venus Multiprobe entry event; this would provide a back-up analog recording in the event receiver or telemetry system lock was lost.

Reference

1. Miller, R. B., Pioneer 10 and 11 Mission Support, in Technical Report 32-1526, Volume 16, *The Deep Space Network Progress Report*, pp. 15-21. Jet Propulsion Laboratory, Pasadena, Calif., August 15, 1973.

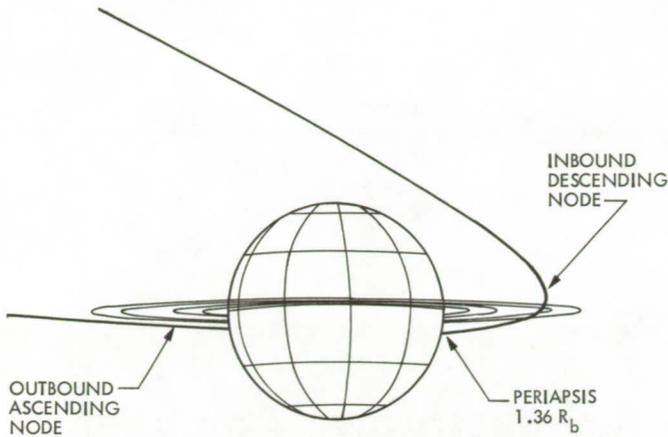
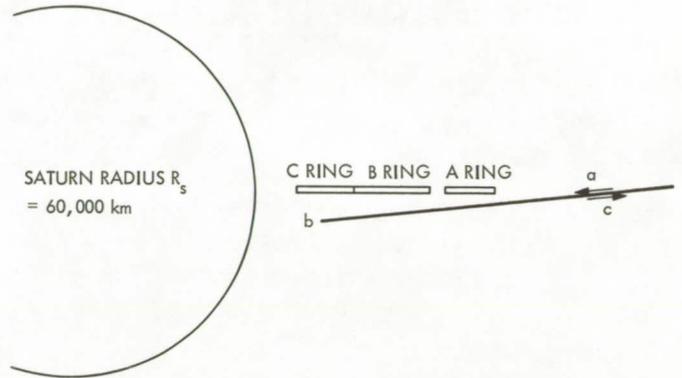


Fig. 1. Pioneer 11 Saturn flyby – balanced ring plane crossings at $2.87 R_s$, view from Earth



a. INBOUND RING CROSSING	E - 2 hr	$2.87 R_s^*$
b. SATURN CLOSEST APPROACH	E - 0 hr	$1.36 R_s$
c. OUTBOUND RING CROSSING	E + 2 hr	$2.87 R_s$

*(SAME AS VOYAGER JSX)

Fig. 3. Pioneer 11 Saturn trajectory in rotating equatorial coordinates

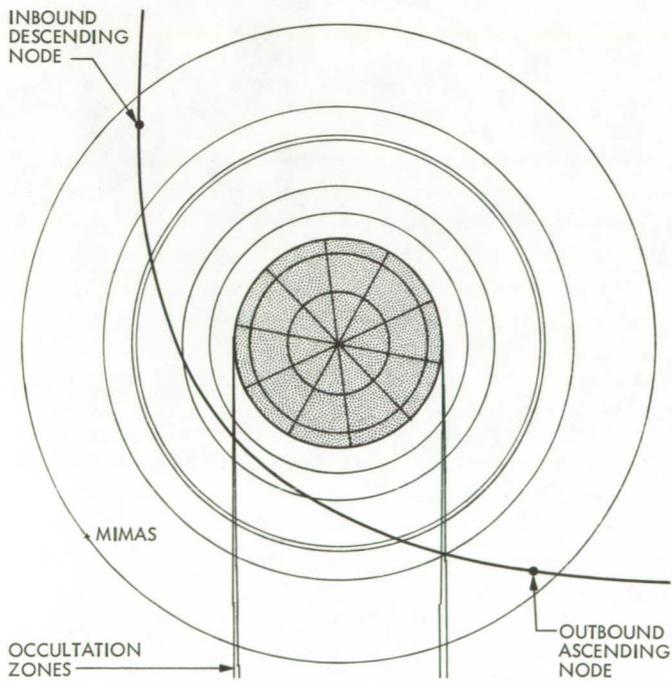


Fig. 2. Pioneer 11 Saturn flyby – balanced ring plane crossings at $2.87 R_s$, north polar view

The Gyrotron: A High-Frequency Microwave Amplifier

A. Kupiszewski

Radio Frequency and Microwave Subsystems Section

The purpose of this article is to introduce a proposed microwave amplifier mechanism for future generations of millimeter high-power uplinks to spacecraft and planetary radar transmitters. Basic electron-electromagnetic field interaction theory for RF power gain is explained, and the starting point for general analytical methods leading to detailed design results is presented.

I. Introduction

The next generation of spacecraft uplinks and planetary radar transmitters will probably operate at higher frequencies in the 20- to 35-GHz range. Present crowding at S- and X-bands as well as increased resolution and ranging available at higher frequencies make such a shift desirable. However, a cursory examination of readily available microwave amplifiers shows that there are no conventional high-power (> 10 kW) tubes useful for such purposes. Recently, much effort has been directed towards development of these higher frequency amplifiers. One such device of high potential is the gyrotron, a cyclotron resonance maser utilizing magnetic coupling between cavity or waveguide fields close to cutoff and relativistic electrons gyrating in large dc magnetic fields. Its basic advantage over klystrons and TWTs is that the electronic interaction volume does not scale down with wave length, hence high power can be obtained without the conventional heat transfer problems inherent in small interactive volumes.

The history of the gyrotron is in itself quite interesting. Theory for the device was developed as far back as 1958-59 by R. Q. Twiss and J. Schneider (Ref. 1), but power output levels were then measured in milliwatts. Interest in the U.S.

lagged, and funding was shifted elsewhere. In the U.S.S.R., however, research continued unabated, and some impressive results were achieved, for example 22 kW CW at 2 mm (Ref. 2). Due to such success, attention in the U.S. has again focused on this class of amplifier with contemplated applications including plasma heating and mm-wave radar.

II. Description

Gyrotrons may be operated over a wide range of frequencies from 2 to 235 GHz, depending only on selection of cavity and magnetic field strength. Figure 1 shows a typical arrangement for a gyrotron. Electrons from an annular cathode biased at several thousand volts form a hollow beam. (Hollow, circular beams are used in the interest of efficiency, since there is no RF component on the axis of symmetry. Breakup is inhibited by the short drift space and relatively high voltages involved, Ref. 3). The beam is then compressed by an axisymmetric magnetic field according to the adiabatic invariant $v_{\perp}^2/B = \text{constant}$, where v_{\perp} is the Larmor orbit velocity and B is the magnetic field strength. This is a magnetic mirror effect similar to that used to confine plasmas, where perpendicular, azimuthal electron energy is increased at the expense

of longitudinal drift energy. In the interaction space, which has a gently varying cross section to maximize efficiency, the electrons are guided by highly uniform fields. Magnetic coupling to the RF cavity fields is responsible for the microwave amplification, and, after completing interaction, the electrons settle on an extended collector surface in a region of weaker magnetic field.

In these respects, the gyrotron is just like klystrons and TWTs. An electron beam provides energy for microwave amplification. Spent electrons impact on an extended, water-cooled collector surface, which is the bulkiest, heaviest part of the tube. There the similarity ends. Instead of longitudinal bunching for microwave amplification, the gyrotron uses cyclotron resonance and azimuthal bunching.

One of the most important characteristics of gyrotrons is the high efficiencies theoretically obtainable. With proper gun design and good electron optics, the axial symmetry favors circumstances where all electrons interact with the RF fields under identical conditions. High efficiencies are contingent upon such relative uniformity. Losses in the magnetic mirrors and the interaction space are minimized due to solenoid design, which ideally reduces magnetic field variations to within a few tenths of one percent of maximum field strength. An example of a high-efficiency tube design with all of the aforementioned characteristics is the gyrotron TWT designed by Chu, Drobot, Granatstein, and Sefor (Ref. 4). The important parameters given are an output power of 342.5 kW CW with 51.0 percent calculated efficiency for a beam voltage and current of 70.82 kV and 9.48 A, respectively, and a magnetic field strength of 17.87 kG. The output frequency is 35 GHz with a maximum gain of 20 dB and a -3 dB bandwidth of 910 MHz.

III. Interaction Mechanism

To gain an understanding of the interaction mechanism, including theoretical results for bandwidth and efficiency, it is necessary to employ classical calculations based on Vlasov theory. The scope of such an endeavor, however, exceeds the aims of this introductory report, so only an overview of the process for arriving at a condition for amplification in a cylindrical cavity gyrotron will be given. More details will be addressed in future articles.

Starting with the Vlasov Equation (a continuity equation relating spacial changes in the electron distribution function to variations in time), one has the following:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f + \frac{q}{m} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \nabla_v f = 0 \quad (1)$$

where

f is the electron distribution function

\mathbf{v} is the velocity vector

q is the charge on the electron

\mathbf{E} is the electric field

\mathbf{B} is the magnetic field

m is the electron mass

The distribution function and the cavity field can be split into

$$f = f_0(\mathbf{x}, \mathbf{v}) + f_1(\mathbf{x}, \mathbf{v}, t) \quad (2)$$

$$\mathbf{E} = \mathbf{E}_1 \quad (\text{the electrons have already been accelerated before entering the cavity})$$

$$\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_1 \quad (\mathbf{B}_0 \text{ is the dc magnetic field}) \quad (3)$$

Using perturbation theory and linearizing to first order, the expression for the perturbed distribution is obtained.

$$f_1(\mathbf{x}, \mathbf{v}, t) = - \int_{-\infty}^t \frac{q}{m} (\mathbf{E}_1 + \mathbf{v} \times \mathbf{B}_1) \cdot \nabla_v f_0 dt' \quad (4)$$

And, for the TE_{011} mode in cylindrical geometry,

$$\begin{aligned} \mathbf{B}_1 = & \hat{\mathbf{r}} E_0 \frac{k_{\parallel}}{\omega} J_1(k_{\perp} r) \cos k_{\parallel} z \sin \omega t \\ & - \hat{\mathbf{z}} E_0 \frac{k_{\perp}}{\omega} J_0(k_{\perp} r) \sin k_{\parallel} z \sin \omega t \end{aligned} \quad (5)$$

$$\mathbf{E}_1 = \hat{\theta} E_0 J_1(k_{\perp} r) \sin k_{\parallel} z \cos \omega t \quad (6)$$

The time averaged power absorbed by the electrons is given by

$$\langle P_{abs} \rangle_{time} = \langle Nq \int \bar{d}\mathbf{v} \mathbf{v} f_1 \cdot \mathbf{E}_1 \rangle_{time} \quad (7)$$

Further integration yields the following:

$$\langle P_{abs} \rangle_{time} = \frac{q^2 E_0^2 \pi \rho(k_{\perp} a) v_{\perp}^3}{2m k_{\parallel}^2 v_{\parallel}^3} \cdot \frac{2k_{\parallel} v_{\parallel}}{\omega} G_{\omega}(v_{\parallel}, v_{\perp}) \cdot [\beta + Q_{\omega}(X)] \quad (8)$$

where E_0 is the RF field amplitude

Ω is the cyclotron frequency in the rest frame

ω is the frequency in radians of the cavity field

$N(r, \theta)$ is the electron density

c is the speed of light in vacuum

v_{\parallel} is the drift velocity, k_{\parallel} is the parallel wave number

v_{\perp} is the Larmor orbit velocity, k_{\perp} is the perpendicular wave number

a is the cavity radius

$$\rho(k_{\perp} a) \equiv \int_0^{2\pi} d\theta \int_0^a r dr N(r, \theta) J_1^2(k_{\perp} r) \quad (9)$$

$$Q_{\omega}(X) = X - \frac{1}{2G} \frac{\partial G}{\partial X} \left[\frac{\Omega^2}{k_{\parallel}^2 c^2} - X^2 \right] \quad (10)$$

$$G_{\omega}(X) = \text{lineshape function} = \frac{\cos^2 \frac{\pi X}{2}}{(1 - X^2)^2} \quad (11)$$

$$X = \frac{\omega - \Omega}{k_{\parallel} v_{\parallel}} \quad (12)$$

$$\beta = \frac{\Omega v_{\parallel} c}{k_{\parallel} c v_{\perp}^2} \quad (13)$$

By inspection, it can be seen that the condition for amplification is

$$-\beta > Q_{\omega}(X) \quad (14)$$

Although the process outlined above gives detailed, correct results for the lineshape function, it is easier to understand the amplification mechanism if it is viewed quantum mechanically. Power absorbed by the electrons traversing the cavity can be expressed (Refs. 5 and 6) as

$$P = N \hbar (\omega_{n,n+1} \rho_{n,n+1} - \omega_{n,n-1} \rho_{n,n-1}) \quad (15)$$

where N is the number of electrons

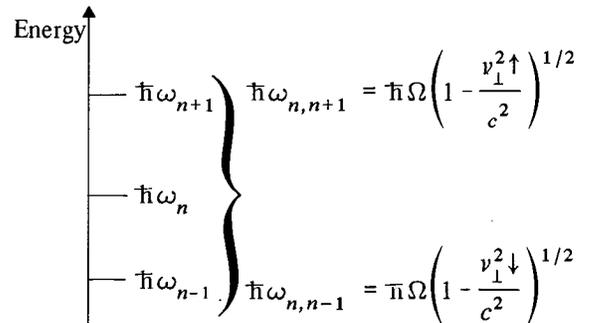
\hbar is Plank's constant divided by 2π

$\omega_{n,n+1}$ is the frequency in radians of an upward transition from level n to $n+1$

$\omega_{n,n-1}$ is the frequency in radians of a downward transition from level n to $n-1$

$\rho_{n,n+1}$ and $\rho_{n,n-1}$ are the respective transition probabilities

For free-free, higher-order transitions (i.e., n very large), it follows that $\rho_{n,n+1} \approx \rho_{n,n-1}$. Amplification, i.e., $P < 0$, occurs only if $\omega_{n,n-1} > \omega_{n,n+1}$. With even a mildly relativistic electron beam (~ 10 kV), such a condition can be easily achieved due to Lorentz corrections to the cyclotron frequency. Looking at the energy levels, one finds that differences can be expressed as Lorentz corrected percentages of the cyclotron frequency.



The lower energy levels have a smaller v_{\parallel} than the higher levels, which produce an unequal energy spacing between levels. This unequal spacing accounts for the emission exceeding absorption. It should be noted, however, that pursuit of this line of reasoning leads to Lorentzian lineshapes and not $G_{\omega}(X)$, which is confirmed by experiments (Ref. 7). Hence, present work in this field is based almost exclusively on classical calculations utilizing relativistic corrections.

IV. Summary

Gyrotrons are cyclotron resonance masers that show great promise in becoming the next generation of deep space transmitters due to high efficiencies at K-band frequencies with high-power outputs. Similarities to klystrons enhance the basic appeal of gyrotrons. Handling and maintenance technolo-

gies should not differ much except in the area concerning cryogenically-cooled superconducting magnets. Further research continues to produce impressive results with respect to power and efficiency at ever higher frequencies. Future reports on this subject will address new advances, efficiency calculations, and potential integration problems with existing systems.

References

1. Hirshfield, J. L. and Granatstein, V. L., "The Electron Cyclotron Maser – An Historical Survey," *IEEE MTT-25*, No. 6, June 1977, p. 522.
2. Flyagin, Gaponov, Petelin, and Yulpatov, "The Gyrotron," *IEEE MTT-25*, No. 6, June 1977, p. 514.
3. Kyhl, R. L. and Webster, H. F., "Breakup of Hollow Cylindrical Electron Beams," *IRE Transactions on Electron Devices*, October 1956, p. 172.
4. Chu, Drobot, Granatstein, and Seftor, "Characteristics and Optimum Operating Parameters of a Gyrotron Traveling Wave Amplifier," *IEEE MTT-27*, No. 2, February 1979, p. 178.
5. Schneider, Jürgen, "Stimulated Emission of Radiation by Relativistic Electrons in a Magnetic Field," *Phys. Rev. Let.*, 2, 504, 1959.
6. Schneider, Jürgen, NP-7916 *Duke Microwave Laboratory Report No. 26 for May 1, 1959, through August 1, 1959*, Duke University 1959 UC-2.
7. Hirshfield, Bernstein, and Wachtel, "Cyclotron Resonance Interaction of Microwaves with Energetic Electrons," *IEEE QE-1*, No. 6, September 1965, p. 237.

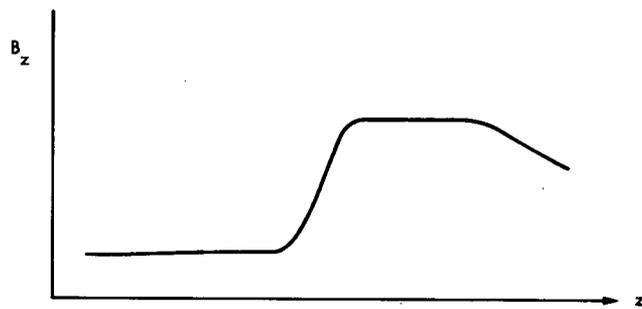
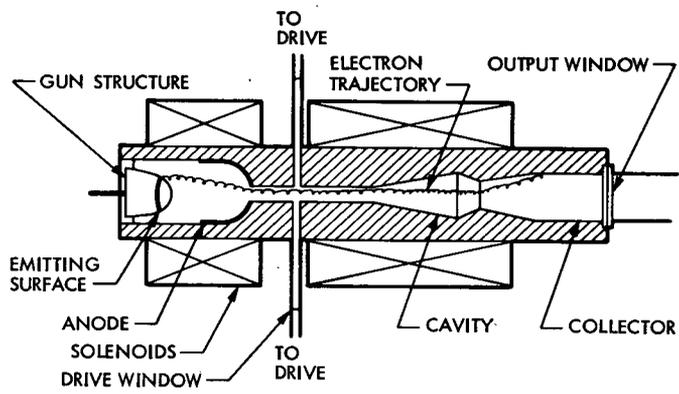


Fig. 1. Gyrotron amplifier

Connection and Validation of Narrow-Band Δ VLBI Phase Observations

S. C. Wu

Tracking Systems and Applications Section

Two-station narrow-band Δ VLBI requires phase connections between consecutive scans. This article presents an efficient computer-aided scheme for this purpose. This scheme is an iteration process alternating between a grand fit on many scans and integer quantization of the phase-shift cycles to be assigned to the scans. Only linear simultaneous equations of a few unknowns need to be solved. A simulation analysis indicates that faultless phase connection can be expected when there is no localized systematic noise. When systematic noise of moderate level exists, the possible incorrect phase connection can be detected and corrected for by comparing the connected phases from the two alternating observations, after removing the residual diurnal effects.

I. Introduction

Narrow-band Δ VLBI (Differential Very Long Baseline Interferometry) has been considered as one of the possible accurate spacecraft tracking systems. A demonstration plan has been made for Voyager spacecraft during Jupiter encounters in 1979 (Ref. 1). As is well known, narrow-band Δ VLBI requires a long pass (3-4 hr) of tracking so that its diurnal signature can resolve, in two orthogonal directions, the angular separation between spacecraft (S/C) and an angularly nearby extragalactic radio source (EGRS). On the other hand, with 2-station mode, which is most likely the case, S/C and EGRS observations are taken alternately and gaps in each data stream are inevitable. The resulting interferometric phase of each continuous data segment (scan) has a $2n\pi$ ambiguity with different n for different segments. To recover the correct diurnal signature of each data stream, the interferometric phases of the corresponding scans have to be connected with correct $2n\pi$ phase shift assigned to each scan.

Such phase connection, if done manually by "eyeballing" the variations of the phases, is not only time consuming and tedious but may result in incorrect integer n since the phase change across the gap may be as large as tens of cycles.

This article presents an efficient computer-aided phase-connection scheme. A simulation analysis shows that the scheme works faultlessly when there is no systematic noise such as clock drift or localized media effects. When such localized fluctuation exists, the scheme may assign incorrect integer cycle adjustments to some scans. This can easily be detected and corrected for by a phase validation scheme.

It should be pointed out that for a successful phase connection, either manually or by a computer, the effect of the earth's spin should be modeled out and only the "residual phases" need be connected.

II. The Grand Fit

When manually connecting the phases, one would fit each scan by a straight line and then try to connect two consecutive scans at one time by shifting one of them up or down by $2n\pi$ such that they look continuous. This appears no problem only when the two fitted straight lines have the same slopes; ambiguity arises when the slopes are different, as may happen in practice. Higher-order fit may seem to resolve this problem. However, an independent higher-order fit on each scan is prone to error due to a limited number of noise-affected data points in a scan. Consequently, a grand fit on as many scans as possible simultaneously is desirable for a computer-aided phase-connection scheme.

It is well known that most, if not all, fitting (optimization) algorithms require the solved-for parameters to be continuous over the region of interest. However, the integers n of concern are discrete numbers. A grand fit with such constraint imposed¹ becomes nonlinear and annoying. The present scheme does the grand fit free from such constraint and allows the parameters n to assume any value. An iterative process is then performed to successively quantize these values of n into integers. The iterative process will be discussed in Section III.

Let the measured residual phase of the i th data point in the j th scan be denoted as $\phi_{i,j}$; let there be J scans to be included in a single grand fit and I_j data points in the j th scan. Then for a grand fit of order M one needs to minimize the following function with respect to a_m and x_j :

$$f = \sum_{j=1}^J \sum_{i=1}^{I_j} \left\{ \sum_{m=0}^M a_m t_{i,j}^{m+r} - (\phi_{i,j} + 2\pi x_j) \right\}^2 \quad (1)$$

where $t_{i,j}$ is the time. The symbol x_j has been used in place of n to denote its role as a continuous variable.

Since only relative phases are needed we can arbitrarily set $x_1 = 0$. Then (1) is replaced by

$$f = \sum_{i=1}^{I_1} \left(\sum_{m=0}^M a_m t_{i,1}^m - \phi_{i,1} \right)^2 + \sum_{j=2}^J \sum_{i=1}^{I_j} \left(\sum_{m=0}^M a_m t_{i,j}^m - \phi_{i,j} - 2\pi x_j \right)^2 \quad (2)$$

To minimize (2) its partial derivatives with respect to a_0, a_1, \dots, a_M and x_2, x_3, \dots, x_J are set to zero. With some manipulations these become

$$\sum_{m=0}^M \sum_{j=1}^J \left(\sum_{i=1}^{I_j} t_{i,j}^{m+r} \right) a_m - 2\pi \sum_{j=2}^J \left(\sum_{i=1}^{I_j} t_{i,j}^r \right) x_j = \sum_{j=1}^J \sum_{i=1}^{I_j} t_{i,j}^r \phi_{i,j} \quad (3)$$

$$r = 0, 1, \dots, M$$

and

$$\sum_{m=0}^M \left(\sum_{i=1}^{I_j} t_{i,j}^m \right) a_m - (2\pi I_j) x_j = \sum_{i=1}^{I_j} \phi_{i,j}, \quad j = 2, 3, \dots, J \quad (4)$$

Equations (3) and (4) are $M + J$ simultaneous linear equations for the $M + J$ unknowns a_m and x_j . A further reduction of this system of linear equations can be done by substituting each x_j from (4) into (3). This yields

$$\sum_{m=0}^M \left\{ \sum_{i=1}^{I_1} t_{i,1}^{m+r} + \sum_{j=2}^J \left[\sum_{i=1}^{I_j} t_{i,j}^{m+r} - \left(\sum_{i=1}^{I_j} t_{i,j}^m \right) \left(\sum_{i=1}^{I_j} t_{i,j}^r \right) / I_j \right] \right\} a_m = \sum_{i=1}^{I_1} t_{i,1}^r \phi_{i,1} + \sum_{j=2}^J \left[\sum_{i=1}^{I_j} t_{i,j}^r \phi_{i,j} - \left(\sum_{i=1}^{I_j} t_{i,j}^r \right) \left(\sum_{i=1}^{I_j} \phi_{i,j} \right) / I_j \right] \quad (5)$$

$$r = 0, 1, \dots, M$$

Hence, a system of $M + 1$ linear equations for a_m is resulted. This can be written in the following matrix form:

$$[s] \{a\} = \{b\} \quad (6)$$

¹For instance, a penalty function $(1 - \cos^{2m} \pi x)$ can be added onto the function to be minimized (cf. Eq. 2).

where $[]$ denotes a square matrix and $\{ \}$ a column matrix. The elements of $[s]$ and $\{b\}$ are

$$S_{r,m} = S_{m,r} = \sum_{i=1}^{I_1} t_{i,1}^{m+r} + \sum_{j=2}^J \left[\sum_{i=1}^{I_j} t_{i,j}^{m+r} - \left(\sum_{i=1}^{I_j} t_{i,j}^m \right) \left(\sum_{i=1}^{I_j} t_{i,j}^r \right) / I_j \right] \quad (7)$$

and

$$b_r = \sum_{i=1}^{I_1} t_{i,1}^r \phi_{i,1} + \sum_{j=2}^J \left[\sum_{i=1}^{I_j} t_{i,j}^r \phi_{i,j} - \left(\sum_{i=1}^{I_j} t_{i,j}^r \right) \left(\sum_{i=1}^{I_j} \phi_{i,j} \right) / I_j \right] \quad (8)$$

In practice, $M = 2$ or 3 will suffice and the solutions of a_m are straightforward. Once the coefficients a_m have been determined, the phase shift parameters x_j are calculated according to (4):

$$x_j = \frac{1}{2\pi I_j} \left[\sum_{m=0}^M \left(\sum_{i=1}^{I_j} t_{i,j}^m \right) a_m - \sum_{i=1}^{I_j} \phi_{i,j} \right], j = 2, 3, \dots, J. \quad (9)$$

III. Iteration Process of Integer Quantization for x_j

The phase-shift parameters x_j determined by a grand fit in the preceding section will in general differ from integers. If the differences are all small, one may conceivably set them equal to the integers they approximate. However, if the differences are not all small, one may hesitate to equate x_j to an integer. However, the following iteration bears out the method.

From (1) it is obvious that any variation in x_j from correct integers induces corresponding changes in the fitted coefficients a_m . Let the change in each a_m be σ_m , $m = 0, 1, \dots, M$. Then the variation in x_j can be expressed, according to (9), as

$$\sigma_j = \frac{1}{2\pi I_j} \sum_{m=0}^M \left(\sum_{i=1}^{I_j} t_{i,j}^m \right) \sigma_m \quad (10)$$

This implies that the variation x_j from the correct integers increases with $t_{i,j}$. In other words, the first few x_j are much closer to the correct integers than the remaining x_j . Hence, one can comfortably set the first few x_j to the nearest integers. With these integers fixed, another grand fit is performed to improve a_m and the remaining x_j . The first few improved x_j will now be closer to the correct integers and can be quantized with greater confidence. This process is repeated until all x_j are quantized into integers.

After each grand fit, the criterion of setting the first x_j to the nearby integer can be more relaxed, as this x_j is more likely to be very nearly an integer. On the other hand it is more probable for the remaining x_j 's to miss the correct integer values by an amount greater than $1/2$ and the criterion should be made more stringent. In the proposed scheme the allowable variation is chosen to be 0.4 for the first x_j after each grand fit; for each of the remaining x_j , both x_j and x_{j-1} must have a variation < 0.1 .

IV. Simulation Analysis

The flow chart of the phase-connection scheme is shown in Fig. 1. The order of the polynomial to be fitted is assigned. In most cases a second- or third-order polynomial will be sufficient. The simulated phase data are generated with the following parameters:

data density = 1 per minute

gap width = W = an integer multiple of a minute

number of data points per scan = I , a variable parameter

data noise = σ , a variable parameter

pass length = 3 hours

number of scans = $J = (3 \times 60)/(W + I)$ rounded to the next lowest integer

$a_1 = 0.5$ rad/s

$a_2 = 0, -5 \times 10^{-5}, -1 \times 10^{-4}$ rad/s²

$a_3 = 0, 5 \times 10^{-9}$ rad/s³

$\phi_{i,j} = \sum_{m=1}^3 a_m t_{i,j}^m + \sigma - 2n_j\pi$ with n_j to be such that

$0 \leq \phi_{1,j} < 2\pi$.

Both $M = 2$ and $M = 3$ are tried for the fit. The following inferences are drawn from the simulation:

- (1) While higher order M for the polynomial to be fitted may be chosen even when $a_2 = a_3 = 0$ in the simulated

phase data, a lower-order polynomial is less susceptible to data noise. Hence one may start with a higher-order polynomial and decrease the order upon failure in phase connection.

- (2) As few as 2 data points per scan are allowable for a successful phase connection provided the gap width and the data noise are reasonably small.
- (3) Given a data density and a gap-width-to-data-span ratio the success in phase connection degrades as data noise σ increases but seems independent of the gap width, at least for gap width ≤ 10 minutes. This relaxes the gap width limit for minimum data loss when time offset between S/C and EGRS observations is called for (Ref. 2).
- (4) The maximum data noise above which phase connection may fail decreases with increasing gap-width-to-data-span ratio. Figure 2 is an example with 5-minute data span (per scan) over three hours. When there is no localized fluctuation due to systematic noise, a typical Δ VLBI pass² will fall well within the convergence range, as shown in the figure.

V. Grouping of Scans in a Single Grand Fit

As pointed out in Section II, it is desirable to include as many scans as possible in a single grand fit. However, a problem arises when sizable localized phase fluctuations exist among scans included in a single grand fit: A low-order polynomial may lose track of these fluctuations while a polynomial of higher order may degrade the convergence of the iteration process. Hence an appropriate grouping of scans in a single fit is essential.

A criterion to determine the grouping of scans is the change in slope of the straight lines fitted to the consecutive scans. Two consecutive scans are to be grouped together if the slopes of their fitted straight lines satisfy

$$|S_j - S_{j-1}| - (\sigma_j^2 + \sigma_{j-1}^2)^{1/2} \lesssim 1 \text{ mHz} \quad (11)$$

where S_j and σ_j are the slope and its uncertainty of scan j . The number of scans J in a group increases until (11) fails. When this number of scans J is determined, the order of polynomial is selected to be

$$\begin{aligned} M &= J, & \text{if } J \leq 5 \\ M &= 5, & \text{if } J > 5 \end{aligned} \quad (12)$$

²S/N = 5 corresponding to $\sigma = 0.2$ rad, and gap width/data span = 1.2.

The order of fit M is to be decreased by 1 upon failure of convergence in the iteration and a new fit is tried. If the iteration does not converge for all $M \geq 2$ the last scan in the group is removed from the fit, the largest M according to (12) is selected and the process repeats.

It is obvious that phase connection is redundant for a group with fewer than two scans. When the number of scans reduces to 2, due to the failure of either (11) or in the iteration process, the slope test of (11) is to be ignored. Also, the allowable deviation of the fitted x_j away from an integer is to be relaxed to 0.5 to exclude any failure in the iteration process. The actual deviation ϵ_{x_j} of each fitted x_j from its quantized value can serve as an indication of the degree of confidence in the resulting phase connection.

VI. Validation of Connected Phases

The phase-connection scheme discussed in the preceding sections works faultlessly when there is no localized phase fluctuations. Because of irregularities in transmission media (ionosphere in particular) and instability of electronic signal path, localized fluctuations do exist. When such fluctuations are sizable, not only do we need to divide the scans into several groups in connecting phases, but also some of the scans may be assigned incorrect cycle adjustments. To guard against such danger, the following phase validation process will be necessary.

Since the residual phases from S/C and EGRS are each independently connected, an approach of phase validation is to bring the two connected phases together and examine whether their variation is continuous over the pass. However, such examination is indicative only if the two residual phases have identical residual model errors. The residual model errors in frequency offset are identical since common frequency standards are used for both signals. Baseline error is not identical even though the same baseline is used, owing to different baseline projections onto the planes of sky in the two different directions of the two sources. Angular position (right ascension and declination) error is also different. Since both baseline error and angular position error appear in the form of residual diurnal variation, they can be easily removed by fitting $A \cos \omega_e t + B \sin \omega_e t$ to each of S/C and EGRS residual phases, where ω_e is the earth's spinning rate. In practice, two more terms, $Ct + D$, are needed to remove the residual frequency offset and phase offset. This is a simple linear least-squares fit with four degrees of freedom.³

³Alternatively, a simultaneous fit for S/C and EGRS phases with common frequency offset term can be adopted. The degrees of freedom become 7.

After removing the residual model errors, the S/C and EGRS phases are plotted on the same graph. This can be done handily by print-plotting on standard computer print. Now comes the time for inspection. Unfortunately, this has to be done by eyeballing at the present time. This is best illustrated by an example of the actual Δ VLBI pass. Figure 3 shows the connected phases of Voyager 1 and OJ 287 with their residual model errors removed. Two consecutive scans from the same source (S/C or quasar) are considered "continuous" if the phase change across the gap is consistent with the phase change in the intervening scan from the other source. A quick glance over the graph reveals that there are discontinuities between scans 2 and 3 of S/C phase and between the last two scans of quasar phase. When these two 1-cycle adjustments are made, the phases after removing the updated residual model errors are shown in Fig. 4. No obvious discontinuity remains and the adjustments are validated. Normally only one or two iterations of adjustment are needed to eliminate all offensive discontinuities.

There has been concern that when a clock without extremely high stability is used, phase connection may fail. By the above phase validation scheme, incorrectly connected phases can still be detected and corrected as long as the phase

drift due to clock instability is continuous (i.e., no phase jumps occur). Figure 5 is an example where rubidium clock was used at one end of the baseline. A discontinuity is detected between the last second and third scans of the S/C phase. After a 1-cycle adjustment the phases are shown in Fig. 6. Again, no obvious discontinuity remains. It should be noted that the gross effect of the clock drift has been absorbed by the fitting functions and only the residual "random walk" effect remains.

VII. Conclusions

A computer-aided, phase-connection scheme has been developed. A simulation analysis shows that, when there is no sizable localized phase fluctuation, this scheme connects VLBI residual phases faultlessly. When there are localized phase fluctuations the scheme may assign incorrect cycle adjustments to some of the scans. This can easily be detected and corrected for by a phase validation scheme. Currently this validation process requires an eyeballing inspection. A means to validate the connected phases automatically is currently being sought.

References

1. Brunn, D. L., et al., DSN Progress Report, 42-45, 15 June 1978, pp. 111-132.
2. Wu, S. C., IOM 315.1-296, June 12, 1978 (JPL internal document).

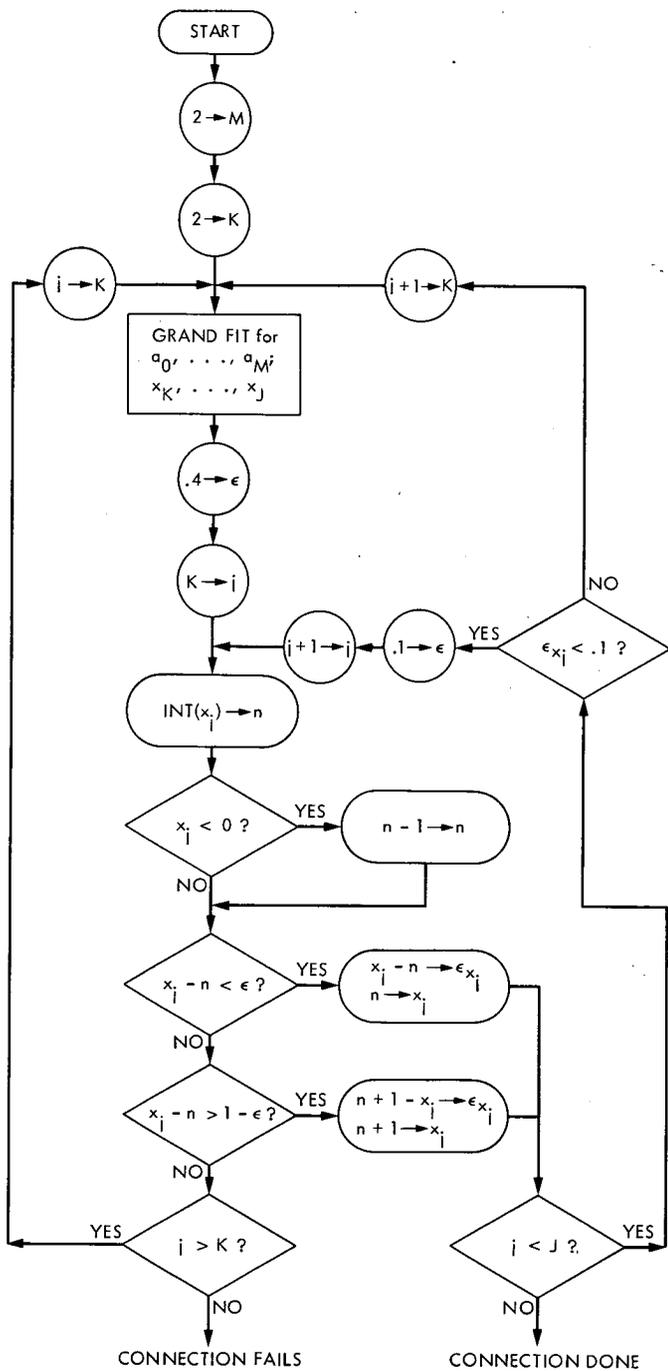


Fig. 1. Flow chart of phase-connection scheme

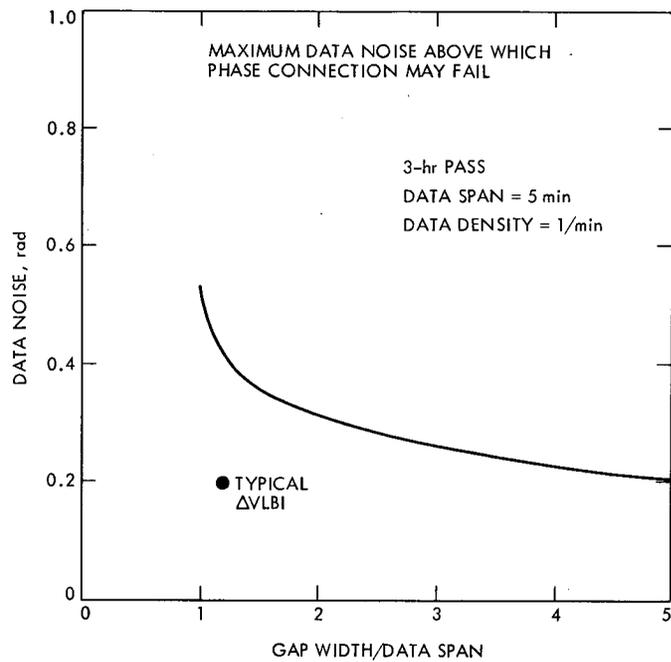


Fig. 2. Phase-connection threshold for a scheme in Fig. 1

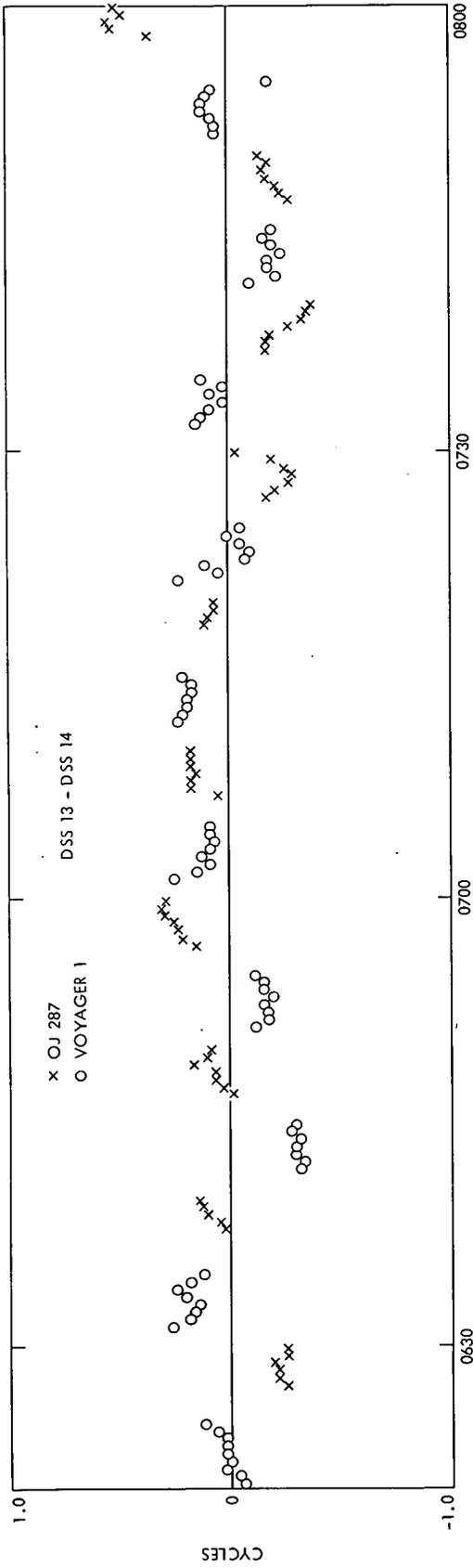


Fig. 3. Connected phases after removing constant rate and diurnal effects

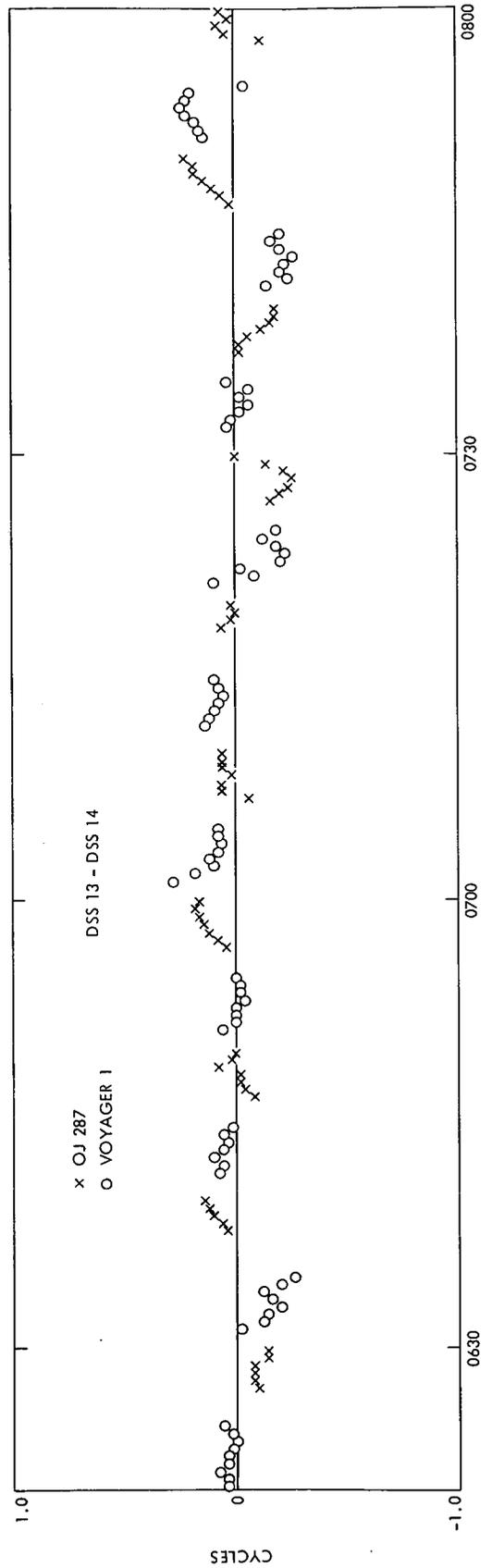


Fig. 4. Same as Fig. 3 except with cycle corrections

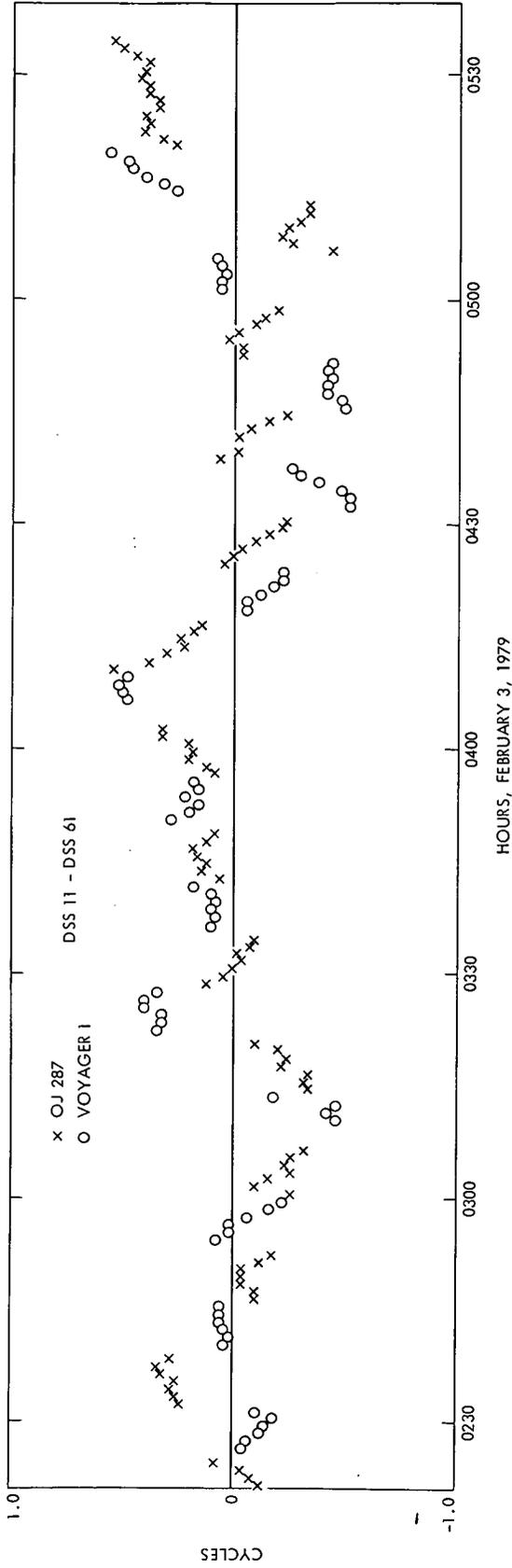


Fig. 5. Connected phases after removing constant phase rate and diurnal effects

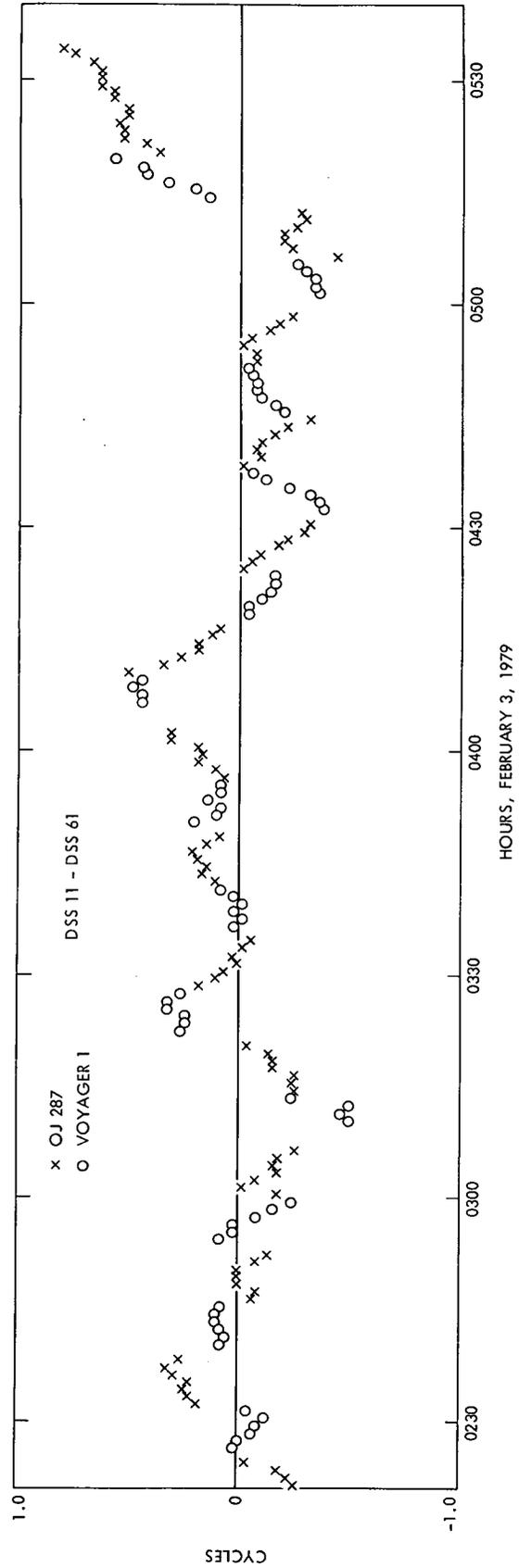


Fig. 6. Same as Fig. 3 except with cycle corrections

An Improved Storage Bulb Mount for DSN Hydrogen Masers

P. R. Dachel, D. P. Russell, T. K. Tucker, and L. B. Stratman
Communications Systems Research Section

This article compares the presently used JPL hydrogen maser suspended atomic storage bulb to a new, rigid, single-plane mounted bulb. The new bulb incorporates three major design changes:

- (1) *Mounting design.*
- (2) *Alterations to the collimator.*
- (3) *Decrease in mass.*

These design changes are expected to increase the long-term stability of the frequency standard by reducing its sensitivity to vibration and thermal effects.

I. Introduction

In June of 1976, an experiment to test the gravitational redshift using a hydrogen maser in a suborbital rocket probe was successfully performed jointly by the Smithsonian Astrophysical Observatory (SAO) and NASA's Marshall Space Flight center (Ref. 1). The design of the space maser was the result of a program begun under NASA sponsorship in 1962 (Ref. 2), and one of the results of this design effort was the single-plane mounted atomic hydrogen storage bulb. This mount consisted of a solid quartz collar with one rim fused to the quartz storage bulb and the other fixed to the RF cavity base plate. By lightening and refining this basic collar design, SAO designed the tetrahedral spoke support system that is used in the SAO VLG-11 today (Ref. 3).

The success of the solid-mounted SAO storage bulb design prompted JPL to test this concept in one of two prototype masers that was disassembled for repair. The rebuilt prototype masers (designated P₂ and P₃) will be used as reference

standards in the Frequency Timing System (FTS) test facility at JPL. The rigid mount bulb to be tested is shown in Fig. 1.

II. Two-Point Suspension Storage Bulb

In the present JPL Hydrogen Maser design, the storage bulbs are suspended in the microwave cavity by a two-point mounting system (Fig. 2). The upper support is a quartz rod fixed to the center line of the storage bulb. This rod extends through the cavity frequency adjustment plate into a spring-loaded mounting fixture. The spring pressure on the rod holds the storage bulb securely against the cavity base plate. The lower support is a Teflon¹ plug that is fixed to the cavity base plate and inserted into the neck of the storage bulb. The press fit of the Teflon plug into the neck of the storage bulb pro-

¹ Teflon is a Dupont name for tetrafluoroethylene fluorinated ethylene propylene copolymer.

vides support on a radial and transverse plane. This plug is center bored so that it also serves as a collimator for the storage bulb. One difficulty with this mounting scheme is that the bulb moves if the maser physics package receives a slight jolt. The magnitude of the displacement is determined by the clearance required in the slip joint of the spring mount, and it is this clearance that allows the storage bulb to rock slightly on the Teflon plug. Although the displacement of the storage bulb is small (0.0025 cm, 0.001 inches), it is sufficient to change the output frequency by approximately a part in 10^{-12} . Since it is practically impossible to avoid some form of physical shock at different times during maser operation, this shift poses a real problem.

The collimator consists of a center-bored Teflon plug with a 0.953-cm (0.375-inch) diameter orifice. The dimensions of the collimator are primary factors in determining the atomic storage time in the microwave cavity. The equation for storage time for the special case of a spherical storage bulb with a cylindrical collimator is:

$$T_b = \frac{2a^3 \ell}{\bar{v} b^3} \quad (1)$$

where:

T_b is the storage time

ℓ is the length of the collimator

b is the radius of the collimator

a is the radius of the storage bulb

\bar{v} is the mean velocity of the hydrogen atoms entering the storage bulb

Since the collimator is part of the storage area, it has an effect on the wall shift. In fact, since the collimator is a solid plug of Teflon, it has a different wall shift than the storage bulb, which is coated with Teflon. The equation for wall shift in a storage area coated with a single material (Ref. 4) is

$$\Delta\omega = \frac{\phi}{t_0} \quad (2)$$

where:

$\Delta\omega$ is the shift in line frequency

ϕ is the phase shift in atomic wave function per collision

t_0 is the mean time between collisions

III. Rigid Single Plane Mount Bulb

Comparing Figs. 2 and 3, it can be seen that with the removal of the support rod, there is no direct connection between the frequency adjustment plate and the rigid single-plane mount bulb. All support is through the base plate. Also, it can be seen in Fig. 3 that the collimator is no longer used for support. Freed from this duty, the collimator can be made a permanent part of the storage bulb. This improves the characteristics of the wall shift since the entire storage bulb substrate is quartz, coated with a continuous film of Teflon. Due to the rigid mounting design, resettability of the maser is improved since there is no chance of changing 1) the alignment of the collimator with respect to the storage bulb, and 2) the quartz (dielectric) position within the cavity electromagnetic field.

Following the SAO design, the collimator has been lengthened and its inside diameter decreased to 0.559 cm (0.220 inch). From Eq. (1), it is seen that this will increase the storage time by a factor of five, approximately. This increase in storage time will manifest itself as an increase in atomic line Q . From the equation for cavity pulling (Ref. 4):

$$\Delta\omega = \Delta\omega_c \frac{Q_c}{Q_\ell} \quad (3)$$

where:

$\Delta\omega$ is the shift in the output frequency

$\Delta\omega_c$ is the shift in the cavity resonant frequency

Q_ℓ is the atomic line Q

Q_c is the cavity Q

it is seen that the higher the line Q , the smaller the resulting shift in the master output frequency (i.e., the output frequency is more resistant to physical changes in the cavity).

Preliminary test results have shown a factor of 2 increase in line Q using this new bulb configuration.

One additional modification has been incorporated into the rigid mount bulb. The two-point suspension storage bulbs weigh between 280 and 300 grams. Small thermal fluctuations (0.001°C) present in the maser can change the dielectric constant of this mass. Since the storage bulb resides in a region of high electromagnetic fields in the resonant cavity, these changes can affect the output frequency of the maser. By reducing the amount of quartz to 230 grams, it is hoped to reduce this frequency shift.

IV. Conclusion

The P₃ prototype maser is now under construction with the rigid storage bulb mount. The following modifications are necessary to fit the new bulb to the JPL maser:

- (1) The material of the cavity bottom plate has been changed from aluminum to quartz (this modification
- (2) The thickness of the base plate has been increased.
- (3) A base support groove has been added to the quartz plate (Section A-1, Fig. 3).

also improves the thermal properties of the microwave cavity by decreasing the coefficient of thermal expansion of the base plate).

References

1. Vessot, R. F. C., and Levine, M. W. "A Test of the Equivalence Principle Using a Space-Borne Clock," *General Relativity and Gravitation*, Vol. 10, No. 3 (1979), pp. 181-204.
2. Vessot, R. F. C., Levin, M. W., Mattison, E. M., Hoffman, T. E., Imbier, E. A., Tetu, M., and Nystrom, G. "Space-Borne Hydrogen Maser Design," *Proceedings of the Eighth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*: November 30-December 2, 1976, U.S. Naval Research Laboratory, Washington, D.C. (X-814-77-149), pp. 277-333.
3. Levine, M. W., Vessot, R. F. C., Mattison, E. M., Blomberg, E., Hoffman, T. E., Nystrom, G., Graveline, D. F., Nicoll, R. F., Dovidio, C., and Brymer, W. "A Hydrogen Maser Design for Ground Applications," *Proceedings of the Eighth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*: November 30-December 2, 1976, U.S. Naval Research Laboratory, Washington, D.C. (X-814-77-149), pp. 249-276.
4. Kleppner, D.; Goldenberg, H. M., Ramsey, N. F. "Theory of the Hydrogen Maser," *Physical Review*, Vol. 126, No. 2, p. 603, 1962.

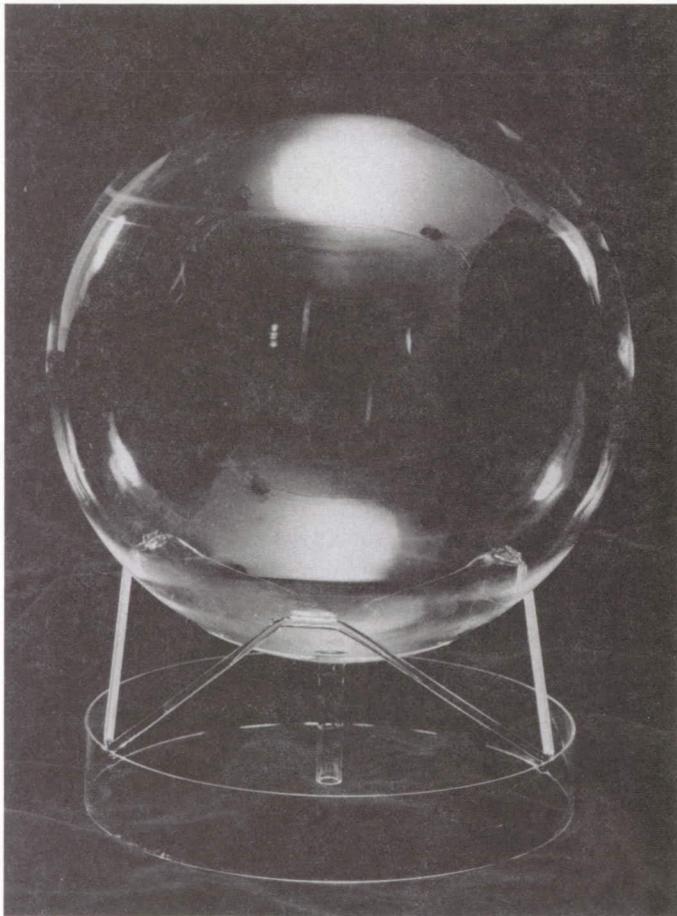


Fig. 1. JPL hydrogen maser

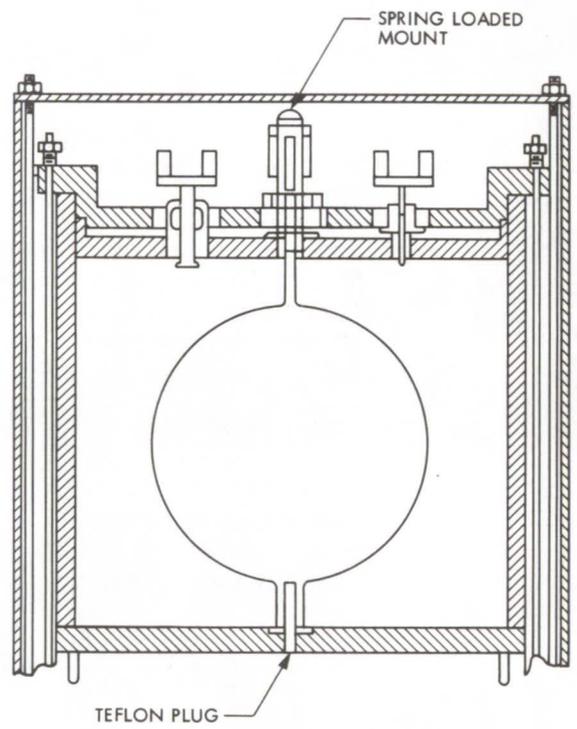
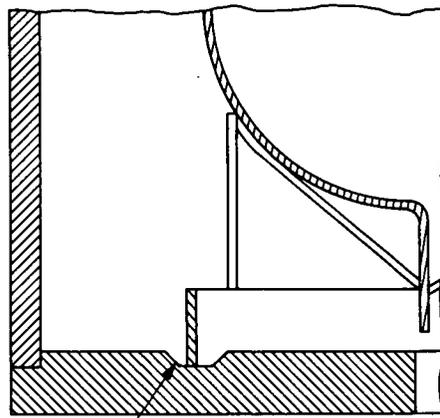
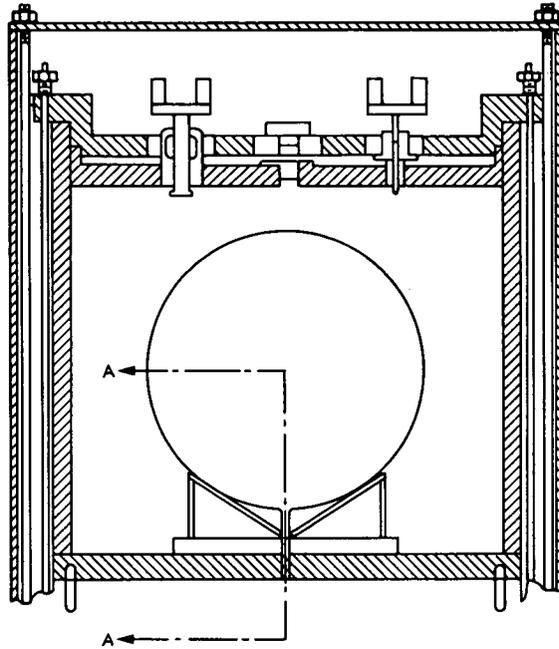


Fig. 2. Two-point suspension



SUPPORT GROOVE
SECTION A-A

Fig. 3. Rigid mount

A Generalization of Binary Minimum Shift Keying and Staggered Quadrphase Shift Keying Modulation

J. L. Massey

Consultant to
Communications Systems Research Section

A generalized modulation scheme, which includes minimum shift keying (MSK) and staggered quadrphase shift keying (SQPSK) as special cases, is analyzed. The general modulator can be realized as a one-input, two-output sequence transducer whose outputs select the carrier signal for each baud. This form of the modulator has the practical advantage of not requiring any RF filtering since there is no actual mixing of the carriers with the modulating signals.

It is shown that the optimum demodulator (whether hard-decision or soft-decision) always can make its decisions from the received waveform over two bauds when the interference is additive white Gaussian noise, thus generalizing a well-known result for hard-decision demodulation of MSK and SQPSK signals. The power spectra of MSK and SQPSK signals are derived to isolate the role played by coherency between the modulating signals and the carriers.

I. Introduction

In this report, we study a generalized modulation scheme that includes minimum shift keying (MSK) and staggered quadrphase shift keying (SQPSK) modulation as special cases. The generalized modulator, whose form has certain practically advantageous features, is introduced in Section II as a specific one-input, two-output sequence transducer whose outputs select the carrier signal for each baud. The resulting modulated signal is described by means of a waveform "trellis" associated with the "trellis" of the sequence transducer. The particular choice of carrier signals to obtain SQPSK and MSK modulation are specified.

In Section III, we use the waveform trellis of the modulator in a "magic genie" argument to prove that the optimum demodulator, whether hard-decision or soft-decision, can restrict its observations to two consecutive bauds when the interference is additive white Gaussian noise. This generalizes a well-known fact about MSK and QPSK modulation.

In Section IV, we derive the power spectra for both MSK and SQPSK signals, paying special attention to the required relationships between the carriers and the modulating signals. Finally, in Section V, we make some concluding remarks and give some suggestions for future research.

II. The Modulator

A. Sequence Transducer and Trellis

Let $\mathbf{d} = \cdots d_{-1}, d_0, d_1, d_2, \cdots$ be the *data sequence* that constitutes the modulator input, where each d_i has value +1 (corresponding to a binary 0) or -1 (corresponding to a binary 1). The heart of our modulator structure will be the *sequence transducer*, shown in Fig. 1, whose output sequences x and y are given by

$$x_i = \frac{1}{2} (d_i + d_{i-1}) \quad (1a)$$

and

$$y_i = \frac{1}{2} a_i (d_i - d_{i-1}) \quad (1b)$$

for $-\infty < i < +\infty$, where \mathbf{a} is the *alternating sequence* defined by

$$a_i = (-1)^{i+1} \quad (2)$$

From Eqs. (1) and (2), we see that the components of the output sequences x and y take values in the set $\{-1, 0, +1\}$, i.e., these sequences are ternary-valued. Moreover, for each i , either $x_i = 0$ (corresponding to $d_i \neq d_{i-1}$) or $y_i = 0$ (corresponding to $d_i = d_{i-1}$), but not both. Thus x_i and y_i can (and soon will) be used to amplitude-moderate the “quadrature components” of a carrier in such a way that one and only one component will be present for each i .

A convenient way to display the input/output relationship of the sequence transducer of Fig. 1 is by means of the *trellis* shown in Fig. 2. The nodes of this trellis at depth i correspond to the possible states of the sequence transducer at time instant i . From Eq. (1), we see that this *state* can be chosen as the value of d_{i-1} ; the nodes in Fig. 2 have been labeled with this choice of state. The branches leaving each state at depth i correspond to the possible state transitions, the upper branch being the transition when the input d_i equals -1 and the lower branch being the transition when the input d_i equals +1; the branches are labeled with the value of (x_i, y_i) corresponding to the output pair for that transition. We see, for instance, that if the transducer is in state +1 at time instant 0, then the input $d_0 = -1$ will cause a transition to state -1 at time instant 1 and the accompanying outputs will be $(x_0, y_0) = (0, +1)$; similarly, the input $d_1 = -1$ will also cause a transition from state +1 at time 1 to state -1 at time 2, but the accompanying outputs will be $(x_1, y_1) = (0, -1)$. Thus, each path through the entire trellis corresponds to a particular input sequence \mathbf{d} , and the

labels on this path specify the resulting output sequences x and y .

The trellis of Fig. 1 is, of course, closely akin to the “trellis” introduced by Forney to represent convolutional codes (Ref. 1). The only essential difference is that the “sections” between the nodes at each depth are identical in convolutional code trellises because the encoder is a time-invariant transducer. The sequence transducer of Fig. 1, however, is *time-varying* because of the effect of the alternating sequence \mathbf{a} , which has period 2. In fact, the trellis of Fig. 2 is *periodic* with period 2 in the sense that the section beginning at time instant i is the same as that beginning at time instant $i + 2$ for all i .

We remark that the sequence transducer shown in Fig. 3 is equivalent to that in Fig. 1, as is easily seen from the facts that $d'_i = a_i d_i$ and that $d'_{i-1} = a_{i-1} d_{i-1} = -a_i d_{i-1}$, since $a_i = -a_{i-1}$. Note that if \mathbf{d} is a *random data sequence*, i.e., if the random variables d_i are statistically independent and identically distributed (i.i.d.) with $P(d_i = +1) = P(d_i = -1) = 1/2$, then \mathbf{d}' is also a random data sequence. From Fig. 3, we can see that the sequences x and y will have identical statistics in this case.

B. Modulator and Trellis

We are now ready to introduce our proposed modulator, whose structure is shown in Fig. 4. Here, $c_x(t)$ and $c_y(t)$ are *carrier waveforms* that will be specified later for specific modulation schemes. The *baud length* will be denoted as T . The function of the “RF selector switch” in Fig. 4 is to select one of the carrier inputs for transmission in each baud according to the following rule:

$$s(t) = \begin{cases} +c_x(t), & \text{if } x_i = +1 \\ -c_x(t), & \text{if } x_i = -1 \\ +c_y(t), & \text{if } y_i = +1 \\ -c_y(t), & \text{if } y_i = -1 \end{cases}, iT \leq t < iT + T \quad (3)$$

We emphasize that no addition of RF signals and *no RF filtering is required* in the modulator, since there is no actual mixing of the carriers with modulating signals.

Because of Eq. (3), we see that the trellis of Fig. 2 can be modified, as shown in Fig. 5, to show the input/output structure of the modulator of Fig. 4. The trellis in Fig. 5 differs from that in Fig. 4 only in that the node depth is labeled by the time iT instead of by the time instant i , and that the transitions are now labeled with the value of $s(t)$ in

the baud $iT \leq t < (i+1)T$ rather than with (x_p, y_p) . Thus, each path through the entire trellis of Fig. 5 still corresponds to a particular input data sequence \mathbf{d} , but the labels on the path now specify the resulting *modulated signal* $s(t)$ from the modulator of Fig. 4.

C. MSK Operation

We show now that the modulator of Fig. 4 realizes binary *minimum shift keying* (MSK) modulation (Ref. 2) when the carriers are selected in the manner

$$c_x(t) = A \sin \left[\left(\omega_0 + \frac{\Delta\omega}{2} \right) t + \theta \right] \quad (4a)$$

and

$$c_y(t) = A \sin \left[\left(\omega_0 - \frac{\Delta\omega}{2} \right) t + \theta \right] \quad (4b)$$

where A and θ are an arbitrary amplitude and arbitrary phase, respectively, where ω_0 is the carrier center frequency, and where

$$(\Delta\omega)T = \pi \quad (5)$$

Since, in each baud, the modulated signal $s(t)$ will be one of $+c_x(t)$, $-c_x(t)$, $+c_y(t)$ and $-c_y(t)$, it suffices to show that the phase of $s(t)$ is continuous at the transitions between bauds (i.e., at times $t = iT$ for all i and for all choices of the data sequence \mathbf{d}).

From the trellis of Fig. 5, we see that the phase of $s(t)$ is certainly continuous at $t = iT$ if $s(t)$ does not switch between the two carriers at time iT , since then $s(t)$ will be the same one of $+c_x(t)$, $-c_x(t)$, $+c_y(t)$, or $-c_y(t)$ in the two bauds adjacent at time iT . It remains to show that the phase is continuous when $s(t)$ switches from either $+c_x(t)$ or $-c_x(t)$ to either $+c_y(t)$ or $-c_y(t)$, or from either $+c_y(t)$ or $-c_y(t)$ to either $+c_x(t)$ or $-c_x(t)$, at $t = iT$.

Suppose first that i is even. We see then from Fig. 5 that the only such carrier-switching transitions possible at $t = iT$ for $s(t)$ are:

- (i) from $-c_x(t)$ to $-c_y(t)$,
- (ii) from $+c_x(t)$ to $+c_y(t)$,
- (iii) from $-c_y(t)$ to $-c_x(t)$, or
- (iv) from $+c_y(t)$ to $+c_x(t)$.

But we see from Eqs. (4) and (5) that the phase difference at time $t = iT$ between $c_x(t)$ and $c_y(t)$ is $(\Delta\omega)iT = i\pi$, which is a multiple of 2π when i is even; thus the phase of $s(t)$ is continuous for all four of the above transitions.

Suppose on the other hand that i is odd. We see now from Fig. 5 that the only carrier-switching transitions that are possible at $t = iT$ for $s(t)$ are:

- (i) from $-c_x(t)$ to $+c_y(t)$,
- (ii) from $+c_x(t)$ to $-c_y(t)$,
- (iii) from $-c_y(t)$ to $+c_x(t)$, or
- (iv) from $+c_y(t)$ to $-c_x(t)$.

But, from Eqs. (4) and (5), we see that the phase difference at time $t = iT$ between $c_x(t)$ and $-c_y(t)$ is $(\Delta\omega)iT + \pi = (i+1)\pi$, which is a multiple of 2π when i is odd; thus again the phase of $s(t)$ is continuous for all four possible transitions. This completes the demonstration that the modulation is indeed binary MSK.

More precisely, the above argument shows that the modulated signal $s(t)$ is the same as for MSK modulation. However, in what is commonly considered to be MSK modulation, the data sequence controls the modulated waveform in the manner that when the i^{th} data digit is $+1$ or -1 , then the modulated signal in the i^{th} baud is $\pm A \sin [(\omega_0 + \Delta\omega/2)t + \theta]$ or $\pm A \sin [(\omega_0 - \Delta\omega/2)t + \theta]$, respectively, when the appropriate sign is chosen to maintain phase continuity. In other words, the data digit directly controls the baud frequency. We see from Eq. (1a), however, that x_i (which controls the sinusoid of frequency $\omega_0 + \Delta\omega/2$) is nonzero if and only if $d_i = d_{i-1}$; and that y_i (which controls the sinusoid of frequency $\omega_0 - \Delta\omega/2$) is nonzero if and only if $d_i \neq d_{i-1}$. But the i^{th} bit in the first difference of the binary data stream is 0 (corresponding to $+1$) when $d_i = d_{i-1}$ and is 1 (corresponding to -1) when $d_i \neq d_{i-1}$. Thus, the modulator of Fig. 4 actually realizes *differential* MSK since the differences in the data sequence form the sequence that actually controls the baud frequency. However, we shall continue to say simply "MSK modulation," rather than the more precise "differential MSK modulation," to describe the modulation performed by the modulator of Fig. 4 when the carriers are specified as in Eq. (6).

D. SQPSK Operation

We show now that the modulator of Fig. 4 realizes *staggered-quadrature shift-keying* (SQPSK) modulation (Ref. 3), also called offset-keyed quadrature shift-keying (OKQPSK), when the carriers are selected in the manner

$$c_x(t) = A \cos(\omega_0 t + \theta) \quad (6a)$$

and

$$c_y(t) = A \sin(\omega_0 t + \theta) \quad (6b)$$

where again A and θ are an arbitrary amplitude and arbitrary phase, respectively.

By definition, SQPSK is four-phase modulation in which the phase can change by either 0 or $\pm\pi/2$ between bauds, but never by $\pm\pi$. (This is usually accomplished by modifying an ordinary QPSK modulator for a baud of length $2T$ in which two independent data sequences binary-antipodally modulate the quadrature carrier components so that one data sequence makes its transitions mid-way between the other's transitions – the advantage over ordinary QPSK is a reduction of cochannel interference due to nonideal effects in demodulation.) But this restricted phase-changing is obviously achieved by the modulator of Fig. 4 since, as we see from Fig. 5, $s(t)$ can never make a transition between $+c_x(t)$ and $-c_x(t)$ or between $+c_y(t)$ and $-c_y(t)$, which are the only transitions according to Eq. (6) for which the phase of $s(t)$ would change by $\pm\pi$.

Again, we remark that we have really shown here only that the modulated signal $s(t)$ is the same as for SQPSK modulation. Just as was the case for MSK modulation, however, when the correspondence between the data sequence and the modulated signal is considered, one finds that the modulator of Fig. 4 actually realizes *differential* SQPSK.

III. The Demodulator

We now consider demodulation for the modulator of Fig. 4 when the received signal is

$$r(t) = s(t) + n(t) \quad (7)$$

where $n(t)$ is additive white Gaussian noise (AWGN). We will show the rather remarkable fact that the hard-decision demodulator shown in Fig. 6 is *optimum* (in the sense of minimizing the *probability of error* in the decision \hat{d}_i for d_i , for all i) *regardless of the choice* of $c_x(t)$ and $c_y(t)$ (i.e., these two carriers need have no special orthogonality properties in each baud) provided only that, for every i , $c_x(t)$ and $c_y(t)$ have the *same energy* in the baud $iT \leq t < iT + T$ (but the energy could depend on i) and that \mathbf{d} is a *random data sequence*. Henceforth, we assume that this energy condition and data condition are satisfied.

To demonstrate the optimality of the hard-decision demodulator of Fig. 6, we exploit the “magic genie” approach of Wozencraft and Jacobs (Ref. 4, p. 419). Suppose we wish to estimate d_i where i is even. Suppose further that the genie is kind enough to tell us both the state, σ_i , of the modulator at time instant i and also the state, σ_{i+2} , at time instant $i+2$. If the genie says $\sigma_i = +1$ and $\sigma_{i+2} = +1$, for instance, we see from the trellis of Fig. 5 that $d_i = +1$, would imply $s(t) = +c_x(t)$ for $iT \leq t < iT + 2T$, whereas $d_i = -1$ would imply $s(t) = +c_y(t)$ for $iT \leq t < iT + 2T$. Moreover, any permissible choice of $s(t)$, for $t < iT$ and for $t \geq iT + 2T$, when $d_i = +1$ is also permissible when $d_i = -1$, since the only requirement is that the state be $+1$ at time instant i and again $+1$ at time instant $i+2$. Thus, the decision problem for d_i (with genie's help) reduces to deciding only whether $c_x(t)$ or $c_y(t)$ was transmitted in the interval $iT \leq t < iT + 2T$. But this is the classical problem of deciding between two equally-likely equal-energy signals in the presence of AWGN, and the well-known (Ref. 4, pp. 238-239) optimum decision rule is: Choose $\hat{d}_i = +1$ if and only if

$$\int_{iT}^{iT+2T} r(t)c_x(t) dt \geq \int_{iT}^{iT+2T} r(t)c_y(t) dt \quad (8)$$

Defining

$$X_i = \int_{iT}^{iT+T} r(t)c_x(t) dt \quad (9a)$$

and

$$Y_i = \int_{iT}^{iT+T} r(t)c_y(t) dt \quad (9b)$$

we see that Eq. (8) can be written as

$$X_i + X_{i+1} \geq Y_i + Y_{i+1} \quad (10)$$

But now suppose instead that the genie had told us that $\sigma_i = -1$ and $\sigma_{i+2} = +1$. Recalling that i is even, we see from the trellis of Fig. 5 that $d_i = +1$ corresponds to $s(t) = -c_y(t)$ for $iT \leq t < iT + T$ and $s(t) = c_x(t)$ for $iT + T \leq t < iT + 2T$. Similarly, $d_i = -1$ corresponds to $s(t) = -c_x(t)$ for $iT \leq t < iT + T$ and $s(t) = c_y(t)$ for $iT + T \leq t < iT + 2T$. Thus, the optimum (genie-aided) decision rule is: Choose $\hat{d}_i = +1$ if and only if

$$-\int_{iT}^{iT+T} r(t)c_y(t) dt + \int_{iT+T}^{iT+2T} r(t)c_x(t) dt \geq \int_{iT}^{iT+T} r(t)c_x(t) dt + \int_{iT+T}^{iT+2T} r(t)c_y(t) dt$$

$$\geq - \int_{iT}^{iT+T} r(t)c_x(t) dt + \int_{iT+T}^{iT+2T} r(t)c_y(t) dt \quad (11)$$

Using Eq. (9), we can write Eq. (11) as

$$-Y_i + X_{i+1} \geq -X_i + Y_{i+1} \quad (12)$$

which we see is precisely the same condition as Eq. (10)!

Similar analyses for the case $\sigma_i = +1$ and $\sigma_{i+2} = -1$ and for the case $\sigma_i = -1$, $\sigma_{i+2} = -1$ show that the optimum genie-aided decision rules are again: Choose $\hat{d}_i = +1$ if and only if Eq. (10) is satisfied. But the four cases considered exhaust the possible values for σ_i and σ_{i+2} . We conclude that we can *exorcise the genie*; we have no use for his information since the optimum decision rule for d_i , when i is even, is independent of his information and is just: Choose $\hat{d}_i = +1$ if and only if Eq. (10) is satisfied. Note that this is precisely the rule for the demodulator in Fig. 6 since, for i even, $-a_i = +1$ so that $A_i = X_i + X_{i+1}$ and $B_i = Y_i + Y_{i+1}$, where A_i and B_i are defined as

$$A_i = X_i + X_{i+1} \quad (13a)$$

$$B_i = -a_i(Y_i + Y_{i+1}) \quad (13b)$$

Next, suppose i is odd and the genie tells us that $\sigma_i = +1$ and $\sigma_{i+2} = +1$. From Fig. 5, we see that $d_i = +1$ corresponds to $s(t) = c_x(t)$ for $iT \leq t < iT + 2T$, whereas $d_i = -1$ corresponds to $s(t) = -c_y(t)$ for $iT \leq t < iT + 2T$. The same argument as before now shows the optimum genie-aided decision rule to be: Choose $\hat{d}_i = +1$ if and only if

$$X_i + X_{i+1} \geq -Y_i - Y_{i+1}$$

Analysis of the three remaining choices of σ_i and σ_{i+2} shows that this same decision rule is also optimum. Thus, the genie can again be exorcised. We note further that this optimum decision rule is the rule for the demodulator in Fig. 6 since, for i odd, $-a_i = -1$ so that $A_i = X_i + X_{i+1}$ and $B_i = -Y_i - Y_{i+1}$.

This completes our proof that the hard-decision demodulator in Fig. 6 is optimum in the sense of minimizing $\Pr(\hat{d}_i \neq d_i)$ when the noise is AWGN, when \mathbf{d} is a random data sequence and, when for each i , the carriers $c_x(t)$ and $c_y(t)$ have the same energy in the baud $iT \leq t < iT + T$. Note that the optimum decision \hat{d}_i for d_i depends only on $r(t)$ over the two-baud interval $iT \leq t < iT + 2T$; this is a well-known fact for both binary MSK and SQPSK, but we have now demonstrated that this "two-baud optimality" is independent of

whether $c_x(t)$ and $c_y(t)$ are orthogonal over each baud as they are in both MSK and SQPSK.

We have argued elsewhere (Refs. 5 and 6) that demodulators should be designed to maximize the *cut-off rate*, R_0 , of the discrete channel created by the modulator, waveform channel and demodulator, rather than to minimize bit error probability for a hard-decision demodulator. Thus, we find it much more satisfactory than showing that the demodulator of Fig. 6 is the optimum hard-decision demodulator to show that:

When $A_i - B_i$ is taken as the output, the demodulator of Fig. 6 is optimum in the sense of maximizing the cut-off rate, R_0 , of the discrete channel between the modulator input and demodulator output (and also in the sense of maximizing the capacity, C , of this channel).

To prove this claim, we must show that the demodulator with output $A_i - B_i$ *preserves the likelihood ratio* for the decision on d_i , since any operation on $r(t)$ reduces R_0 (and also C) unless and only unless this likelihood ratio is preserved (Ref. 5).

We begin by letting \mathbf{r} , \mathbf{s} and \mathbf{n} be the vector representations of $r(t)$, $s(t)$ and $n(t)$, respectively, in some appropriate Euclidean space ("signal space"). Given that \mathbf{r} is received and that a magic genie informs the receiver that $d_j = \delta_j$ for all $j \neq i$, the likelihood ratio for the decision on d_i is

$$\Lambda_i = \frac{p(\mathbf{r} | d_i = +1, d_j = \delta_j \text{ for } j \neq i)}{p(\mathbf{r} | d_i = -1, d_j = \delta_j \text{ for } j \neq i)} \quad (14)$$

where p is a conditional probability density function. Equation (14) can be rewritten as

$$\Lambda_i = \frac{p(\mathbf{r} | \mathbf{s}_{+1})}{p(\mathbf{r} | \mathbf{s}_{-1})} \quad (15)$$

where \mathbf{s}_{+1} and \mathbf{s}_{-1} are the signal space representations of the waveforms $s_{+1}(t)$ and $s_{-1}(t)$ assumed by $s(t)$ for the data sequence $d_j = \delta_j$ for $j \neq i$ with d_i equal to $+1$ and -1 , respectively. Since $\mathbf{r} = \mathbf{s} + \mathbf{n}$ and the noise $n(t)$ is AWGN, Eq. (15) becomes

$$\Lambda_i = \frac{e^{-\frac{1}{N_0} |\mathbf{r} - \mathbf{s}_{+1}|^2}}{e^{-\frac{1}{N_0} |\mathbf{r} - \mathbf{s}_{-1}|^2}} \quad (16)$$

where N_0 is the one-sided noise power spectral density. The assumption that, in each baud, $c_x(t)$ and $c_y(t)$ have the same energy implies that $s_{+1}(t)$ and $s_{-1}(t)$ have the same energy or, equivalently, $|s_{+1}|^2 = |s_{-1}|^2$. Thus, Eq. (16) becomes

$$\Lambda_i = e^{\frac{2}{N_0} [\mathbf{r} \cdot (s_{+1} - s_{-1})]} \quad (17)$$

By the correspondence of "dot product" in signal space to correlation of waveforms,

$$\mathbf{r} \cdot (s_{+1} - s_{-1}) = \int_{-\infty}^{+\infty} r(t) [s_{+1}(t) - s_{-1}(t)] dt \quad (18)$$

Because the data sequences yielding $s_{+1}(t)$ and $s_{-1}(t)$ differ only in the value of d_i , we see from the trellis of Fig. 5 that $s_{+1}(t) = s_{-1}(t)$ for $t < iT$ and $t \geq iT + 2T$ so that Eq. (18) can be written as

$$\mathbf{r} \cdot (s_{+1} - s_{-1}) = \int_{iT}^{iT+2T} r(t) [s_{+1}(t) - s_{-1}(t)] dt \quad (19)$$

It remains to evaluate Eq. (19). From Eqs. (2) and (9) and the trellis of Fig. 5, we find that, for all four possible values of d_{i-1} and d_{i+1} ,

$$\int_{iT}^{iT+T} r(t) [s_{+1}(t) - s_{-1}(t)] dt = X_i + a_i Y_i$$

and

$$\begin{aligned} \int_{iT+T}^{iT+2T} r(t) [s_{+1}(t) - s_{-1}(t)] dt &= X_{i+1} + a_{i+1} Y_{i+1} \\ &= X_{i+1} - a_i Y_{i+1} \end{aligned}$$

This, together with Eq. (13), implies that Eq. (19) can be written as

$$\mathbf{r} \cdot (s_{+1} - s_{-1}) = A_i - B_i \quad (20)$$

(which shows again that we can exorcise the genie, since A_i and B_i can be formed without the genie's help.) Finally, substitution of Eq. (20) into Eq. (17) gives

$$\Lambda_i = e^{\frac{2}{N_0} (A_i - B_i)} \quad (21)$$

which shows that the output $A_i - B_i$ does indeed preserve the value of the likelihood ratio, as was to be shown.

We note that, since the demodulator of Fig. 6 with $A_i - B_i$ taken as the output preserves the likelihood ratio (and hence *all* the statistical information about d_i), the *optimum* (in the sense of maximizing R_0 or C) *soft-decision modulator* with any *specified number of demodulator output levels* can be obtained simply by appropriately quantizing the output Z_i (cf. Refs. 5 and 7 for details of this procedure).

IV. Modulation Spectra

A. General Considerations

We now examine the spectral properties of the signal produced by the modulator of Fig. 4.

With the sequence \mathbf{x} of Eq. (1a), we associate the pulsed waveform $x(t)$ defined by

$$x(t) = x_i \quad iT \leq t < (i+1)T \quad (19)$$

In the same way, we associate pulsed waveforms $y(t)$, $d(t)$, and $a(t)$ with the sequences \mathbf{y} , \mathbf{d} , and \mathbf{a} , respectively. Note that $a(t)$ is a "square wave" of period $2T$.

For purposes of spectral analysis, we need now to randomize the time origin so that all time signals under consideration become stationary random processes. We accomplish this by replacing all time functions [e.g., $x(t)$ and $s(t)$] with their translations by t_0 seconds [e.g., $x(t+t_0)$ and $s(t+t_0)$] where t_0 is a random variable uniformly distributed over the interval $0 \leq t_0 < 2T$ and statistically independent of the data pulse values. We will write $R_x(\tau)$ and $R_s(\tau)$, for instance, to denote the resulting autocorrelation functions

$$\begin{aligned} R_x(\tau) &= E[x(t+t_0+\tau)x(t+t_0)] \\ &= E[x(t_0+\tau)x(t_0)] \end{aligned}$$

and

$$\begin{aligned} R_s(\tau) &= E[s(t+t_0+\tau)s(t+t_0)] \\ &= E[s(t_0+\tau)s(t_0)] \end{aligned}$$

respectively. It should be noted that, since *all* time signals are translated by t_0 seconds, this randomization of the time origin neither introduces nor removes any "coherence" between the modulating signals and the carriers.

Note that the data signal $d(t + t_0)$ is the familiar telegraph wave (with baud length T and amplitude ± 1). Thus, $R_d(\tau)$ is the triangular autocorrelation function shown in Fig. 7; and the power spectral density (the Fourier transform of $R_d(\tau)$) is the familiar

$$S_d(f) = T \operatorname{sinc}^2(fT) \quad (20)$$

where

$$\operatorname{sinc}(z) = \frac{\sin(\pi z)}{\pi z}$$

The spectrum $S_d(f)$ is also shown in Fig. 7.

Recalling Eqs. (1a) and (1b), we see that

$$x(t) = \frac{1}{2} [d(t) + d(t - T)] \quad (21a)$$

and

$$y(t) = \frac{1}{2} a(t) [d(t) - d(t - T)] \quad (21b)$$

From Eqs. (21a) and (21b), we find by a simple calculation that

$$\begin{aligned} R_x(\tau) &= R_y(\tau) = \frac{1}{4} R_z(\tau + T) + \frac{1}{2} R_z(\tau) + \frac{1}{4} R_z(\tau - T) \\ &= \frac{1}{2} R_d\left(\frac{\tau}{2}\right) \end{aligned} \quad (22)$$

which further implies

$$S_x(f) = S_y(f) = S_d(2f) = T \operatorname{sinc}^2(2fT) \quad (23)$$

These autocorrelation functions and spectra are shown in Fig. 8. Inasmuch as $x(t)$ and $y(t)$ are ternary-valued (0, +1 or -1), it is a bit surprising that their autocorrelation functions are of the same form as the binary-valued (± 1) data signal $d(t)$.

Recall from our discussion in Section II about the sequences x and y that, for each i , one of x_i and y_i will be 0 and the other will be +1 or -1. Thus, the random processes $x(t + t_0)$ and $y(t + t_0)$ are highly dependent. It is quite surprising then to find that *the random processes $x(t + t_0)$ and $y(t + t_0)$ are uncorrelated*, i.e., that

$$R_{xy}(\tau) = E[x(t + t_0 + \tau)y(t_0 + t)] \equiv 0 \quad (24)$$

as can be verified by a direct (but somewhat tedious) calculation from Eqs. (21a) and (21b).

We now turn to the main task at hand, finding the spectrum of the modulated signal, i.e., of the random process $s(t + t_0)$. From Eq. (3), we see that

$$s(t) = x(t)c_x(t) + y(t)c_y(t) \quad (25)$$

which is our starting point. To proceed further, we must consider specific carrier signals. One general note of caution is in order, however, at this point. Because of Eq. (24), it might seem that $s(t)$ in Eq. (25) can be treated as the sum of two independently modulated signals. This is not true in general, however, because some types of coherence between $c_x(t)$ and $c_y(t)$ can cause $x(t + t_0)c_x(t + t_0)$ to be correlated with $y(t + t_0)c_y(t + t_0)$, even though $x(t + t_0)$ and $y(t + t_0)$ are uncorrelated. We shall soon see that this type of coherence between $c_x(t)$ and $c_y(t)$ does not occur with SQPSK modulation, but does occur with MSK modulation.

B. SQPSK Spectrum

As we saw in Subsection II.D., the signal $s(t)$ is an SQPSK modulated signal when the carriers are specified by Eqs. (6a) and (6b). From these equations and Eq. (25), we find in this case that

$$\begin{aligned} s(t + t_0) &= A x(t + t_0) \cos[\omega_0(t + t_0) + \theta] \\ &\quad + A y(t + t_0) \sin[\omega_0(t + t_0) + \theta] \end{aligned} \quad (26a)$$

It follows from Eq. (26a) that

$$\begin{aligned} R_s(\tau) &= E[s(t + t_0 + \tau)s(t + t_0)] \\ &= \frac{1}{2} A^2 R_x(\tau) \cos(\omega_0\tau) + \frac{1}{2} A^2 R_y(\tau) \cos(\omega_0\tau) \\ &\quad + A^2 R_{xy}(\tau) \sin(\omega_0\tau) \end{aligned} \quad (27)$$

(Where the vanishing of the double carrier frequency terms under expectation strictly speaking requires T to be an integer multiple of the carrier period $2\pi/\omega_0$, but is virtually the case whenever $\omega_0 \gg 1/T$.) By virtue of Eq. (24), we see that the term resulting from the cross-correlation of the two carriers vanishes (i.e., Eq. (26) represents independent modulation of orthogonal carriers) so that Eq. (27) becomes

$$R_s(\tau) = A^2 R_x(\tau) \cos(\omega_0 \tau) \quad (28)$$

where we have also made use of Eq. (22). From Eqs. (28) and (23), it follows that

$$S_s(f) = \frac{1}{2} A^2 T \{ \text{sinc}^2 [2(f - f_0) T] + \text{sinc}^2 [2(f + f_0) T] \} \quad (29)$$

where $f_0 = \omega_0/2\pi$ is the carrier frequency in hertz. Equation (29) gives, of course, the well-known spectrum of SQPSK, which coincides with that of ordinary QPSK. The purposes of our deriving this result here are to emphasize the role played by the vanishing of $R_{xy}(\tau)$ and to point out that, because the phase angle θ in Eq. (26) is arbitrary, no special phase relationship is required between the carriers and the modulating signals. The spectrum $S_s(f)$ of Eq. (29) is sketched in Fig. 9.

C. MSK Spectrum

When the carriers in Eq. (25) are specified as in Eqs. (4a) and (4b), $s(t + t_0)$ becomes the MSK signal

$$s(t + t_0) = A x(t + t_0) \sin \left[\left(\omega_0 + \frac{\Delta\omega}{2} \right) (t + t_0) + \theta \right] + A y(t + t_0) \sin \left[\left(\omega_0 - \frac{\Delta\omega}{2} \right) (t + t_0) + \theta \right] \quad (26b)$$

Applying familiar trigonometric identities, we can rewrite Eq. (26b) as

$$s(t + t_0) = A [x(t + t_0) + y(t + t_0)] \cdot \cos \left[\frac{\Delta\omega}{2} (t + t_0) \right] \sin [\omega_0 (t + t_0) + \theta] + A [x(t + t_0) - y(t + t_0)] \cdot \sin \left[\frac{\Delta\omega}{2} (t + t_0) \right] \cos [\omega_0 (t + t_0) + \theta] \quad (26c)$$

We are thus led to define the signals

$$d_1(t) = x(t) + y(t) \quad (30a)$$

and

$$d_2(t) = x(t) - y(t) \quad (30b)$$

With the aid of Eqs. (21a) and (21b), we see from Eqs. (30a) and (30b) that

$$d_1(t) = \begin{cases} d(t), & iT \leq t < (i+1)T \text{ and } i \text{ odd} \\ d(t-T), & iT \leq t < (i+1)T \text{ and } i \text{ even} \end{cases} \quad (30c)$$

and

$$d_2(t) = \begin{cases} d(t-T), & iT \leq t < (i+1)T \text{ and } i \text{ odd} \\ d(t), & iT \leq t < (i+1)T \text{ and } i \text{ even} \end{cases} \quad (30d)$$

In other words, $d_1(t)$ and $d_2(t)$ are *independent amplitude random data signals* of baud length $2T$, *staggered* so that the transitions in $d_1(t)$ occur at times iT with i odd whereas the transitions in $d_2(t)$ occur at times iT with i even. The data values (+1 or -1) in $d_1(t)$ correspond to those of $d(t)$ in the bauds $iT \leq t < (i+1)T$ with i odd, while the data values in $d_2(t)$ correspond to those of $d(t)$ in the bauds $iT \leq t < (i+1)T$ with i even. In particular, it follows that

$$R_{d_1}(\tau) = E[d_1(t + t_0 + \tau) d_1(t + t_0)] = R_{d_2}(\tau) = R_d(\tau/2) \quad (31)$$

Substituting Eqs. (30a) and (30b) into Eq. (26c), we find

$$s(t + t_0) = A d_1(t + t_0) \cos \left[\frac{\Delta\omega}{2} (t + t_0) \right] \sin [\omega_0 (t + t_0) + \theta] + A d_2(t + t_0) \sin \left[\frac{\Delta\omega}{2} (t + t_0) \right] \cos [\omega_0 (t + t_0) + \theta] \quad (32)$$

It follows from the statistical independence (and zero means) of the amplitudes of $d_1(t + t_0)$ and $d_2(t + t_0)$ that the cross-correlation between the two terms on the right in

Eq. (32) vanishes. Moreover, each of the two terms clearly has the same autocorrelation function so that

$$\begin{aligned} R_s(\tau) &= E[s(t+t_0+\tau)s(t+t_0)] \\ &= A^2 \cos(\omega_0\tau) R_m(\tau) \end{aligned} \quad (33)$$

where $R_m(\tau)$ is the autocorrelation function of the random process

$$m(t+t_0) = d_1(t+t_0) \cos\left[\frac{\Delta\omega}{2}(t+t_0)\right] \quad (34)$$

We now proceed to find

$$R_m(\tau) = E[m(t+t_0+\tau)m(t+t_0)] \quad (35)$$

$$= E[m(t_0+\tau)m(t_0)] \quad (36)$$

where we have also used the fact that $(\Delta\omega)T = \pi$. Taking Fourier transforms (recalling that $R_m(\tau) = R_m(-\tau)$) in Eq. (38) yields

$$\begin{aligned} S_m(f) &= \frac{T}{2} \left[\text{sinc}\left(2fT - \frac{1}{2}\right) + \text{sinc}\left(2fT + \frac{1}{2}\right) \right]^2 \\ &= \frac{T}{2} \left\{ \text{sinc}\left[2\left(f - \frac{\Delta f}{2}\right)T\right] + \text{sinc}\left[2\left(f + \frac{\Delta f}{2}\right)T\right] \right\}^2 \end{aligned} \quad (39)$$

where

$$\Delta f = \frac{\Delta\omega}{2\pi} = \frac{1}{2T} \quad (40)$$

Equation (39) is our desired expression for the spectrum of $m(t)$.

It is convenient to take first the conditional expectation in Eq. (36) for a fixed t_0 , which gives

$$\begin{aligned} E\{m(t_0+\tau)m(t_0) | t_0\} &= E\left\{d_1(t_0+\tau)d_1(t_0) \cos\left[\frac{\Delta\omega}{2}(t_0+\tau)\right] \cos\left[\frac{\Delta\omega}{2}t_0\right] | t_0\right\} \\ &= \begin{cases} \cos\left[\frac{\Delta\omega}{2}(t_0+\tau)\right] \cos\left[\frac{\Delta\omega}{2}t_0\right], & 0 \leq t_0 < T-\tau \text{ and } T \leq t_0 < 2T \\ 0, & T-\tau \leq t_0 < T \end{cases} \end{aligned} \quad (37)$$

Recalling that t_0 is uniformly distributed over $0 \leq t_0 < 2T$, we find that with the aid of Eq. (37) that

$$\begin{aligned} R_m(\tau) &= \frac{1}{2T} \int_0^{2T} E\{m(t_0+\tau)m(t_0) | t_0\} dt_0 \\ &= \begin{cases} \frac{1}{2} \left(1 - \frac{\tau}{2T}\right) \cos\left(\frac{\Delta\omega}{2}\tau\right) + \frac{1}{2\pi} \sin\left(\frac{\Delta\omega}{2}\tau\right), & 0 \leq \tau < 2T \\ 0, & \tau \geq 2T \end{cases} \end{aligned} \quad (38)$$

From Eqs. (33) and (39), we have finally

$$\begin{aligned}
S_s(f) = & \frac{1}{4} A^2 T \left\{ \text{sinc} \left[2 \left(f - f_0 - \frac{\Delta f}{2} \right) T \right] \right. \\
& + \text{sinc} \left[2 \left(f - f_0 + \frac{\Delta f}{2} \right) T \right] \left. \right\}^2 \\
& + \frac{1}{4} A^2 T \left\{ \text{sinc} \left[2 \left(f + f_0 - \frac{\Delta f}{2} \right) T \right] \right. \\
& + \text{sinc} \left[2 \left(f + f_0 + \frac{\Delta f}{2} \right) T \right] \left. \right\}^2
\end{aligned} \quad (41)$$

which is the well-known spectrum of MSK, but in a form perhaps more transparent than the usual one. This spectrum is sketched in Fig. 9 for comparison with that of SQPSK.

We remark here that the phase angle θ in Eq. (26c) is arbitrary. Thus, the MSK spectrum of Eq. (41) does not require any special phase relationship between the modulating signal and the carriers $c_x(t)$ and $c_y(t)$.

The MSK spectrum does, of course, depend on the coherency between the modulating signals, $x(t)$ and $y(t)$, and the carriers, $\sin[(\omega_0 + \Delta\omega/2)t + \theta]$ and $\sin[(\omega_0 - \Delta\omega/2)t + \theta]$, in Eq. (25). If there were no cross-correlation between the two terms on the right in Eq. (26b), then the autocorrelation function of $s(t + t_0)$ would be simply

$$\begin{aligned}
& \frac{1}{2} A^2 R_x(\tau) \cos \left[\left(\omega_0 + \frac{\Delta\omega}{2} \right) \tau \right] + \frac{1}{2} A^2 R_y(\tau) \cos \left[\left(\omega_0 - \frac{\Delta\omega}{2} \right) \tau \right] \\
& = \frac{1}{2} A^2 R_x(\tau) \left\{ \cos \left[\left(\omega_0 + \frac{\Delta\omega}{2} \right) \tau \right] + \cos \left[\left(\omega_0 - \frac{\Delta\omega}{2} \right) \tau \right] \right\}
\end{aligned}$$

and hence the power spectrum of $S(t + t_0)$ would be

$$\begin{aligned}
& \frac{1}{4} A^2 \left[S_x \left(f - f_0 - \frac{\Delta f}{2} \right) + S_x \left(f - f_0 + \frac{\Delta f}{2} \right) \right. \\
& + S_x \left(f + f_0 + \frac{\Delta f}{2} \right) + S_x \left(f + f_0 - \frac{\Delta f}{2} \right) \left. \right] \\
& = \frac{1}{4} A^2 T \left\{ \text{sinc}^2 \left[2 \left(f - f_0 - \frac{\Delta f}{2} \right) T \right] \right. \\
& + \text{sinc}^2 \left[2 \left(f - f_0 + \frac{\Delta f}{2} \right) T \right] \left. \right\}
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{4} A^2 T \left\{ \text{sinc}^2 \left[2 \left(f + f_0 + \frac{\Delta f}{2} \right) T \right] \right. \\
& + \text{sinc}^2 \left[2 \left(f + f_0 - \frac{\Delta f}{2} \right) T \right] \left. \right\}
\end{aligned} \quad (42)$$

Comparing Eqs. (41) and (42), we see that the cross-correlation between the two terms on the right in Eq. (27) has the effect of *changing the spectrum from a sum of the squares of two sinc functions* (as in Eq. (42)) to the square of the sum of the same two sinc functions. Moreover, outside the central lobe of $S_x(f)$ (i.e., outside the range $f_0 - 3/4T < f < f_0 + 3/4T$), one of the sinc functions in Eq. (41) is positive and the other is negative for all f , as follows from the fact that their arguments differ by $2(\Delta f) T = 1$. Thus, one of these sinc functions always tends to cancel the other outside the central lobe, which is the reason that the MSK spectrum drops off much more rapidly than that of SQPSK outside the central lobe.

We have already mentioned that the phase angle θ between the carriers and the modulating signals has no effect on the MSK spectrum. In fact, more is true. The phase angle θ in Eq. (26b) can even be time-varying, i.e., a random process $\theta(t + t_0)$, and the above analysis goes through unchanged provided only that $\theta(t)$ not change appreciably over one baud (say, by 0.1 radians or less.) This latter condition is required to insure the vanishing of the expectation of the double carrier frequency terms when computing $R_s(\tau)$.

V. Concluding Remarks

It is well-known (Ref. 8) that MSK modulation can be viewed as "offset QPSK" modulation in which the modulating signals are sinusoidal pulses in each band rather than rectangular pulses – this follows immediately from Eq. (32) above. Simon (Ref. 9) has offered a generalization of MSK, based on this fact, by considering more general pulse shaping.

By expanding on the left and using the usual trigonometric identities, one easily proves the identity

$$\left[\text{sinc} \left(Z - \frac{1}{2} \right) + \text{sinc} \left(Z + \frac{1}{2} \right) \right]^2 = \frac{1}{2\pi^2} \frac{1 + \cos(2\pi Z)}{\left(Z^2 - \frac{1}{4} \right)^2} \quad (43)$$

With the aid of Eq. (43), Eq. (39) can be rewritten as

$$S_m(f) = \frac{4T}{\pi^2} \frac{1 + \cos(4\pi f T)}{(16f^2 T^2 - 1)^2} \quad (44)$$

which is the usual baseband form of the spectrum of MSK and shows that the spectrum drops off as $1/f^4$ for large f (in contrast to that of QPSK which drops off only as $1/f^2$).

Our aim in this report has been to study MSK in particular from a new standpoint with the hope of isolating the fundamental structure that accounts for its spectral improvement over QPSK. The "sequence transducer" approach that we have taken appears amenable to new generalizations of MSK signalling by different choices of the sequence transducer and carriers. This seems to us to warrant further investigation.

Finally, we remark that several recent papers, e.g., (Refs. 10 and 11), have considered a "phase trellis" approach to MSK

and other forms of continuous phase modulation. (Our approach is closely related to these but somewhat more general since we can admit arbitrary "carrier signals" when converting the trellis of the sequential transducer to a "waveform trellis.") Anderson and Taylor (Ref. 11) in particular show substantial power improvements (2 to 4 dB) over QPSK together with a smaller bandwidth. However, their modulation scheme introduces "convolutional code type" constraints into the modulated signal so that demodulation over a fixed number of bauds is not optimal (as it is for MSK). At this writing, it is not clear whether these schemes actually exploit the potential "coding gain" of simpler modulation formats and thus cannot be further encoded efficiently, or whether in fact they actually utilize bandwidth more efficiently than simpler modulation schemes. It seems to us that this is an important question to settle.

Acknowledgment

The author is indebted to Dr. James R. Lesh for several helpful suggestions in the course of the research described here.

References

1. G. D. Forney, Jr., "The Viterbi Algorithm," *Proc. IEEE*, Vol. 61, pp 268-273, March 1973.
2. M. L. Doelz and E. H. Heald, *Minimum-Shift Data Communications*, U.S. Patent No. 2,977,417, March 28, 1961.
3. M. K. Simon and J. G. Smith, "Offset Quadrature Communications with Decision Feedback," *IEEE Trans. Comm.*, Vol. COM-22, pp. 1576-1584, Oct. 1974.
4. J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley 1965.
5. J. L. Massey, "Coding and Modulation in Digital Communications," in *Proc. 1974 Int. Zurich Seminar on Digital Comm.*, pp. E2(1)-E2(4), March 12-15, 1974.
6. J. L. Massey, "Comparison of Phase Modulation Systems," in *The Deep Space Network Progress Report 42-49*, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 1979, pp. 57-63.
7. L. Lee, "On Optimum Soft-Decision Modulation," *IEEE Trans. Info. Th.*, Vol. IT-22, pp. 437-444, July 1976.
8. S. A. Gronemeyer and A. L. McBride, "MSK and Offset QPSK Modulation," *IEEE Trans. Comm.*, Vol. COM-24, pp. 809-820, Aug. 1976.
9. M. K. Simon, "A Generalization of Minimum-Shift-Keying (MSK)-type Signalling Based Upon Input Data Symbol Pulse Shaping," *IEEE Trans. Comm.*, Vol. COM-24, pp. 845-856, Aug. 1976.
10. T. Aulin and C. E. Sundberg, *Binary CPFSK Type of Signalling with Input Data Symbol Pulse Shaping – Error Probability and Spectrum*, Technical Report TR-99, TTT, Dept. of Elec. Engr., Lund University, Sweden, April 1978.
11. J. B. Anderson and D. P. Taylor, "A Bandwidth-Efficient Class of Signal-Space Codes," *IEEE Trans. Info. Th.*, Vol. IT-24, pp. 703-712, Nov. 1978.

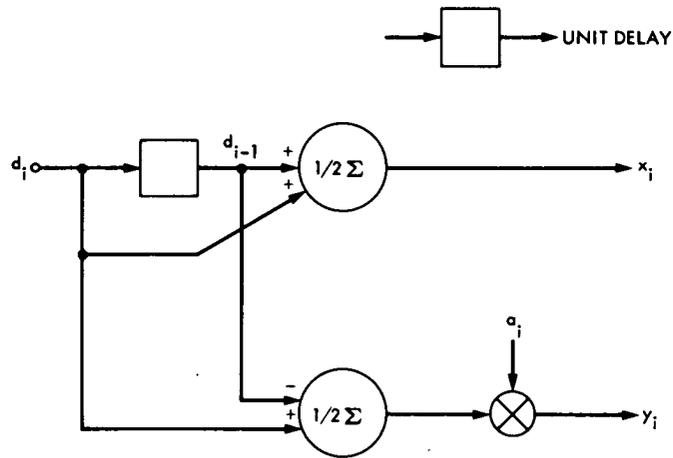


Fig. 1. The sequence transducer

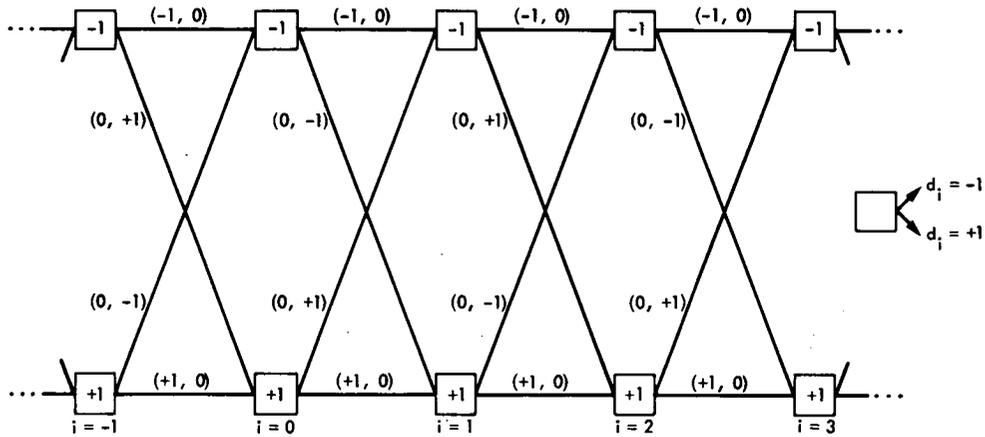


Fig. 2. Trellis for the sequence transducer of Fig. 1

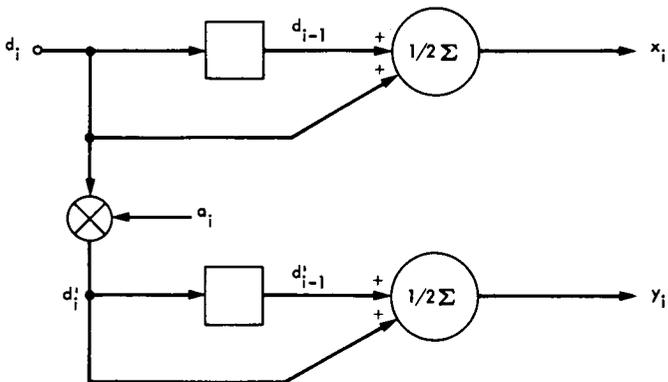
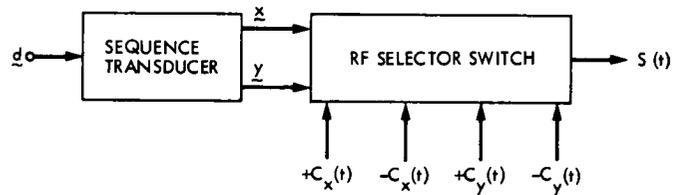


Fig. 3. An alternative sequence transducer equivalent to that in Fig. 1



NOTE: SEE EQUATION (3) FOR SELECTION RULES

Fig. 4. Proposed modulator

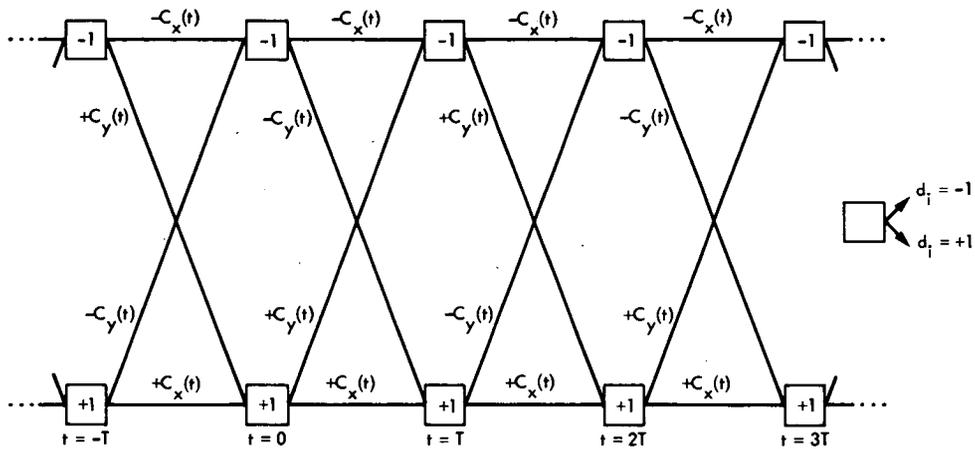


Fig. 5. Waveform trellis for the modulator of Fig. 4

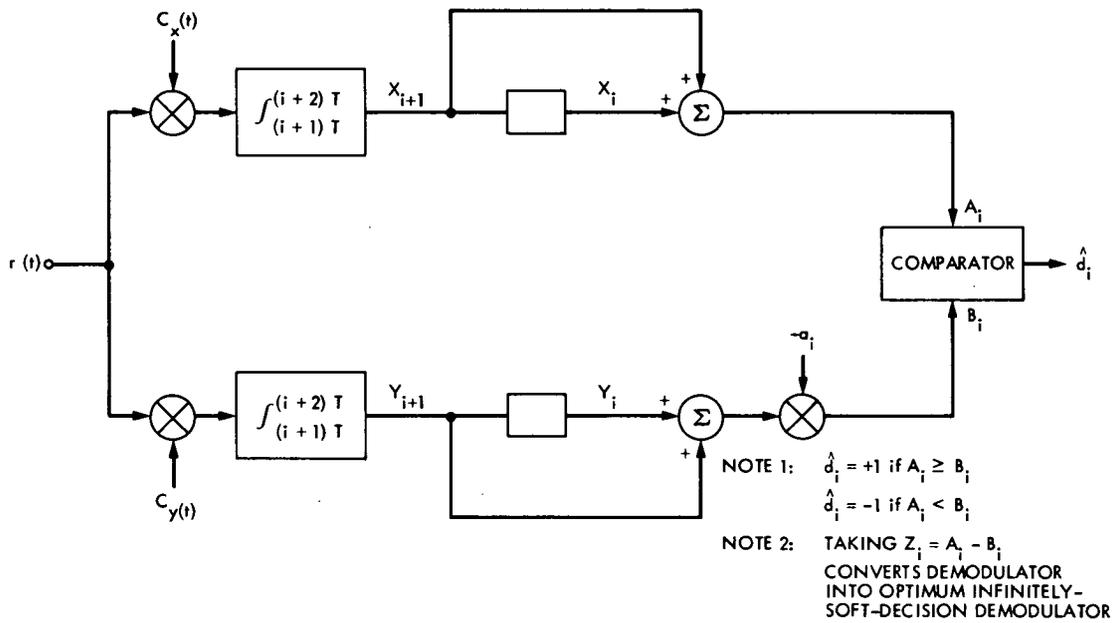


Fig. 6. Optimum hard-decision demodulator for the modulator of Fig. 4

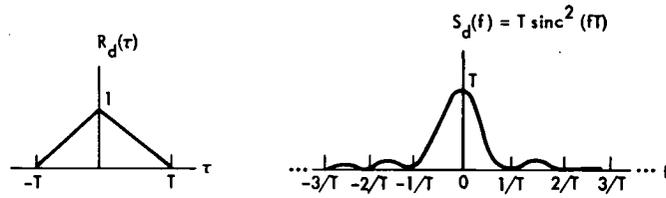


Fig. 7. The autocorrelation function $R_d(\tau)$ and the power spectral density $S_d(f)$ of the random data signal $d(t + t_0)$

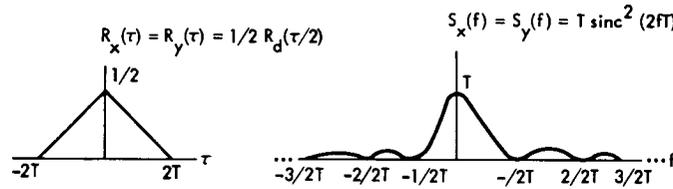


Fig. 8. The autocorrelation functions and power spectra of the outputs $x(t)$ and $y(t)$ of the sequence transducer of Fig. 1

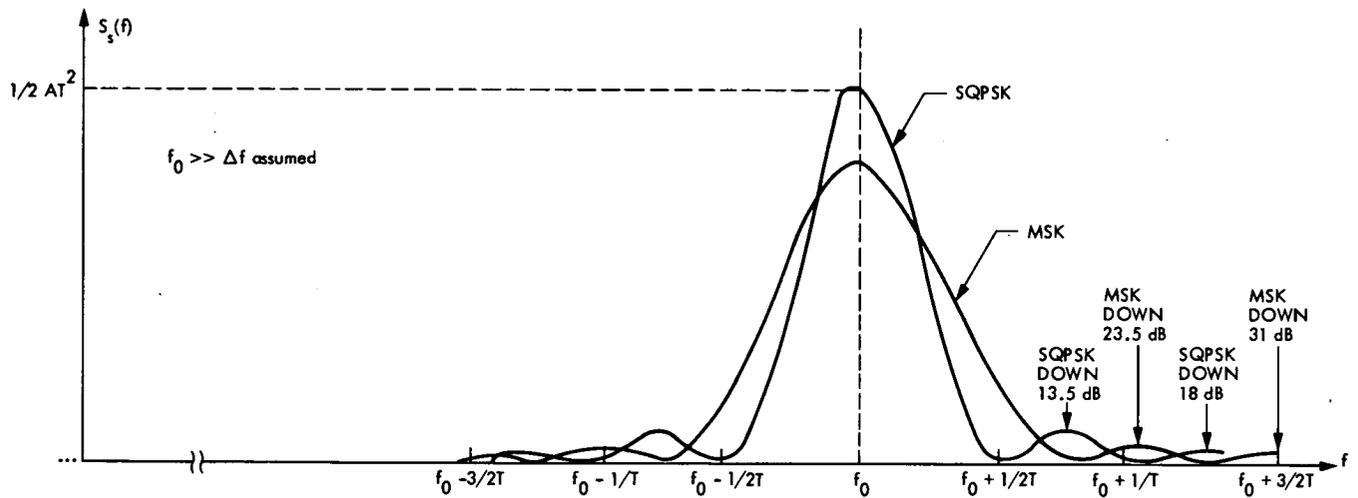


Fig. 9. Power spectra of MSK and SQPSK showing attenuation relative to spectral peak at sidelobe center frequencies

Internal Noise of a Phase-Locked Receiver With a Loop-Controlled Synthesizer

C. A. Greenhall

Communications Systems Research Section

We propose a local oscillator design that uses a digitally programmed frequency synthesizer instead of an analog VCO. The integral of the synthesizer input, the "digital phase," is a convenient measure of integrated doppler. We examine the internal noise of such a receiver. At high carrier margin, the local oscillator phase noise equals that of the Block IV receiver, about 2 deg rms at S-band, whereas the digital phase noise is about 0.5 deg rms.

I. Introduction

A simplified diagram of the carrier tracking function of the Block IV receiver is shown in Fig. 1. To obtain the local oscillator output, the doubled output of the voltage-controlled crystal oscillator (VCXO) is mixed with the output of the Dana Synthesizer and multiplied up to the first heterodyne frequency at S-band or X-band. The synthesizer is programmed to follow the previously estimated carrier frequency, and the difference between the estimated and actual carrier phases is made up by the loop-controlled VCXO.

It has been proposed that the VCXO be replaced by a loop-controlled digital frequency synthesizer such as the Dana Synthesizer already used in the programmed oscillator. This idea leads to the local oscillator design shown in Fig. 2. The 50-MHz reference clocks the synthesizer and is also mixed with the synthesizer output; the purpose of mixing here is to reduce the factor by which synthesizer phase noise is multiplied. The gain applied to the digital loop error signal is chosen to make the loop transfer function essentially the same as that of the present Block IV receiver, for a chosen bandwidth setting.

An important feature is the *digital phase*. At present, doppler phase is extracted by mixing the local oscillator output with a signal from the exciter. In the proposed design, the frequency programmed into the synthesizer is integrated digitally to produce a phase that is in step with the contents of the synthesizer's internal phase register. The synthesizer itself is a phase-locked loop that tries to keep its output phase in step with this internal register. Thus the digital phase can be expected to stay close to the local oscillator phase.

The main purpose of this article is to compare the digital phase noise of the proposed receiver to the local oscillator phase noise of the present Block IV receiver. The models include only the "internal" phase noises due to the Block IV VCXO and the Dana Synthesizers; in particular, thermal noise and the jitter of the station reference timing signal are ignored. Effects of sampling and quantization in the proposed receiver are also ignored.

We shall see that, for high carrier margin (at least 40 dB above threshold), most of the noise in the Block IV local oscillator comes from the Dana Synthesizer. Thus, the local

oscillator phase noises of the present and proposed receivers are almost equal, about 2 deg rms at S-band. The rms noise of the digital phase of the proposed receiver is about 0.5 deg, but grows very slowly as integration time increases.

II. Properties of the Dana Synthesizer

Since the Dana Synthesizer appears in both the present and the proposed local oscillators, we shall first set down our best estimates of its transfer function and phase noise spectrum. Assume that the reference timing signal has no jitter. Let $\nu(t)$ be the frequency programmed into the synthesizer. (Actually, $\nu(t)$ is sampled every 10 μ s but we shall treat time as continuous.) Then the synthesizer output phase is

$$2\pi \frac{L_d(s)}{s} \nu(t) + n_d(t)$$

where $L_d(s)$ is the phase-to-phase transfer function of the device, and $n_d(t)$ is the additive phase noise.

A gross approximation of the transfer function is given by

$$L_d(s) = \frac{1 + t_0 s}{1 + t_0 s + t_0^2 s^2 / 2}, \quad (1)$$

where $t_0 = 165 \mu$ s. The one-sided fiducial bandwidth of the filter $L_d(s)$ is $0.75/t_0 = 4500$ Hz, which is large enough so that we may replace $L_d(s)$ by 1 when using the synthesizer as a loop-controlled oscillator in a narrow-band tracking loop.

Our estimate of the one-sided spectral density of the phase noise n_d in rad^2/Hz is given by

$$\begin{aligned} S_d(2\pi f) &= \frac{N_d}{f}, & 0 < f \leq 100 \text{ Hz} \\ &= \frac{N_d}{100}, & 100 \text{ Hz} \leq f \leq 10 \text{ kHz} \\ &= \frac{N_d 10^{10}}{f^3}, & 10 \text{ kHz} \leq f \leq 50 \text{ kHz} \\ &= 0, & f > 50 \text{ kHz} \end{aligned} \quad (2)$$

where

$$N_d = 1.5 \times 10^{-8}.$$

This noise is nonstationary because of the $1/f$ term at low frequencies. The noise variance for Fourier frequencies f above 1 Hz is

$$\sigma_d^2 = 2.3 \times 10^{-6} \text{ rad}^2. \quad (3)$$

Jitter in the reference is not included. Both N_d and σ_d^2 have a ± 3 dB uncertainty.

The following explains how these values were arrived at; the uninterested reader can skip to Section III.

We first carried out a theoretical computation of the Dana Synthesizer phase noise spectrum. The model consisted of a VCO with $1/f^3$ phase noise spectrum (Ref. 1), controlled by a second-order loop with perfect integrator. This yielded a synthesizer phase noise spectral density that behaves like f at low frequencies. Because two references (Refs. 1 and 2) report spectral densities approximately like $1/f$ at low frequencies, we were forced to abandon our model and rely wholly on an understanding of the synthesizer as a black box.

The shape of the spectrum in (2) comes from a paper by G. Gillette of Dana Laboratories (Ref. 1); he plots the spectral density down to $f = 1$ Hz. Meyer and Sward of JPL (Ref. 2) measured a spectral density like $1/f^{0.75}$ for $4 \text{ Hz} \leq f \leq 80 \text{ Hz}$. We shall assume a $1/f$ behavior down to zero frequency. This assumption may be dangerous.

Because a frequency synthesizer derives its time base from a reference oscillator, whether internal or external, one must enquire whether published reports of synthesizer phase noise include the jitter of the reference. Meyer and Sward used a technique that cancels reference jitter. Dana's manual for its synthesizer (Ref. 3) gives a method for measuring total phase noise variance (discussed below); this method also cancels reference jitter. We therefore assume that Gillette's graph does not include reference phase noise.

Since Refs. 1 and 2 deal with an old model of the Dana Synthesizer, we used Gillette's graph to get the shape of the phase noise spectrum, but used phase noise specifications of a current model to obtain the normalizing constant N_d . Reference 3 specifies a total phase noise of $-54 \text{ dB} = 4 \times 10^{-6} \text{ rad}^2$ in the band $f_0 \pm 15 \text{ kHz}$, excluding $f_0 \pm 1 \text{ Hz}$, where f_0 is the programmed output frequency. Figure 3 shows in simplified form the setup used to verify this specification. The reference signal of synthesizer 2 is slaved to that of synthesizer 1.

Suppose that the synthesizers have outputs

$$x_1(t) = \cos(2\pi f_1 t + \theta_1(t) + \theta_0)$$

$$x_2(t) = \cos(2\pi f_2 t + \theta_2(t)).$$

where f_1 and f_2 are the programmed frequencies, $\theta_1(t)$ and $\theta_2(t)$ are independent, zero-mean phase noises, and θ_0 is a constant phase offset. The rms voltmeter sees the signal

$$y(t) = K \sqrt{2} \cos(2\pi(f_1 - f_2)t + \theta(t) + \theta_0),$$

where $\theta(t) = \theta_1(t) - \theta_2(t)$, and K is an unknown constant. The first thing we do in the test is to set $f_1 = f_2 + 1$ kHz. Then the voltmeter reading is K . This is the reference level. Next, we set $f_1 = f_2 + f$, where $0.05 \text{ Hz} \leq f \leq 0.1 \text{ Hz}$, and observe the *peak reading* of the rms voltmeter, which now sees the signal

$$y(t) = K \sqrt{2} [\cos(2\pi f t + \theta_0) \cos \theta(t) - \sin(2\pi f t + \theta_0) \sin \theta(t)].$$

The phase noise specification is defined to be the square of the ratio of peak reading to reference level. Over the short run, say up to one minute, we can assume $\theta(t) \ll 1$ rad; hence

$$y(t) \approx K \sqrt{2} [\cos(2\pi f t + \theta_0) - \theta(t) \sin(2\pi f t + \theta_0)].$$

The voltmeter, insensitive below 1 Hz (Fig. 3), does not respond to the first term. The second term causes the voltmeter reading to rise and fall every $1/(2f)$ seconds; the peak reading, $K \sqrt{2} \sigma(\theta)$, occurs when the synthesizers are in quadrature. Dividing by the reference level K , we conclude from the published phase noise specification that

$$2\sigma^2(\theta) = 2(\sigma^2(\theta_1) + \sigma^2(\theta_2)) = 4 \times 10^{-6} \text{ rad}^2.$$

Therefore, the larger of $\sigma^2(\theta_1)$ and $\sigma^2(\theta_2)$ is between 10^{-6} rad^2 and $2 \times 10^{-6} \text{ rad}^2$. We shall use

$$\sigma_{\text{spec}}^2 = 2 \times 10^{-6} \text{ rad}^2.$$

To compute N_d we set

$$\int_1^{15000} S_d(2\pi f) df = \sigma_{\text{spec}}^2.$$

Integrating (2) gives $\sigma_{\text{spec}}^2 = 131.4 N_d$, whence $N_d = 1.5 \times 10^{-8}$.

A more direct estimate of N_d comes from the single side band (SSB) phase noise spectral density specification in Ref. 3. The phase noise in the band $f_0 \pm 15$ kHz, excluding $f_0 \pm 10$ Hz, is given as -90 dB/Hz maximum. The SSB spectral density, denoted by $\mathcal{L}(f)$ in the literature, is asymptotically related to the baseband phase noise spectral density $S_d(2\pi f)$ by

$$\mathcal{L}(f) \sim \frac{1}{2} S_d(2\pi f),$$

provided f is large enough so that

$$\int_f^\infty S_d(2\pi f') df' \ll 1 \text{ rad}^2$$

(Ref. 4). From (2) we set $\frac{1}{2} S_d(20\pi) = 10^{-9}$. Then $N_d/20 = 10^{-9}$; $N_d = 2 \times 10^{-8}$. This agrees well with the 1.5×10^{-8} computed earlier. We settle on the latter figure with a ± 3 -dB uncertainty.

Finally, the value of σ_d^2 in (3) is obtained from σ_{spec}^2 by including the extra power from 15 kHz to 50 kHz.

III. Internal Noise of Present Block IV Receiver

The rms internal local oscillator phase noise at S-band is given in Table 1 as a function of design point bandwidth b_{LO} and carrier margin. The total noise is the RSS of the noises from the VCXO and the Dana Synthesizer. At high carrier margin, most of the noise, about 1.7 deg, comes from the Dana Synthesizer and is almost independent of bandwidth. Most of the local oscillator noise is outside the loop passband. Derivations of these results follow.

From Fig. 1 we see that the noise in the local oscillator phase $\hat{\theta}$, with the VCXO input shorted, is

$$M(n_d(t) + 2n_v(t)),$$

where n_d and n_v are the phase noises of the Dana Synthesizer and the VCXO, and $M = 20$ for S-band. The noise n_d is described in Section II. We shall assume that n_v has a one-sided spectral density N_v/f^3 ; the constant N_v will be calibrated from receiver specifications.

When the phase-locked loop is allowed to control the local oscillator, its phase noise becomes

$$(1 - L(s))M(n_d(t) + 2n_v(t)),$$

where

$$L(s) = \frac{1 + \tau_2 s}{1 + \tau_2 s + \tau_2^2 s^2 / r}, \quad (4)$$

$$\tau_2 = 0.75/b_{LO},$$

and r depends on the one-sided threshold loop bandwidth b_{LO} and on the carrier margin. At threshold, $r = 2$. At high margin,

$$\begin{aligned} r &= 32 \text{ for } 2b_{LO} = 1, 10, 100 \text{ Hz} \\ &= 18.5 \text{ for } 2b_{LO} = 3, 30, 300 \text{ Hz} \end{aligned}$$

We are using the perfect integrator approximation of the loop filter, $F(s) = (1 + \tau_2 s)/(\tau_1 s)$.

The high-pass filter $1 - L(s)$, acting on n_d , cancels the $1/f$ spectrum at low frequencies and passes most of the rest. Therefore, the rms contribution of the Dana Synthesizer to local oscillator phase noise is essentially

$$M\sigma_d = 3.0 \times 10^{-2} \text{ rad} = 1.7 \text{ deg}.$$

We shall ignore the decrease at the higher loop bandwidths since it is at most about 0.1 deg.

The variance σ_v^2 of $(1 - L(s))n_v$ is given on p. 32 of Ref. 5. It is convenient to recast those formulas as follows:

$$\sigma_v^2 = \frac{N_v \tau_2^2}{r^2 \sqrt{-D}} \sin^{-1} \left(\frac{r \sqrt{-D}}{2} \right), \quad 2 \leq r < 4$$

$$= \frac{N_v \tau_2^2}{8}, \quad r = 4$$

$$= \frac{N_v \tau_2^2}{r^2 \sqrt{D}} \ln \left(\frac{r}{2} - 1 + \frac{r \sqrt{D}}{2} \right), \quad r > 4$$

where $D = 1 - 4/r$.

The specifications of the Block IV Receiver-Exciter Subsystem give a rms phase jitter of 9 deg when $\tau_2 = 1.5$ s, $r = 2$, $M = 20$. (This results from operating the VCXO in a phase-locked loop with these parameters and a strong signal, so that thermal noise is absent.) This calibrates N_v and allows us to fill in the VCXO noise column of Table 1, which lists $2M\sigma_v$ at threshold and at high margin. When the receiver is switched to wide mode, N_v is multiplied by 100, τ_2 is divided by 10, and r is unchanged. Therefore, b_{LO} is multiplied by 10 and σ_v is unchanged.

Table 1 also lists the total rms internal noise $\sigma(\hat{\theta})$, where

$$\sigma^2(\hat{\theta}) = (M\sigma_d)^2 + (2M\sigma_v)^2.$$

IV. Internal Noise of Proposed Receiver

We now have to consider two phases, the local oscillator output phase $\hat{\theta}$ and the digital phase $\tilde{\theta}$. The rms noise $\sigma(\hat{\theta})$ is about 1.7 deg at S-band, the same as the Dana Synthesizer component of the noise in the present Block IV receiver. Since the VCXO is absent, $\sigma(\hat{\theta})$ does not rise as carrier margin decreases. Again, most of the $\hat{\theta}$ noise is outside the loop passband.

On the other hand, most of the noise in the digital phase $\tilde{\theta}$ is inside the loop passband. This noise turns out to be nonstationary, with a spectral density like $1/f$ at low frequencies. Thus it makes no sense to quote a single variance $\sigma^2(\tilde{\theta})$. Instead, we compute

$$\begin{aligned} \sigma^2(\tilde{\theta}, T) &= \frac{1}{2} \text{var} [\tilde{\theta}(t+T) - \tilde{\theta}(t)] \\ &= \frac{1}{2} \text{var} \int_t^{t+T} \frac{d\tilde{\theta}}{dt} dt, \end{aligned}$$

which is independent of t because $d\tilde{\theta}/dt$ is stationary. The factor $1/2$ is needed to make a fair comparison with $\sigma(\hat{\theta})$, for if we apply the operator $\sigma^2(\cdot, T)$ to the stationary process $\hat{\theta}$, we get

$$\begin{aligned} \sigma^2(\hat{\theta}, T) &= \frac{1}{2} \text{var} (\hat{\theta}(t+T) - \hat{\theta}(t)) \\ &= R(0) - R(T) \end{aligned}$$

$$\rightarrow R(0) = \sigma^2(\hat{\theta}) \quad \text{as } T \rightarrow \infty,$$

where R is the autocovariance function of $\hat{\theta}$. On the other hand, we shall see that

$$\sigma^2(\tilde{\theta}, T) \rightarrow \infty \quad \text{as } T \rightarrow \infty.$$

Table 2 lists the rms noise $\sigma(\tilde{\theta}, T)$ at S-band as a function of T/τ_2 and the loop threshold bandwidth b_{LO} , where $\tau_2 = 0.75/b_{LO}$, for high carrier margin. The values decrease slightly as margin decreases. Since $\tilde{\theta}$ is "almost" stationary, $\sigma^2(\tilde{\theta}, T)$ grows slowly with T , in fact, like $\log T$. We must warn the reader, however, that this result holds only if the spectral density $S_d(2\pi f)$ of the Dana Synthesizer phase noise behaves like $1/f$ at least down to $f = 1/(4T)$. Derivations of these results follow.

Although the loop error signal of the proposed receiver is digital, we shall proceed with an analog analysis. A phase-locked receiver using the local oscillator of Fig. 2 has the closed-loop transfer function.

$$L_1(s) = \frac{1 + \tau_2 s}{1 + \tau_2 s + \tau_2^2 s^2 / (\tau L_d(s))}$$

where $L_d(s)$, given by (1), is the transfer function of the Dana Synthesizer. The one-sided bandwidth of $L_d(s)$ is about 4500 Hz. The internal noise in the local oscillator phase $\hat{\theta}$ is

$$(1 - L_1(s)) Mn_d(t) \quad (5)$$

where $n_d(t)$ is the Dana Synthesizer phase noise. As with the Block IV receiver, the rms value $\sigma(\hat{\theta})$ is approximately $M\sigma_d = 1.7$ deg for S-band.

To obtain the internal noise in the digital phase $\tilde{\theta}$, suppose that $n_d(t)$ is the only active influence on the loop. The frequency (Hz) programmed into the synthesizer is $(1/2\pi M) d\tilde{\theta}/dt$. The local oscillator phase (rad) is

$$\begin{aligned} \hat{\theta} &= M \left(2\pi \frac{L_d(s)}{s} \frac{s\tilde{\theta}}{2\pi M} + n_d \right) \\ &= L_d(s)\tilde{\theta} + Mn_d. \end{aligned}$$

Replacing $\hat{\theta}$ by (5), we get

$$\tilde{\theta} = \frac{\hat{\theta} - Mn_d}{L_d(s)} = \frac{L_1(s)}{L_d(s)} Mn_d.$$

For computation, we shall approximate L_1/L_d by the Block IV transfer function L , given by (4). Then

$$\begin{aligned} \sigma^2(\tilde{\theta}, T) &= \frac{1}{2} \text{var} \int_0^T \frac{d\tilde{\theta}}{dt} dt \\ &= \int_0^\infty (1 - \cos \omega T) S_{\tilde{\theta}}(\omega) \frac{d\omega}{2\pi} \\ &= M^2 \int_0^\infty (1 - \cos \omega T) |L(i\omega)|^2 S_d(\omega) \frac{d\omega}{2\pi} \end{aligned}$$

where $S_d(\omega)$ is given by (2). This leads to the further approximation

$$\sigma^2(\tilde{\theta}, T) = M^2 N_d (J_1 + J_2)$$

where

$$J_1 = \int_0^\infty (1 - \cos \omega T) |L(i\omega)|^2 \frac{d\omega}{\omega} \quad (6)$$

is the main portion, and

$$J_2 = \frac{1}{200\pi} \int_{200\pi}^\infty |L(i\omega)|^2 d\omega \quad (7)$$

is a correction that is significant only for $2b_{LO} = 100$ Hz or 300 Hz. In the Appendix, these integrals are evaluated in closed form in terms of the elementary functions and the exponential integral $\int (e^z/z) dz$. Numerical results are given in Table 2.

V. Conclusions

We have proposed a design for a receiver local oscillator containing a loop-controlled synthesizer. The transfer function of the proposed receiver and its performance in the presence of thermal noise are essentially the same as those of the present Block IV receiver. The proposed design provides a convenient digital estimate of carrier phase, obtained by integrating the synthesizer input frequency.

For S-band and high carrier margin, the rms internal phase noises of the present (Block IV) and proposed receivers are approximately as follows:

1.7 – 2.0 deg Present local oscillator phase noise.
Increases as carrier margin decreases.

1.7 deg

Proposed local oscillator phase noise.
Insensitive to carrier margin.

0.3 – 0.7 deg

Proposed digital phase noise.
Decreases as margin decreases.
Increases as integration time increases.

Appendix

Evaluation of Two Integrals

The integrals in question are J_1 and J_2 , given by (6) and (7). We rewrite them as follows:

$$J_1 = \int_0^{\infty} \frac{1 - \cos ax}{x} f(x) dx$$

$$J_2 = \frac{1}{x_0} \int_{x_0}^{\infty} f(x) dx$$

where

$$f(x) = \frac{1 + rx^2}{1 + (r-2)x^2 + x^4},$$

$$a = \sqrt{r} T / \tau_2, \quad x_0 = 200\pi \tau_2 / \sqrt{r}.$$

From now on, we assume that $r > 4$.

The partial fraction expansion of $f(x)$ is

$$f(x) = \frac{A_1}{x^2 + x_1^2} + \frac{A_2}{x^2 + x_2^2}$$

where

$$A_1 = \frac{1 - rx_1^2}{x_2^2 - x_1^2}, \quad A_2 = \frac{1 - rx_2^2}{x_1^2 - x_2^2}$$

$$x_2^2 = \frac{r}{2} - 1 + \frac{r}{2} \sqrt{1 - 4/r}, \quad x_1^2 = 1/x_2^2.$$

Integrating $f(x)$, we get

$$J_2 = \frac{x_2}{x_0(x_2^4 - 1)} \left[(x_2^2 - r) \text{Tan}^{-1} \frac{1}{x_0 x_2} + (rx_2^2 - 1) \text{Tan}^{-1} \frac{x_2}{x_0} \right].$$

To evaluate J_1 we evaluate the integral

$$I(t) = \int_0^{\infty} \frac{1 - \cos tx}{x(1+x^2)} dx$$

In terms of the exponential integrals

$$E_1(t) = \int_t^{\infty} \frac{e^{-x}}{x} dx,$$

$$\text{Ei}(t) = \int_{-\infty}^t \frac{e^x}{x} dx,$$

which are tabulated in Ref. 6. We use Raabe's integrals (Ref. 7):

$$I_s(t) = \int_0^{\infty} \frac{\sin tx}{1+x^2} dx = \frac{1}{2} \left[e^t E_1(t) + e^{-t} \text{Ei}(t) \right],$$

$$I_c(t) = \int_0^{\infty} \frac{x \cos tx}{1+x^2} dx = \frac{1}{2} \left[e^t E_1(t) - e^{-t} \text{Ei}(t) \right].$$

(The integral form of $I_c(t)$ will not play a part here.) Since

$$I(t) = \int_0^t I_s(u) du, \quad (8)$$

$$\frac{d}{dt} I_c(t) = I_s(t) - \frac{1}{t},$$

we have, for some constant C ,

$$I(t) = I_c(t) + \ln t + C, \quad t > 0.$$

As $t \rightarrow 0$, $I(t) \rightarrow 0$ from (8). Limiting expressions for $E_1(t)$ and $\text{Ei}(t)$ give

$$I_c(t) + \ln t \rightarrow -\gamma \quad \text{as } t \rightarrow 0,$$

$$J_1 = (1+k)I_c(ax_2) - kI_c(a/x_2) + \gamma + (1+2k)\ln x_2 + \ln a$$

where γ is Euler's constant, 0.577216. Therefore, $C = \gamma$.

where

For large t ,

$$I(t) = \ln t + \gamma - 1/t^2 + O(1/t^4).$$

$$k = \frac{2x_2^2 + 1}{x_2^4 - 1}.$$

Finally, we have

These formulas were spot-checked by numerical integration.

References

1. Gillette, G. C., "The Digiphase Synthesizer," *Frequency Technology* 7, pp. 25-29, 1969.
2. Meyer, R., and Sward, A., "Frequency Generation and Control: The Measurement of Phase Jitter," *Space Programs Summary 37-64*, Vol. II, pp. 55-58, Jet Propulsion Laboratory, Pasadena, Calif., 1970.
3. *Dana Series 7000 Digiphase® Frequency Synthesizers (7010-S-241)*, Publication 980428, Dana Laboratories, Inc., Irvine, Calif., April 1970, revised March 1973.
4. Shoaf, J. H., Halford, D., and Risley, A. S., *Frequency Stability Specification and Measurement: High Frequency and Microwave Signals*, NBS Technical Note 632, National Bureau of Standards, 1973.
5. Tausworthe, R. C., *Theory and Practical Design of Phase-Locked Receivers*, Vol. 1, Technical Report 32-819, Jet Propulsion Laboratory, Pasadena, Calif., 1966.
6. Abramowitz, M., and Stegun, I. A., *Handbook of Mathematical Functions*, New York, 1972.
7. Erdélyi, A., et al., *Higher Transcendental Functions*, Vol. II, New York, 1953.

Table 1. RMS internal noise of Block IV receiver at S-band (total noise is RSS of VCXO noise with Dana synthesizer noise $M\sigma_d = 1.7$ deg)

$2b_{LO}$, Hz		Carrier margin	VCXO noise, $2M\sigma_v$, deg	Total noise, $\sigma(\hat{\theta})$, deg
Narrow	Wide			
1	10	1	9	9
		∞	0.9	2.0
3	30	1	3.0	3.5
		∞	0.5	1.8
10	100	1	0.9	2.0
		∞	0.09	1.7
30	300	1	0.3	1.8
		∞	0.05	1.7

Table 2. RMS internal noise of digital phase $\hat{\theta}$ of proposed receiver at S-band as a function of integration time T (high carrier margin assumed)

		$T/\tau_2, \tau_2 = 0.75/b_{LO}$			
		1	10^2	10^4	10^6
$2b_{LO}$, Hz	1	0.29	0.42	0.52	0.60
	3	0.28	0.41	0.51	0.59
	10	0.29	0.42	0.52	0.60
	30	0.29	0.42	0.52	0.60
	100	0.42	0.52	0.60	0.67
	300	0.50	0.58	0.66	0.72

$\sigma(\hat{\theta}, T)$, deg

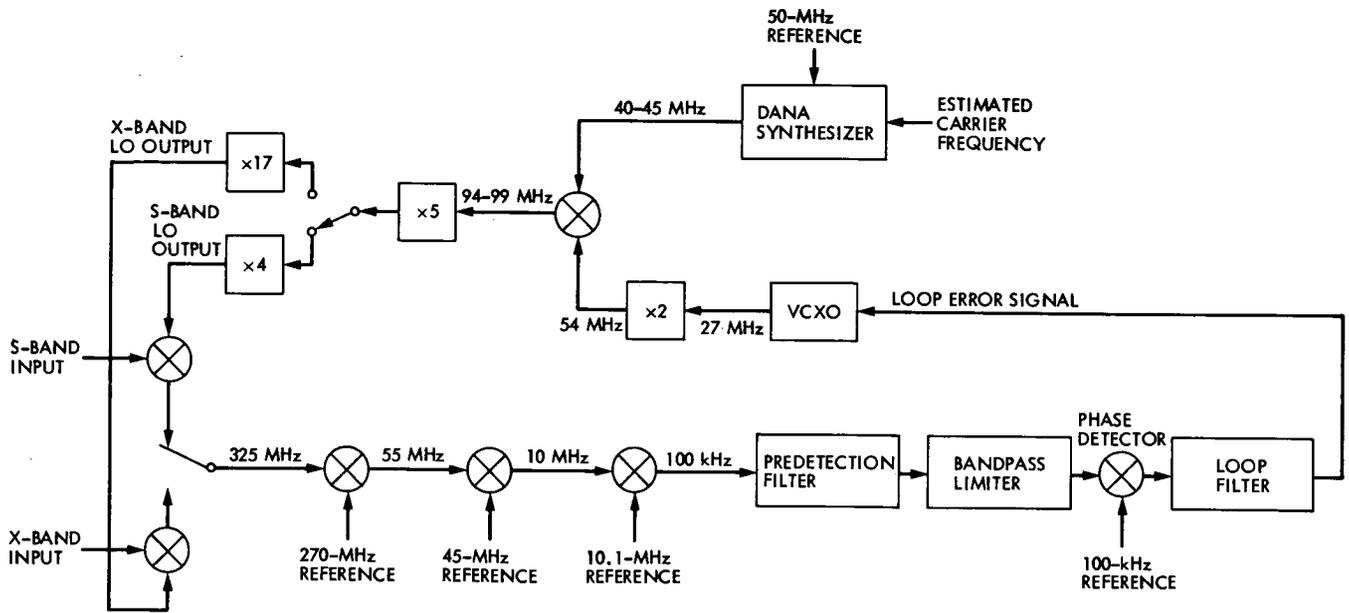


Fig. 1. Block IV carrier tracking receiver

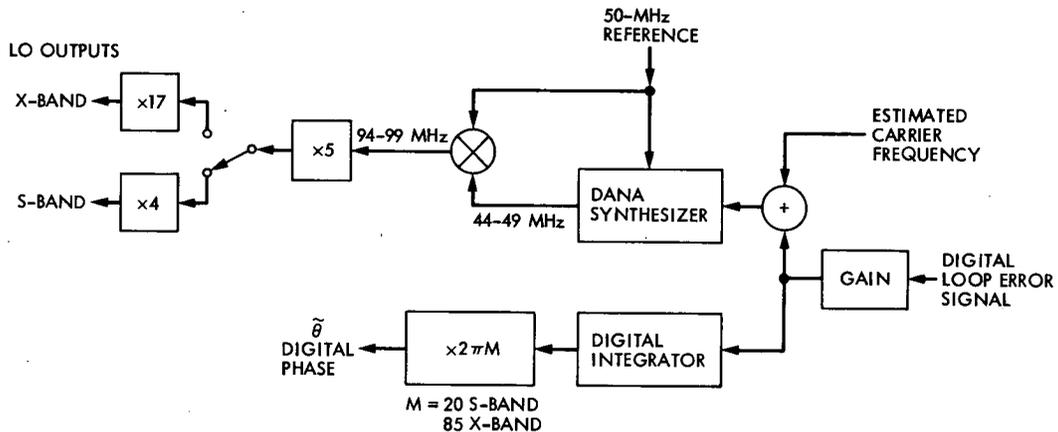


Fig. 2. Proposed local oscillator with Dana Synthesizer as loop-controlled oscillator

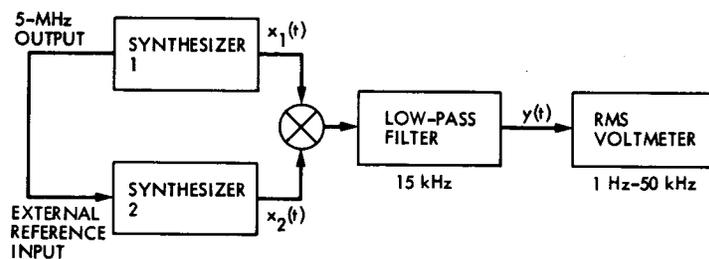


Fig. 3. Measurement of synthesizer phase noise variance

A Prototype DSN X-S Band Feed: DSS 13 Application Status (Third Report)

W. Williams, D. Nixon, H. Reilly, J. Withington, and D. Bathker
Radio Frequency and Microwave Subsystems Section

This article, the third in this series discussing a new prototype X-S band horn feed for future use at various DSN sites, deals with the testing of the final fabricated feed at DSS 13. Measured feedhorn patterns are presented and efficiencies calculated. Preliminary results of system noise temperature and 26-m antenna system gain measurements are presented. Also discussed are some measurements leading to an improved second generation feed.

The preliminary results of the field measurements indicate that this horn will perform as originally specified and required. Also the tests for the second generation feed have indicated the potential cause of minor X-band moding.

I. Introduction

The first report (Ref. 1) of this series discusses the design of a dual-band (X-S) corrugated horn for DSN applications. The radiation patterns and efficiencies of a half-scale model were presented in Ref. 1. The second report (Ref. 2) discusses the design of a diplexer or X-S combiner for injecting signals into the horn. A combiner for the X-S horn was fabricated. Also, efficiency calculations were made for the half-scale model horn patterns scattered from the Venus (DSS 13) site subreflector.

The final full-scale horn fabrication has been completed, and full-scale measurements were made. These measurements indicate some additional problems in hybrid moding, which must be resolved. The full-scale horn was assembled in a feedcone with other electronic equipment and shipped to the

DSS 13 site, where preliminary temperature measurements were made at X-band frequencies. The cone was mounted on the 26-m antenna for final evaluation.

The final full-scale horn pattern measurements were used in the JPL theoretical scattering program (Ref. 3) to determine final efficiencies after reflection from the DSS 13 site hyperboloid with vertex plate. The program (spherical wave expansion) takes into account the fact that the subreflector is at a range somewhat less than D^2/λ .

II. The Final X-S Horn

The final Phase I X-S horn is shown in Fig. 1, and measured radiation patterns of this horn are shown in Figs. 2, 3, and 4. Although these patterns are acceptable, there are some

differences from the measured half-scale model at frequencies in the high end of the X-band region (Fig. 2). There is a marked difference in the E and H plane patterns, the E plane being broader with a 1-dB null (or dip) on axis. This results in more forward spillover from this broader E-plane pattern, and a somewhat reduced efficiency.

The E and H plane patterns (Fig. 3) at 7.175 GHz overlay much better, although there is evidence of some difference near boresight where slight dips occur. Figure 4 indicates a slight difference in the E and H plane patterns at 2.295 GHz, but so minor as to be ignored.

It is next necessary to predict the final RF efficiency of DSS 13 with this feedhorn in place. This is done by decomposing the patterns into Fourier-type components using the spherical wave expansion technique of Ref. 3. These components are then used to determine currents upon the DSS 13 subreflector at the proper range from the horn and hence determine the final scattered pattern that illuminates the 26-m paraboloid. Fifty-eight of these expansion coefficients were determined for each X-band horn pattern, and 45 coefficients determined for the S-band horn pattern. The resulting scatter patterns are shown in Figs. 5, 6, and 7. The result of an unequal E and H horn pattern is noted in Fig. 5 since the scattered E and H plane patterns are also unequal.

Also it is apparent that the DSS 13 vertex plate was not designed for X-band. Geometric shadowing by the subreflector extends to about 8 deg. From Figs. 5 and 6 it is noted that this vertex plate achieves its purpose to about 13 to 15 deg, wasting a significant part of the inner paraboloidal region.

Next, the efficiency of these patterns is calculated as illuminators of the 26-m paraboloid. Values are tabulated in Table 1. Observe that these totals do not include spar blockage or surface tolerance. These are not a function of the X-S feed unit.

It is interesting to understand qualitatively the reasons for the slight differences in efficiencies at the various frequencies. The horn patterns (Figs. 2, 3, and 4) show the angle of 32.7 degrees, which is subtended by the total subreflector, thus representing the total portion of each pattern that is intercepted. The first item noted is the greater spillover at S-band than at X-band. The reason is evident when looking at these radiation patterns: the S-band pattern is triangular in shape, reaching the -40 dB level at nearly 36 deg, while X-band patterns have much steeper sides (slopes) and reach -40 dB at a smaller angle. Thus a greater fractional energy is intercepted at X-band. Next is noted the differences in illumination efficiency. The superiority of S-band can be noted in Figs. 5, 6, and 7. In this case, the large X-band hole due to the vertex

plate has drastically reduced the effectiveness of the reflector area compared to S-band. The superiority of 8.425 GHz illumination efficiency over 7.175 GHz can be seen in Figs. 5 and 6. The taper at 7.175 GHz (Fig. 6) can be seen as greater than 8.425 GHz (Fig. 5), the result of the greater taper on the hyperbola intercept (-16 dB, Fig. 3) compared to -14 dB at 8.425 GHz (Fig. 2). Also, phase efficiency is superior at S-band, again the effect of the vertex plate having the more serious effect at the shorter wavelengths. It becomes apparent that the subreflector could be redesigned for this antenna, if one desired further X-band improvement, to increase illumination and phase efficiency by, perhaps, a total of 5 percent.

III. The Feedcone

The new X-S feedhorn and other necessary equipment (power lines, brackets to hold masers and other electronic gear, etc.) were installed in a feedcone at JPL. The completed unit is shown in Fig. 8. This assembly was sent to DSS 13 for further measurement and installation on the 26-m antenna.

Figure 9 is a photograph of this new installation at DSS 13 and Fig. 10 depicts the older reflex feed as applied to the 34-m modifications.

The plan to evaluate the common aperture X-S horn, as installed in a feedcone system at DSS 13, was to measure total operating system noise temperature (T_{op}) in a normal feedcone-on-ground test mode followed by T_{op} (feedcone antenna mounted) and 26-m antenna area efficiency (η) measurements. Two fundamental problems have prevented the gathering of reliable data.

First, the cone-on-ground S-band measurements are clouded by a problem discovered after the feedcone was antenna mounted. Although JPL Mesa Antenna Range measurements at S-band were done with a proper assembly of the feedhorn yoke assembly (power division for phasing the four S-band horn inputs), an inadvertent reassembly within the feedcone introduced a phasing problem. Resultant 26-m antenna main beam initial measurements showed a highly squinted and elliptical beam prior to proper assembly of the yoke. Thus full confidence was not obtained in the present S-band T_{op} (23 K) since the expected 3 K feedcone OFF/ON antenna value (due to spillover and scatter) has not been demonstrated. The approximate S-band area efficiency under these uncertain conditions is 62 percent.

Secondly, initial feedcone-on-ground X-band T_{op} measurements appeared high (21 K) and not stable. Later, with the feedcone antenna mounted, the measurements were very high (about 30 to 40 K) and very unstable. Additionally, different

T_{op} measurement schemes produced different results. The problem was traced to the X-band TWM, an old unit that suffered from years of thermal cycling. The travelling wave comb structure had become distorted and short circuited in places. Now repaired and reinstalled in the antenna-mounted feedcone, different measurement schemes all indicate T_{op} in the low 20 K class. During the period with a suspect maser, X-band antenna efficiency appeared to be 48 percent. This value now requires another evaluation.

In summary, despite the above problems, the X-S common aperture horn (initial receive-only model) appears to have no major problems; a full evaluation is planned just prior to (and after) the next opportunity for feedcone removal from the 26-m antenna. At that time, feedcone-on-ground T_{op} measurements will be carefully compared with feedcone-on-antenna T_{op} measurements to discern the expected spillover and scatter contributions.

IV. Further Horn Study, Second Generation

The X-band horn has exhibited evidence of an operational problem as noted in the unequal E and H plane patterns of Fig. 2. There appears to be some small amount of moding, i.e., the generation of other corrugated hybrid modes than the desired HE_{11} mode. For instance, the EH_{11} mode may be generated at low X-band when input grooves are less than $\lambda/4$, or higher modes, the HE_{12} and EH_{12} , when the horn is sufficiently large and horn interruptions (discontinuities) may occur. It was not obvious what caused these problems.

Toward the end of solving this problem, several tests were made using segments of the full-scale horn that were left over from the first model development. An X-band input section was available, to about 12.7-cm (5-inch) aperture, and an additional section was built for enlarging the sample and to allow various tests to be made. Also, an additional input piece was made for the X-band, which consisted of a gradual radius change from the X-band cylindrical guide to the 17.1-deg horn

flare instead of the abrupt 17.1-deg horn attached to cylindrical waveguide.

Basically, the following tests were made:

- (1) Metallic taping off two grooves to simulate the combiner input that results in the pitch change for the corrugations.
- (2) Building the additional section at a different flare angle to simulate the change of flare angle for this horn.
- (3) Using the gradual flare change at the X-band input.

The following results were noted:

- (1) A significant (very noticeable) change occurs when the gradual taper is used. The abrupt change from cylindrical to the 17.1-deg flare is a major cause of moding.
- (2) Neither the small flare angle change nor the corrugation interruption caused a significant moding, although some was noticed.

The conclusion was that a new horn will be constructed using the gradual change to the 17.1-deg flare; over about a 10.2-cm (4-inch) transition length. Further investigation into the optimum groove depth and corrugation pitch is being carried on.

Also in this second-generation study, a new diplexer/combiner must be developed so that the S-band system will have sufficient bandwidth for transmission as well as low noise reception. Although the first generation diplexer (now at DSS 13) offered sufficient X-band isolation (Ref. 2), it had an operating bandwidth of less than 100 MHz in S-band. The second-generation diplexer being designed will use the same approach as the first-generation design, except that the radial transmission line for S-band injection will be about 70 percent wider to obtain the bandwidth, and the X-band radial line chokes will be increased from 2 to 6 in number to maintain the high X-band isolation.

References

1. W. F. Williams, "A Prototype DSN X-S Band Feed: DSS 13 First Status Application," *DSN Progress Report 42-44*, Jet Propulsion Laboratory, Pasadena, Calif., Jan.-Feb. 1978.
2. W. F. Williams, "A Prototype DSN X-S Band Feed DSS 13 Application Status (Second Report)," *DSN Progress Report 42-47*, Jet Propulsion Laboratory, Pasadena, Calif., July-Aug. 1978.
3. Arthur C. Ludwig, "Calculations of Scattered Patterns from Asymmetrical Reflectors," Technical Report 32-1430, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1970.

Table 1. Total efficiency as a function of frequency

Efficiency Type	Frequency		
	8425 GHz	7175 GHz	2295 GHz
Total spillover	0.96660	0.97139	0.88693
Illumination	0.86173	0.82389	0.88610
Cross-polarization	0.98861	0.99430	0.99590
Phase	0.89627	0.89301	0.92438
Blockage (subreflector)	1.0000	1.0000	0.96900
Totals	0.73804	0.70687	0.70107

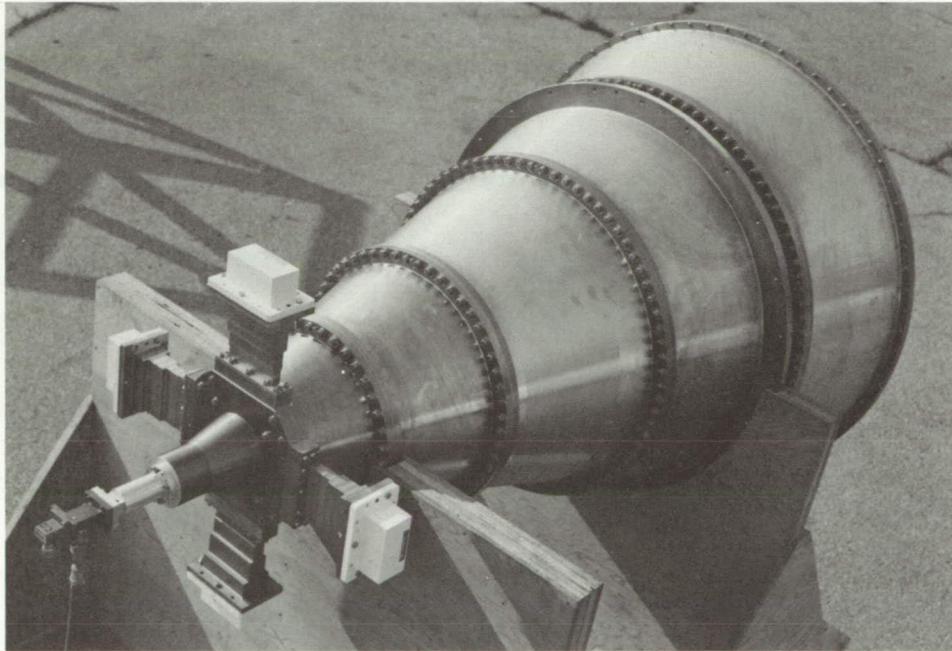


Fig. 1. Final first-generation X-S horn

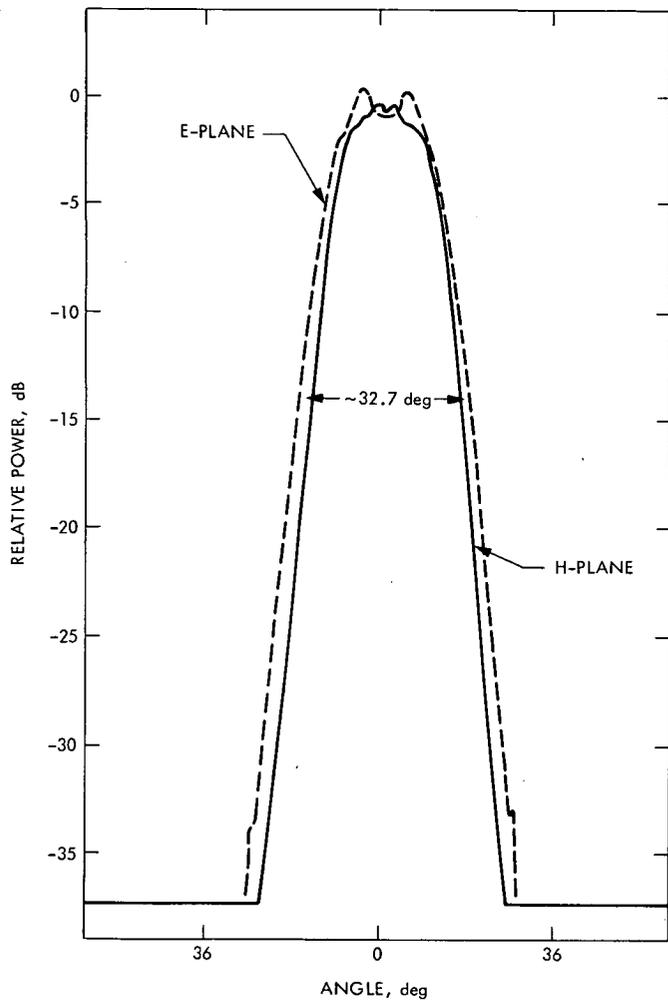


Fig. 2. Final X-S horn pattern at $f = 8.425$ GHz linear polarization

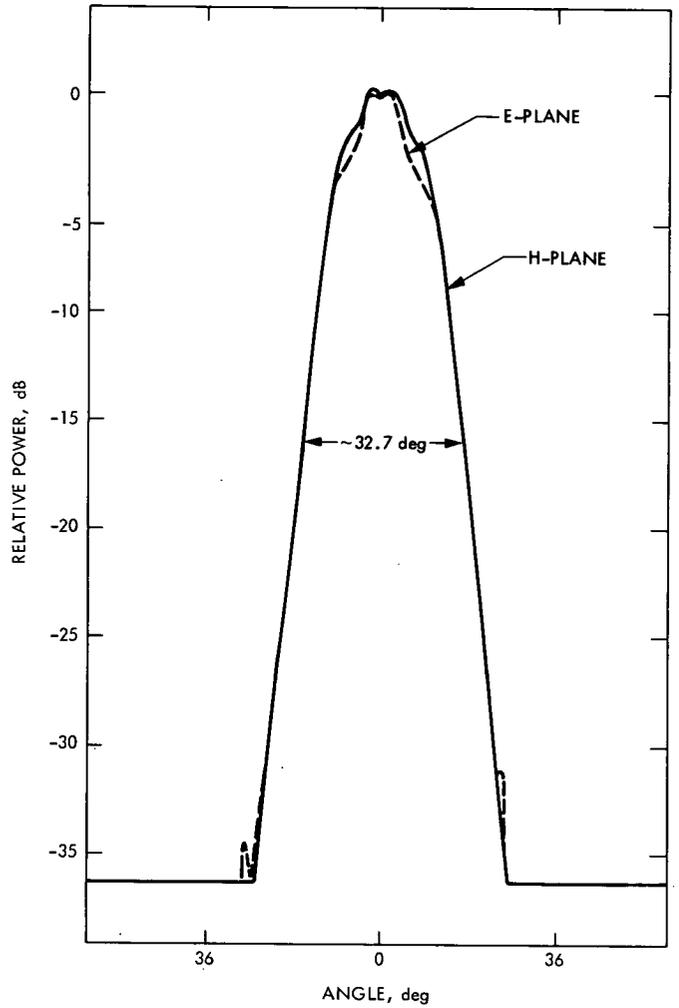


Fig. 3. Final X-S horn pattern at $f = 7.175$ GHz linear polarization

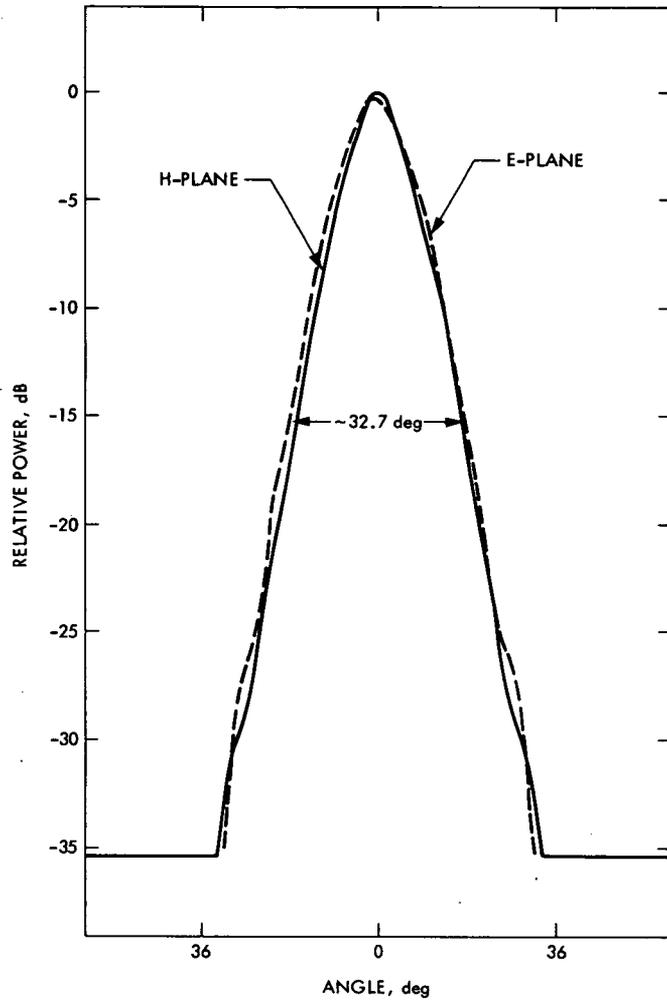


Fig. 4. Final X-S horn pattern at $f = 2.295$ GHz linear polarization

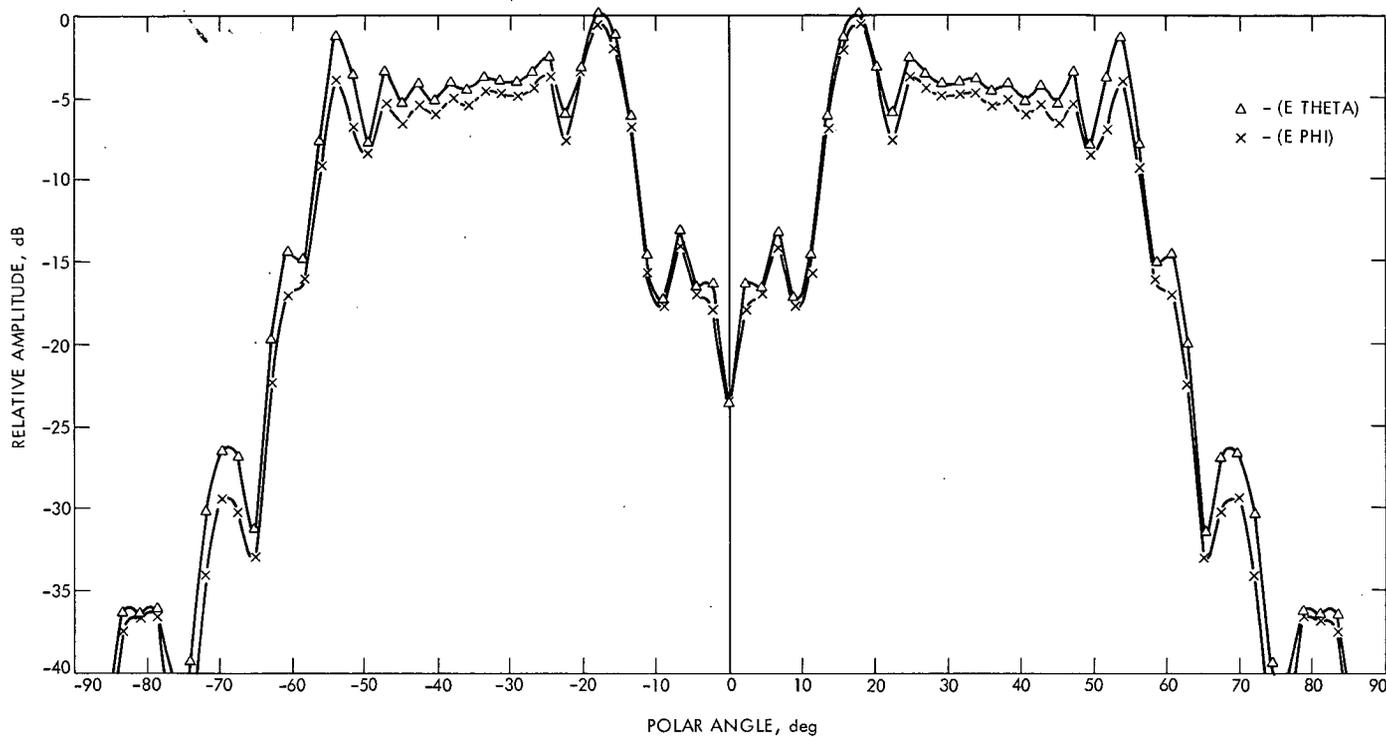


Fig. 5. The 8.425 measured amplitude pattern, 58 modes scattered from DSS 13 subreflector

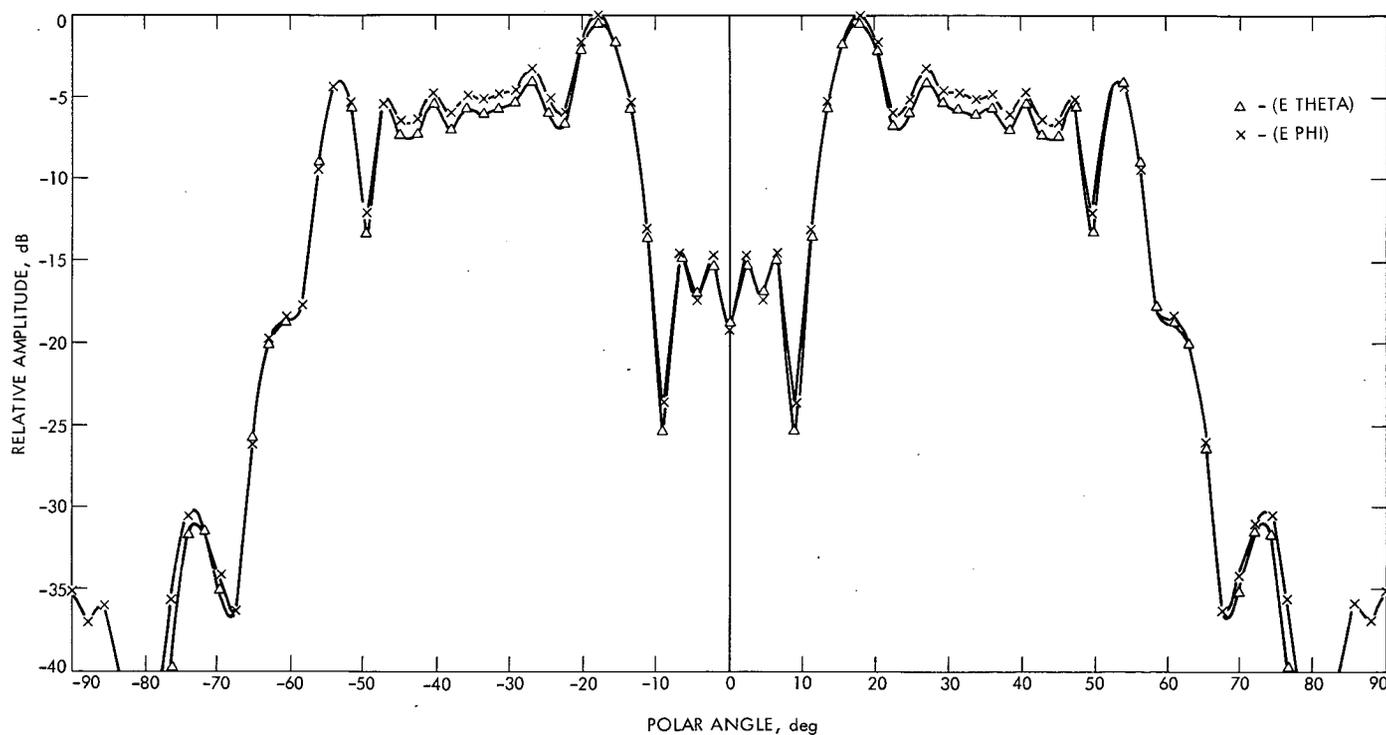


Fig. 6. The 7.175 measured amplitude pattern, 58 modes scattered from DSS 13 subreflector

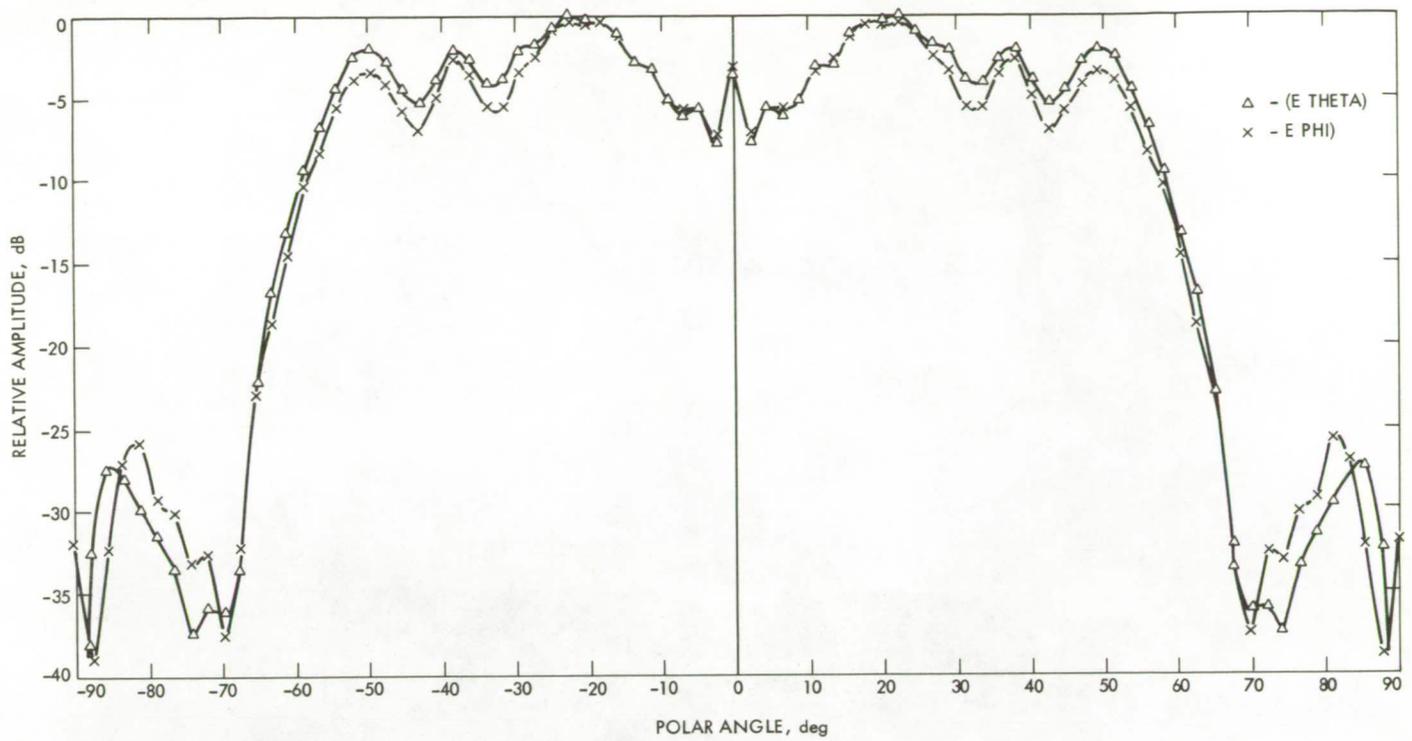


Fig. 7. The final 2.295 measured patterns, 45 modes, scattered from DSS 13 subreflector

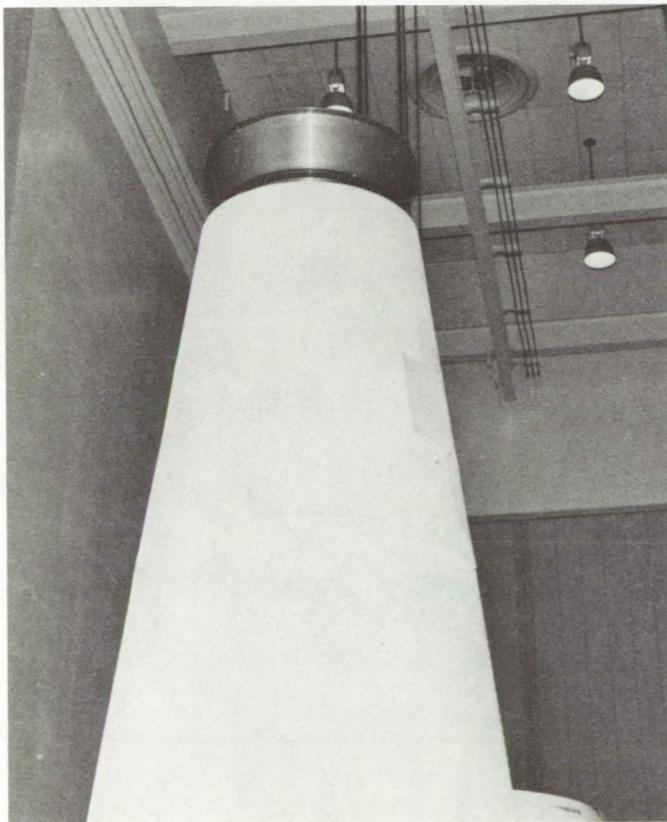


Fig. 8. First generation S-X feed mounted in its feedcone

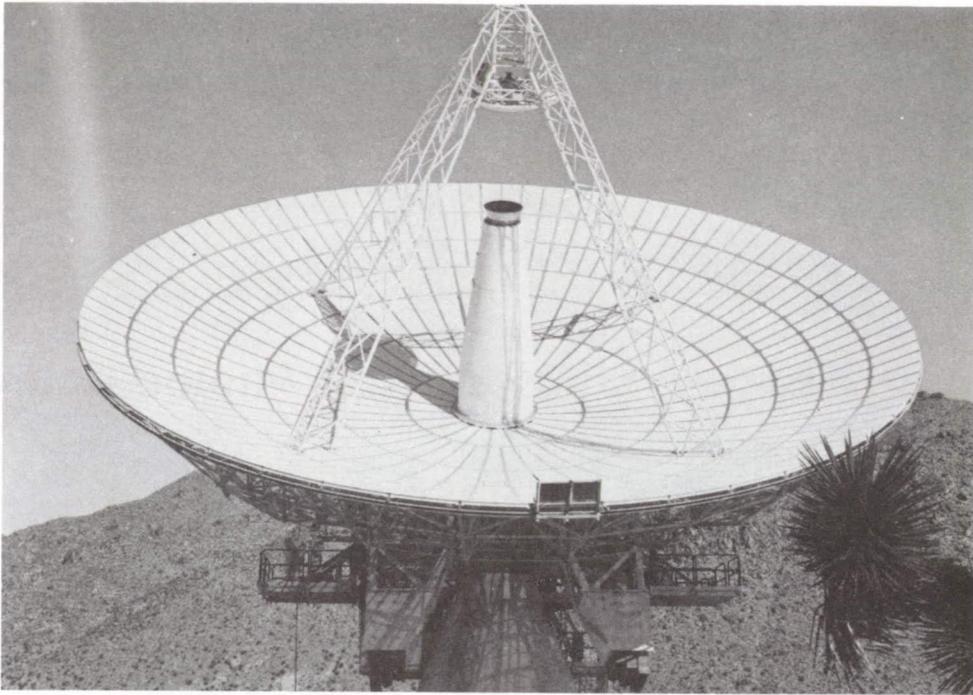


Fig. 9. The first-generation X-S horn at DSS 13

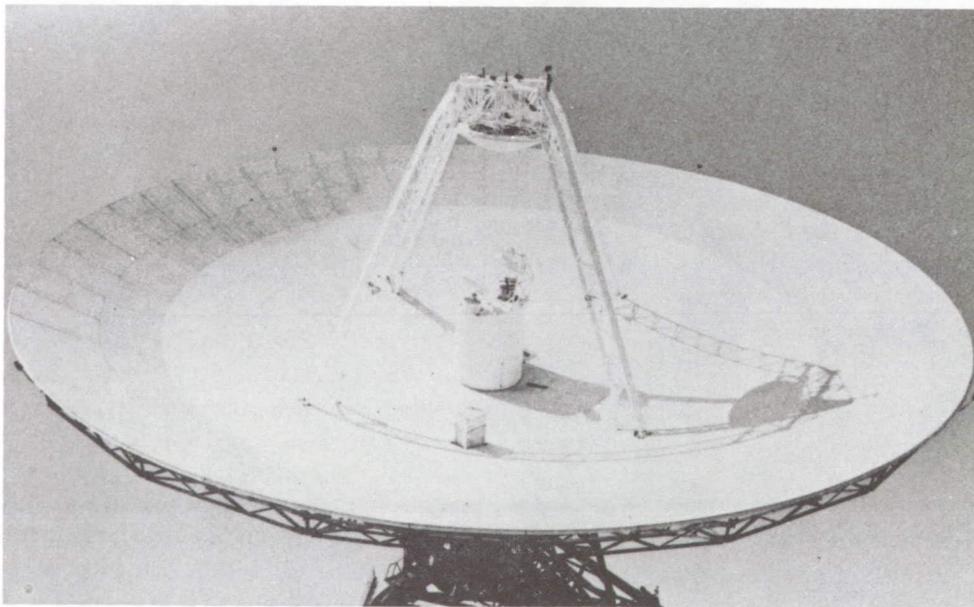


Fig. 10. The S-X reflex feed on the 34-m antenna

Coding for Optical Channels With Photon-Counting

R. J. McEliece and L. R. Welch
Communications Systems Research Section

We study the problem of coding for Pierce's recent model for optical communication. We conclude that for any positive rate ρ (measured in nats per photon), the best code of length n has an error probability bounded by an exponentially decaying function of n ; we exhibit explicit practical schemes for $\rho \leq \sim 1$; and give evidence that $\rho \approx 1$ may be the "practical limit" for optical communication.

I. Introduction

In a recent paper Pierce (Ref. 1) has shown that if one uses photon-counting techniques for communication at optical frequencies, and low noise temperatures, the rate at which one can operate is at most $hf/kt = 4.80 \times 10^{-11} (f/T)$ nats per photon, where f is the photon frequency in hertz and T is the noise temperature in kelvin. This number is typically fairly large; for example $f = 6 \times 10^{14}$ Hz, $T = 400$ K gives a limit of 72 nats per photon. Pierce observed, however, that if one attempts to design a practical communication system for such a channel, one encounters severe problems of coding complexity long before the limiting rate of hf/kT is reached. Moreover, these problems are not caused by thermal noise, but rather by the nature of the photon-counting process itself. We shall now describe Pierce's "noiseless" photon channel, and devote the rest of our article to a discussion of the coding problem for this channel.

We assume that the data to be transmitted consists of a long stream of 0's and 1's: x_1, x_2, x_3, \dots , and that the time interval during which these bits are to be transmitted is divided into small intervals ("slots") of duration t_0 seconds each. The transmitter is a semiconductor laser which is pulsed during the i th time interval if and only if $x_i = 1$. The expected number of photons emitted during such a pulse we denote by

λ . The receiver is a photon counter; it emits a 1 if it is struck by one or more photons in a given time slot, and a 0 if it is not. If the laser is not pulsed, of course it emits no photons; hence a transmitted 0 is always received correctly. On the other hand, if the laser is pulsed, because of the Poisson statistics which govern photon emissions, there is a nonzero probability $e^{-\lambda}$ that no photons will be emitted. Hence a transmitted 1 is received incorrectly as a 0 with probability $e^{-\lambda}$. In effect, then, the photon channel is just the "Z-channel" familiar to information theorists (see Fig. 1), where the crossover probability ϵ is equal to $e^{-\lambda}$.

In Ref. 1 Pierce considered the following coding scheme for this channel. There are n distinct binary codewords, each of length n . Each codeword has a 1 in one coordinate only. Since there are n codewords, each carries $\log n$ nats¹ of information. Furthermore, the transmission of each codeword requires exactly one pulse and hence an average of λ photons. Hence the rate of transmission, measured in nats per photon, is given by $\rho = (\log n)/\lambda$. The probability that the codeword will be received incorrectly is just the probability $e^{-\lambda}$ that no photons will be emitted by the laser during one pulse period. Since $\lambda = \rho^{-1} \log n$, it follows that if we use this

¹Throughout the paper all logarithms will be natural.

scheme at a rate of ρ nats per photon, the error probability will be

$$P_E = n^{-(1/\rho)} \quad (1)$$

Hence for any desired rate of transmission ρ , however large, by choosing n large enough, one can make the receiver error probability arbitrarily small.

In practice, however, Pierce's codes are not of much use, unless ρ is quite small. For example, at $\rho = 1$ nat per photon, and $P_E = 10^{-6}$, Eq. (1) implies that n must be at least 10^6 . The problem is of course that the decay of P_E in n and ρ is very slow. In Ref. 2 it was shown that there exist codes of length n for which P_E decays exponentially as a function of ρ^{-1} , $\rho \rightarrow 0$. In this paper we shall show that for any fixed value of ρ , however large, there exists a sequence of codes of increasing length all having rate at least ρ nats per photon, whose error probabilities decay exponentially as a function of the code length. In Section II we will prove this fact (the main idea of the proof is to use Pierce's coding scheme as a modulation scheme for more complex codes); in Section III we will exhibit some codes which are practical and give good performance for $\rho \leq \sim 1$ nat per photon. (These simple codes will outperform any possible coding strategy if one uses coherent signal amplification rather than photon-counting techniques, because $\rho = 1$ nat/photon is the *capacity* of the optical channel when linear amplification is used (Ref. 1). Finally, in Section IV we will give evidence (but no proof) that it is probably very difficult to design practical systems that operate at rates much larger than 1 nat per photon.

II. An Exponential Error Probability Bound

In this section we will show that there exists a function $E(\rho)$, which is positive for all $\rho > 0$, such that for any $\rho > 0$ there exists a sequence of codes C_1, C_2, C_3, \dots , of increasing block length, each having rate at least ρ nats per photon, such that the decoding error probability for the k th code satisfies

$$P_E \leq \exp[-n_k E(\rho)], \quad (2)$$

where n_k is the length of the k th code.

Thus let $\rho > 0$ be fixed. Choose a positive integer q satisfying

$$q > e^\rho. \quad (3)$$

Having selected q , choose a real number R , $0 < R < 1$, satisfying

$$-\frac{1}{R} \log(1-R) < \rho^{-1} \log q. \quad (4)$$

(This is possible since by Eq. (3) $\rho^{-1} \log q > 1$; and the left side of Eq. (4) is an increasing function of R and approaches 1 as $R \rightarrow 0$).

Now consider a q -ary erasure channel with erasure probability ϵ . This channel has input alphabet $A_q = \{1, 2, \dots, q\}$ and output alphabet $A'_q = A_q \cup \{e\}$, where "e" is an erasure symbol. The channel's transition probabilities are

$$P(y|x) = \begin{cases} 1 - \epsilon & \text{if } x = y \\ \epsilon & \text{if } y = e \\ 0 & \text{otherwise} \end{cases}$$

Thus each symbol is transmitted correctly with probability $1 - \epsilon$, and "erased" with probability ϵ . This channel has capacity $(1 - \epsilon) \log q$, and so according to the noisy-channel coding theorem (see Ref. 3, Chapter 5), for any $R < (1 - \epsilon)$, and any positive integer N there exists a code of length N with at least q^{RN} codewords, whose decoding error probability satisfies

$$P_E \leq \exp[-NE_r(R)], \quad (5)$$

where $E_r(R)$, the random coding exponent for the channel, is a function which is positive for all $R < 1 - \epsilon$.

Our object is now to use these codes, on the "photon Z-channel." To do this we must have a "modulation" scheme which will allow us to transmit symbols from $A_q = \{1, 2, \dots, q\}$ over the Z-channel. The modulation scheme we use is as follows: we assign each symbol from A_q a unique binary vector of length q containing one "1." For example, with $q = 4$ the assignment could be

1 : 1000
2 : 0100
3 : 0010
4 : 0001

(This is exactly Pierce's quantized pulse position modulation described in Ref. 1 and Section I.)

When one of these length q binary patterns is transmitted over the Z-channel, the "demodulator" looks for a "1" among the q received bits; if a "1" is detected in the k th position, the demodulator knows for certain that the transmitted symbol

was k . If, however, no 1's are detected, the demodulator has no idea what was sent and outputs the erasure symbol "e." It follows that when this particular modulation scheme is adopted, the Z-channel becomes a q -ary erasure channel whose erasure probability ϵ is equal to the crossover probability of the Z-channel.

The code of length N has at least q^{RN} codewords and so each codeword conveys at least $RN \log q$ nats of information. Since transmission of each codeword requires exactly N pulses, we are getting at least $R \log q$ nats per pulse. If we are transmitting at rate ρ nats per photon, this means that we need at most,

$$\lambda = \rho^{-1} R \log q \text{ photons per pulse.} \quad (6)$$

Hence if we operate the laser at λ photons per pulse, where λ is given by Eq. (6), our coding scheme will operate at ρ nats per photon. Now according to the Poisson statistics governing photon emission, this means that the erasure probability ϵ for the q -ary erasure channel is

$$\epsilon = q^{-R/\rho} \quad (7)$$

Hence if the code's (q -ary) rate R is less than the channel's (q -ary) capacity $1 - \epsilon$, we can get exponential decay of the error probability. This condition is

$$R < 1 - q^{-R/\rho}, \quad (8)$$

which is just a restatement of Eq. (4). Hence it follows that for every N there exists a code of rate R , whose error probability satisfies Eq. (5). This code, viewed as a code for the Z-channel, has length $n = qN$, and so its error probability satisfies

$$P_E \leq \exp[-nE_r(R)/q]. \quad (9)$$

Since for fixed $\rho > 0$, and q, R satisfying Eqs. (3) and (4) we have $E_r(R)/q > 0$, it follows that there exists a function $E(\rho)$ such that

$$P_E \leq \exp[-nE(\rho)], \quad (10)$$

which is what we set out to prove.

We have made no attempt to give numerical bounds on $E(\rho)$; it would be interesting to do this, however.

In the next section, we will show how the technique just described can be used to design practical coding schemes for the photon channel in the range $\rho \leq \sim 1$ nats/photon:

III. Practical Schemes for $\rho \leq \sim 1$

In this section we will give explicit examples of the coding schemes described in Section II; these will turn out to be easily implemented and have low error probabilities, provided ρ is not much larger than 1 nat/photon.

As before, for a given rate ρ , choose a positive integer $q > \exp(\rho)$. Using the modulation scheme described in Section II, the Z-channel becomes a q -ary erasure channel, and we must design a code over the alphabet A_q capable of correcting as many erasures as possible. An obvious choice is the class of Reed-Solomon codes, which require only that q be a power of a prime.

Over the alphabet A_q (which we now view as a finite field), Reed-Solomon codes are linear cyclic codes of length $n = q - 1$, arbitrary dimension $1 \leq k \leq n - 1$, with minimum distance $d = n - k + 1$. They will correct any pattern of up to $n - k$ erasures, and very efficient encoding and decoding procedures are known, when q is a power of 2 (see Ref. 5, Chapter 8 for details).

If we use an (n, k) Reed-Solomon code for the present application, each of the q^k codewords carries $k \log q$ nats of information, and each codeword requires n pulses. Thus if we are transmitting at ρ nats/photon, the average number of photons per pulse is at most (cf. Eq. (6)).

$$\lambda = \frac{k}{n} \cdot \rho^{-1} \log q \text{ photons per pulse.} \quad (11)$$

It follows that the erasure probability for the corresponding q -ary erasure channel is (at worst)

$$\epsilon = e^{-\lambda} = q^{-R/\rho}, \quad (11a)$$

where $R = k/n$ is the code rate. Since the RS code can correct all patterns of up to $n - k$ erasures, it follows that the decoding error probability P_E satisfies

$$P_E \leq \sum_{j=n-k+1}^n \binom{n}{j} \epsilon^j (1 - \epsilon)^{n-j}. \quad (12)$$

In Fig. 2 we have plotted P_E vs ρ for four typical RS codes. The curve labelled $q = 16$ is a (15,8) RS code with $q = 16$; the others are (31,16) $q = 32$; (63,32), $q = 64$; and (127,64), $q = 128$. It is seen that the performance of all of these codes degrades rapidly as ρ increases beyond 1. (It is to be recalled that as codes for the photon channel, the length is actually $16 \cdot 15 = 240$ for the $q = 16$ codes; $n = 31 \cdot 32 =$

992 for the $q = 32$ code; $n = 4032$ for $q = 64$; and $n = 16256$ for $q = 128$.) Each of these codes is the best of its length, at least in the limit as $\rho \rightarrow 0$, and so no significant improvement could be expected merely by altering the code's rate. We conclude that for $\rho \leq \sim 1$ practical schemes exist. In the next section we give evidence that $\rho \geq \sim 1$ may be very hard to achieve.

IV. R_0 for the Photon-Detection Channel

Associated with any discrete memoryless channel there is a number C , called channel capacity, which represents the theoretical maximum rate at which one can communicate reliably over the channel. However, it is usually extremely difficult to operate at rates very close to C , and it has been conjectured (see e.g., Ref. 4) that another number, R_0 , which is associated with every discrete memoryless channel, is a more realistic measure of the maximum rate of practical reliable communication. In this section we shall show that for our photon-counting Z-channel, the value of ρ corresponding to R_0 is exactly 1 nat/photon; this number, incidentally, is the limiting value (capacity) for coherent optical communication systems (Ref. 1).

Given a discrete memoryless channel with input alphabet A_X , output alphabet A_Y , and transition probabilities $p(y|x)$, we can state the " R_0 -conjecture" as follows. "Any practical communication system, in which the relative frequencies with which the input symbols $x \in A_X$ are used are described by the probability distribution $p(x)$, must operate at a rate R satisfying

$$R < \sim -\log \sum_{y \in A_Y} \left[\sum_{x \in A_X} p(x) \sqrt{p(y|x)} \right]^2 = R_0(p). \quad (13)$$

In the specific case of the Z-channel with erasure probability ϵ , with $p(x=0) = 1 - \delta$, $p(x=1) = \delta$, the quantity on the right side of Eq. (13) is given by

$$R_0(\delta) = -\log [1 - 2\delta(1 - \delta)(1 - \epsilon^{1/2})]. \quad (14)$$

If we are using a code of length n and rate R for the photon-counting Z-channel, in which the average number of 1's per codeword is δn , a calculation similar to that which led to Eq. (16) and Eq. (11) shows that the average number of photons per pulse must be

$$\delta = R/\rho \delta \text{ photons per pulse.} \quad (15)$$

Thus the erasure probability for the Z-channel is

$$\epsilon = \exp(-R/\rho\delta) \quad (16)$$

and the inequality Eq. (13) becomes

$$R < -\log \{1 - 2\delta(1 - \delta) [1 - \exp(-R/2\rho\delta)]\} \quad (17)$$

Our result is the following.

Theorem: There exist real numbers (R, δ) $0 < R < \log 2$, $0 < \delta < 1$ satisfying Eq. (17) if and only if $\rho < 1$. Hence any system for the photon-counting Z-channel for which $\rho \geq 1$ must violate Eq. (13).

Proof: First assume $\rho < 1$, and define $\alpha = R/2\delta$. Then $\alpha > 0$ and Eq. (17) becomes

$$2\alpha\delta < -\log \{1 - 2\delta(1 - \delta) [1 - \exp(-\alpha/\rho)]\}. \quad (18)$$

Now choose α satisfying

$$\alpha < 1 - e^{-\alpha/\rho} \quad (19)$$

(This is possible since $\rho \leq 1$.) For this fixed value of α the right side of Eq. (18) is equal to $2\delta(1 - e^{-\alpha/\rho}) + O(\delta^2)$. It follows that Eq. (18) is satisfied for all sufficiently small $\delta > 0$. Thus with $R = 2\alpha\delta$ and δ small enough, Eq. (17) is satisfied.

Now assume $\rho \geq 1$. Then $\exp(-\alpha/\rho) \geq \exp(-\alpha)$, and the right side of Eq. (18) is bounded as follows:

$$\begin{aligned} & -\log \{1 - 2\delta(1 - \delta)[1 - \exp(-\alpha/\rho)]\} \\ & \leq -\log \{1 - 2\delta(1 - \delta)[1 - \exp(-\alpha)]\} \end{aligned} \quad (20)$$

If we replace the right side of Eq. (18) by its upper bound (Eq. 20)), a little rearrangement yields

$$\frac{1 - \exp(-2\alpha\delta)}{1 - \exp(-\alpha)} < 2\delta(1 - \delta). \quad (21) \quad \text{(The proof of the lemma is elementary and is omitted.)}$$

Lemma: Let $k > 0$, $f(y) = (1 - y^k)/(1 - y)$. Then

$$\inf_{0 < y < 1} f(y) = \begin{cases} k & \text{for } k < 1 \\ 1 & \text{for } k \geq 1. \end{cases}$$

Now let $y = \exp(-\alpha)$, $k = 2\delta$. Then the left side of Eq. (21) is $(1 - y^k)/(1 - y)$. By the lemma, this is $\geq 2\delta$ if $2\delta < 1$; thus Eq. (21) implies $2\delta < 2\delta(1 - \delta)$, a contradiction. If $2\delta \geq 1$, then Eq. (21) gives $1 \leq 2\delta(1 - \delta)$, another contradiction, since $2\delta(1 - \delta) \leq 1/2$ for $0 < \delta < 1$. This completes the proof of the theorem.

References

1. Pierce, J. R., "Optical Channels: Practical Limits with Photon Counting," *IEEE Trans. Comm.*, COM-26 (1978), pp. 1819-1821.
2. Baumert, L. D., McEliece, R. J., and Rumsey, H. C., Jr., "Coding for Optical Channels," *DSN Progress Report 42-49*, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1979, pp. 70-77.
3. Gallager, R. G., *Information Theory and Reliable Communication*. Wiley, New York, 1968.
4. Massey, J. L., "Coding and Modulation in Digital Communications," *Proc. International Zurich Seminar on Digital Communications*, 1974.
5. McEliece, R. J., *The Theory of Information and Coding*. Addison-Wesley, Reading, Mass, 1977.

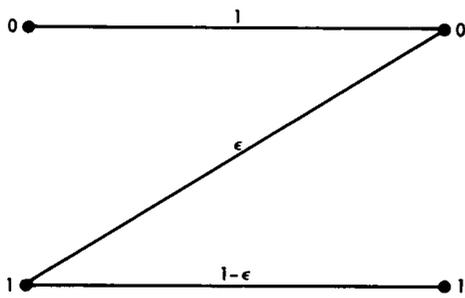


Fig. 1. The Z-channel

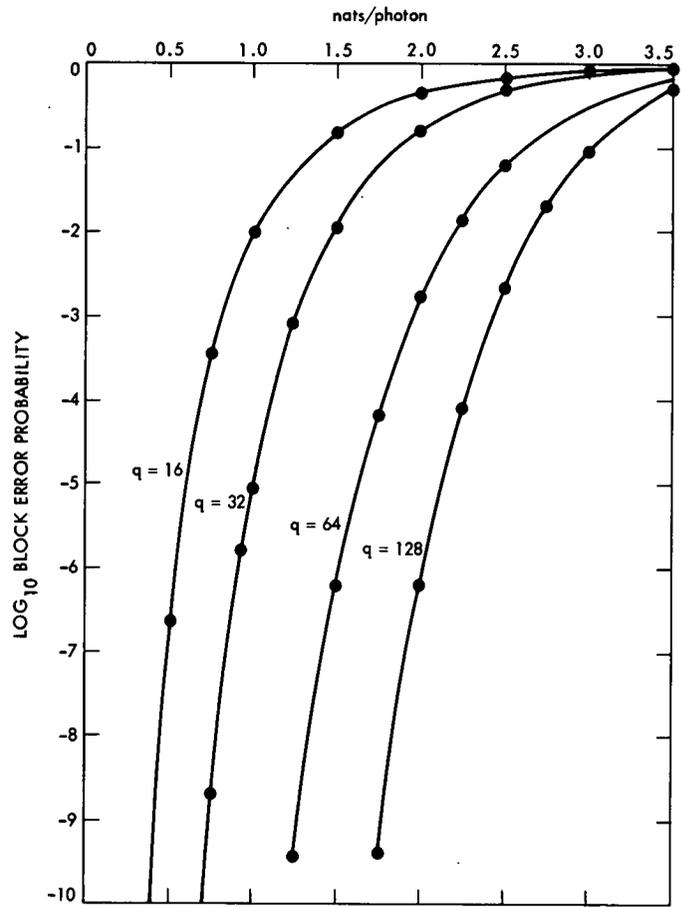


Fig. 2. Performance of some RS codes on the photon channel

A Fast Technique for Computing Syndromes of BCH and RS Codes

I. S. Reed

Department of Electrical Engineering
University of Southern California

T. K. Truong, R. L. Miller

Communication Systems Research Section

In this article, a combination of the Chinese Remainder Theorem and Winograd's algorithm is used to compute transforms of odd length over $GF(2^m)$. Such transforms are used to compute the syndromes needed for decoding BCH and RS codes. The present scheme requires substantially fewer multiplications and additions than the conventional method of computing the syndromes directly.

I. Introduction

It is shown in Ref. 1 that the finite field transform can be used to compute the syndromes for BCH and RS codes. The disadvantage of the transform method over $GF(2^m)$ is that the transform length is an odd number, so that the most efficient FFT algorithm cannot be used. In this article a combination of the Chinese Remainder Theorem and Winograd's algorithm is used to develop a fast finite field transform over $GF(2^m)$. Such a fast transform can be used to compute the syndromes needed when decoding BCH and RS codes (Ref. 2). It is also shown by example that the number of multiplications and additions of this new scheme for computing the syndromes is substantially fewer than used in the direct method.

II. The Computation of the Syndromes for BCH and RS Codes

Let n be the block length of a BCH or RS code of designed distance d in $GF(2^m)$. Also denote the received vector by $\mathbf{r} = (r_0, r_1, \dots, r_{n-1})$. If α^j $1 \leq j \leq d-1$ are roots of the code's

generator polynomial, then the decoder needs to evaluate the syndromes S_j for $1 \leq j \leq d-1$ from the received vector by

$$S_j = \sum_{i=0}^{n-1} r_i \alpha^{ij} \text{ for } 1 \leq j \leq d-1 \quad (1)$$

Evidently, a direct computation of S_j for $1 \leq j \leq d-1$ involves $(d-1)(n-1)$ multiplications and $(d-1)(n-1)$ additions. Using the Chinese Remainder Theorem, it is shown in this section that this number of multiplications and additions can be reduced substantially. To see this, if $n = n_1 n_2 \dots n_r$, where $(n_i, n_j) = 1$ for $i \neq j$, then by the Chinese Remainder Theorem it follows (Refs. 3, 4) that the n -point transform over $GF(2^m)$ defined in (1) can be decomposed into a multidimensional transform over $GF(2^m)$ as follows:

Given an integer j , define $f(j) = (j_1, j_2, \dots, j_r)$, where $j_k = j \bmod n_k$ ($1 \leq k \leq r$). Then $\alpha_k = \alpha^{(0,0, \dots, 0, 1, 0, \dots, 0)}$, where 1 is in the k^{th} position, is a primitive n_k^{th} root of

unity. The computation of S_j given by (1) can now be written as

$$\begin{aligned}
 S_j &= S_{(j_1, j_2, \dots, j_r)} \\
 &= \sum_{i_1=0}^{n_1-1} \sum_{i_2=0}^{n_2-1} \dots \sum_{i_r=0}^{n_r-1} a_{(i_1, i_2, \dots, i_r)} \\
 &\quad \alpha_1^{i_1 j_1} \alpha_2^{i_2 j_2} \dots \alpha_r^{i_r j_r} \\
 &\text{for } 1 \leq j \leq d-1
 \end{aligned} \tag{2}$$

From Eq. (2) this algorithm consists of the following r stages:

Stage 1:

$$\begin{aligned}
 S^1_{(i_1, i_2, \dots, i_r)} &= \sum_{i_r=0}^{n_r-1} a_{(i_1, i_2, \dots, i_r)} \alpha_r^{i_r j_r} \\
 &\text{for } 0 \leq j_r \leq n_r - 1
 \end{aligned}$$

Stage 2:

$$\begin{aligned}
 S^2_{(i_1, i_2, \dots, i_{r-1}, j_r)} &= \sum_{i_{r-1}=0}^{n_{r-1}-1} S^1_{(i_1, i_2, \dots, i_r)} \alpha_{r-1}^{i_{r-1} j_{r-1}} \\
 &\text{for } 0 \leq j_{r-1} \leq n_{r-1} - 1
 \end{aligned}$$

⋮

Stage r :

$$\begin{aligned}
 S_j &= S^r_{(j_1, \dots, j_{r-1}, j_r)} = \sum_{i_1=0}^{n_1-1} S^{r-1}_{(i_1, j_2, \dots, j_r)} \alpha_1^{i_1 j_1} \\
 &\text{for } 1 \leq j \leq d-1
 \end{aligned} \tag{3}$$

Assume the number of multiplications and additions needed to perform the n_i -point transform for $1 \leq i \leq r-1$ is m_i and a_i , respectively. Then by Eq. (3) it is evident that the number of multiplications and additions needed to compute Stage 1 up to Stage $(r-1)$ is given by

$$M = n \sum_{i=1}^{r-1} \frac{m_i}{n_i} \text{ and } A = n \sum_{i=1}^{r-1} \frac{a_i}{n_i}, \text{ respectively.}$$

To compute the syndromes one needs only to compute the first d points of the transform over $GF(2^m)$ defined in Eq. (1). Thus, at the Stage r in Eq. (3), one needs only to compute $S_j = S_{(j_1, j_2, \dots, j_r)}$ for $1 \leq j \leq d-1$. Hence the number of both multiplications and additions needed to compute Stage r in (3) is $(d-1)(n_1-1)$. Combining this with the above results the total number of multiplications and additions needed to compute (1) is finally $M + (d-1)(n_1-1)$ and $A + (d-1)(n_1-1)$, respectively.

Suppose that the transform length given in Eq. (3) of the first $r-1$ stages is small. Then it was shown (Ref. 1) that such short length transforms can often be computed by a modification of Winograd's algorithm. Algorithms for computing a transform over $GF(2^m)$ of n points for $n = 3, 5, 7, 9, 17$ are given in detail in Ref. (1). The detailed algorithms for computing the n -point transform over $GF(2^m)$ for $n = 3, 5$ are given in the appendix. These transforms are used in the following example to compute a transform of 2^8-1 points.

Example 1: Let $n = 255$ be the block length of an RS code of designed distance $d = 33$. This code will correct t errors where $2t < 33$. The first step of the decoding process is to compute the 32 syndromes as follows:

$$S_j = \sum_{i=0}^{255-1} r_i \alpha^{ij} \text{ for } 1 \leq j \leq d-1 = 32 \tag{4}$$

where α is an element of order 255 in $GF(2^8)$. If Eq. (4) is computed directly, the number of multiplications and additions needed is $32(255-1) = 8128$ and $32(255-1) = 8128$, respectively. Since $n = 255 = n_1 n_2 n_3 = 17 \cdot 5 \cdot 3$, (4) for this case reduces to

Stage 1:

$$S^1_{(i_1, i_2, j_3)} = \sum_{i_3=0}^{3-1} a_{(i_1, i_2, i_3)} \alpha_3^{i_3 j_3} \text{ for } 0 \leq j_3 \leq 2$$

Stage 2:

$$S_{(i_1 j_2 j_3)}^2 = \sum_{i_2=0}^{5-1} S_{(i_1, i_2, j_3)}^1 \alpha_2^{i_2 j_2} \text{ for } 0 \leq j_2 \leq 4$$

Stage 3:

$$S_j = S_{(i_1 j_2 j_3)}^3 = \sum_{i_1=0}^{17-1} S_{(i_1, j_2, j_3)}^2 \alpha_1^{i_1 j_1} \text{ for } 1 \leq j \leq 32 \quad (5)$$

In Eq. (5) Stage 1 is a 3-point transform and Stage 2 is a 5-point transform. These two transforms are computed by a

modification of Winograd's algorithm. It is shown in the appendix that the number of multiplications and additions needed to compute the 3-point transform is 1 and 5, respectively. Thus, the number of multiplications and additions needed to compute Stage 1 is $(17)(5)(1) = 85$, and $(17)(5)(5) = 425$, respectively. Again by the appendix one observes that the number of multiplications and additions needed to compute a 5-point transform is 5 and 17, respectively. Thus, the number of multiplications and additions needed to compute Stage 2 is $(17)(3)(5) = 255$ and $(17)(3)(17) = 867$, respectively. Since one needs to compute S_j only for $j = 1, 2, \dots, 32$, the number of multiplications and additions needed to compute Stage 3 is $32(17 - 1) = 512$ and $32(17 - 1) = 512$, respectively. Thus the total number of multiplications and additions needed to compute the syndromes S_j for $1 \leq j \leq 32$ is $85 + 255 + 512 = 852$, and $425 + 867 + 512 = 1804$, respectively.

Appendix

A Summary of Transform Algorithm of n -Points for $n = 3, 5$

For $n=3$, let γ be a primitive cube root of unity in $GF(2^m)$, where $3|(2^m - 1)$. The 3-point transform over $GF(2^m)$ is

$$A_k = \sum_{i=0}^{3-1} a_i \gamma^{ik} \quad \text{for } 0 \leq k \leq 2.$$

Algorithm for $n=3$:

$$s_1 = a_1 + a_2, s_2 = s_1 + a_0, m_0 = 1 \cdot s_2;$$

$$m_1 = s_1 \cdot \gamma^1, m_2 = 1 \cdot a_1, m_3 = 1 \cdot a_2;$$

$$s_3 = m_0 + m_1, s_4 = s_3 + m_2, s_5 = s_3 + m_3;$$

$$A_0 = m_0, A_1 = s_4, A_2 = s_5 \quad (\text{A-1})$$

Hence, from Eq. (A-1), the total number of multiplications and additions needed to perform a 3-point transform is 1 and 5, respectively.

Next consider the case $n=5$. Let γ be a primitive 5th root of unity in $GF(2^m)$, where $5|(2^m - 1)$. The 5-point transform is

$$A_k = \sum_{i=0}^{5-1} a_i \gamma^{ik} \quad \text{for } 0 \leq k \leq 4$$

Algorithm for $n=5$:

$$s_1 = (a_2 + a_3), s_2 = a_1 + a_4, s_3 = a_1 + a_3;$$

$$s_4 = a_2 + a_4, s_5 = s_1 + s_2, s_6 = s_5 + a_0;$$

$$m_0 = 1 \cdot s_6, m_1 = (\gamma^0 + \gamma^3) \cdot s_5, m_2 = (\gamma^3 + \gamma^4) \cdot s_1;$$

$$m_3 = (\gamma^1 + \gamma^3) \cdot s_2, m_4 = (\gamma + \gamma^4) \cdot s_3, m_5 = (\gamma + \gamma^4) \cdot s_4;$$

$$m_6 = 1 \cdot a_1, m_7 = 1 \cdot a_3, m_8 = 1 \cdot a_4, m_9 = 1 \cdot a_2;$$

$$s_7 = m_0 + m_1, s_8 = s_7 + m_2, s_9 = s_7 + m_3, s_{10} = m_5 + m_9;$$

$$s_{11} = m_4 + m_6, s_{12} = m_5 + m_8, s_{13} = m_4 + m_7, s_{14} = s_9 + s_{10};$$

$$s_{15} = s_8 + s_{11}, s_{16} = s_8 + s_{12}, s_{17} = s_9 + s_{13}, A_0 = m_0;$$

$$A_1 = s_{14}, A_2 = s_{15}, A_3 = s_{16}, A_4 = s_{17}. \quad (\text{A-2})$$

Hence, from Eq. (A-2), the total number of multiplications and additions needed to perform a 5-point transform is 5 and 17, respectively.

References

1. Reed, I. S., Truong, T. K., Miller, R. L., and Benjauthrit, B., "Further Results on Fast Transforms for Decoding Reed-Solomon Codes over $GF(2^m)$ for $m = 4, 5, 6, 8$," in *The Deep Space Network Progress Report 42-50*, pp. 132-154, Jet Propulsion Laboratory, Pasadena, Calif., Jan. 15, 1979.
2. Berkekamp, E. R., *Algebraic Coding Theory*, New York, McGraw Hill, 1968.
3. Winograd, S., "On Computing the Discrete Fourier Transform," *Proc. Nat. Acad. Sci. USA*, Vol. 73, 1976, pp. 1005-1006.
4. Reed, I. S., and Truong, T. K., "Fast Mersene-Prime Transforms For Digital Filtering," *Proc. IEE*, Vol. 125, No. 5, May 1978, pp. 433-440.

New X-Band Antenna Feeds for the DSN 64-Meter Stations

R. Hartop

Radio Frequency and Microwave Subsystems Section

New X-band antenna feed assemblies with dual-polarization capability are being implemented in the DSN 64-meter stations. Together with dual X-band travelling wave masers, they permit the simultaneous reception of right- and left-hand circular polarization from the Voyager spacecraft. The new feed also includes a dual hybrid mode feedhorn which increases the antenna gain by 0.36 dB over the present feedhorn.

I. Introduction

The X-band feed cone assembly (Ref. 1) on the DSN 64-meter antennas provides the capability for selectable right-hand (RCP) or left-hand (LCP) circular polarization (Ref. 2). Originally designed to work with a single travelling wave maser (TWM), the XRO cone assembly has recently been upgraded to include dual TWM's (Ref. 3). As part of that reconfiguration, provision was made for a new feed assembly, the Mod III XRO, that would include two basic improvements over the feeds that have been in use since early 1977. The first improvement is the development of an orthogonal-mode transducer that permits simultaneous reception of two different polarizations. The second improvement is the incorporation of a dual hybrid mode feedhorn (Ref. 4) to increase the antenna gain.

II. The Mod III XRO Feed Assembly

The new feed assembly is shown in Figs. 1 and 2. It is designed to mount within the XRO cone assembly on the same

circular mounting plate as the previous (Mod I and Mod II) feeds, so that replacement in the field is expedited. Each output arm of the orthogonal mode transducer has its own waveguide switch for TWM calibration. One of the precision ambient terminations may be seen at the bottom of the feed assembly.

The new orthogonal-mode transducer is shown in Figs. 3 and 4. The unit is based on a design by Microwave Development Laboratories, developed under contract to JPL, with the final units fabricated by the electroforming process. The unit has an isolation of 35 dB or better over the frequency band of 8400 to 8500 MHz, with the return losses shown in Table 1.

III. Noise Temperature Measurements

Using a vertically oriented single hybrid mode (corrugated) horn on top of a TWM for reference, the noise temperature contribution of the Mod III XRO feed was determined in ground tests. By disconnecting the reference horn at the TWM

input and connecting the Mod III XRO feed at the same point, variables such as weather, smog, TWM noise temperature, receiver followup noise temperature, etc., are compared in a relative measurement and the loss contribution of the added feed components can be determined with high precision (est. ± 0.15 K). From the throat of the horn to the TWM interface, the following components contribute to the additional noise temperature compared to the simple horn: rotary joints (2 ea), quarter-wave plate polarizer, orthogonal mode transducer, waveguide twist (in the side path measured), and waveguide switch. The measurements were performed on two different feeds (for DSS 43 and DSS 63) and the resulting 2.43 K increase (over the simple horn) was the same for each feed within experimental tolerances. This indicates that the 64-meter antenna system temperature at zenith should not exceed 25 K, including the dichroic plate losses and the normal antenna and receiver noise contributions.

IV. DSN Implementation

The first Mod III XRO feed assembly has been installed at DSS 14, and measurements (Ref. 5) have confirmed 0.36-dB antenna gain increase. For reasons yet to be determined, the system noise temperature is indicating several kelvins too high. Since the system noise temperature was also too high before the feed change, the problem is thought to be elsewhere, and efforts are underway to locate and correct the problem. Nevertheless, excellent data was received from Voyager 1 as it passed through the Jovian system, verifying the feed capability.

The second and third Mod III feeds have been fabricated and tested and, together with some additional parts for the cone assembly, will be implemented at the overseas sites in late 1979 and early 1980.

References

1. Hartop, R., "X-Band Antenna Feed Cone Assembly," TR 32-1526, Vol XIX, Jet Propulsion Laboratory, Pasadena, Calif., pp. 173-5, 1974.
2. Hartop, R., "Selectable Polarization at X-Band," *DSN Progress Report* 42-39, Jet Propulsion Laboratory, Pasadena, Calif., pp. 177-80, 1977.
3. Hartop, R., "New X-Band Microwave Equipment at the DSN 64-Meter Stations," *DSN Progress Report* 42-48, Jet Propulsion Laboratory, Pasadena, Calif., pp. 126-28, 1978.
4. Thomas, R., and Bathker, D., "A Dual Hybrid Mode Feedhorn for DSN Antenna Performance Enhancement," *DSN Progress Report* 42-22, Jet Propulsion Laboratory, Pasadena, Calif., pp. 101-8, 1974.
5. Freiley, A. J., "Radio Frequency Performance of DSS-14 64 m Antenna at X-Band Using a Dual Hybrid Mode Feed" to be published in *DSN Progress Report* 42-53.

Table 1. Impedance match characteristics of X-band orthogonal mode transducer

Frequency, MHz	Return loss, dB	
	Thru port	Side port
8400	-24.0	-28.0
8425	-25.5	-33.5
8450	-28.5	< -40.
8475	-29.0	-32.0
8500	-26.0	-27.0

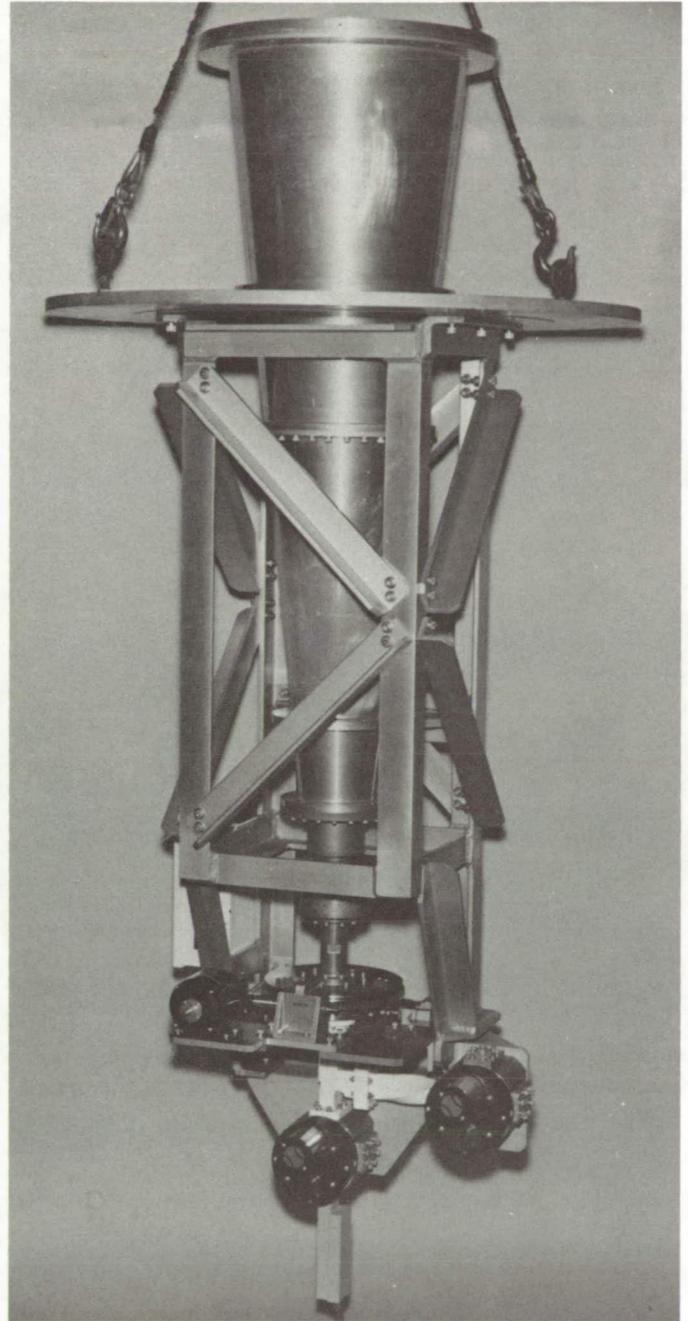


Fig. 1. XRO Mod III feed assembly

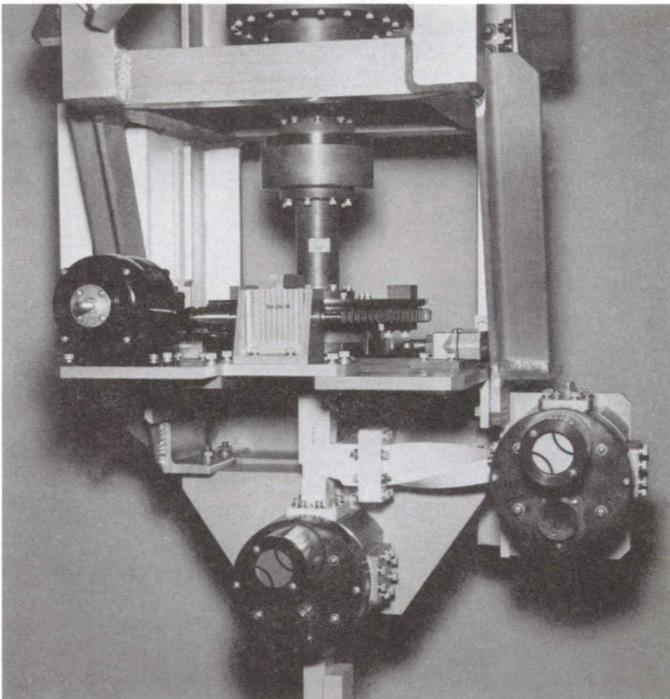


Fig. 2. Close-up of feed assembly showing orthogonal mode transducer and switch

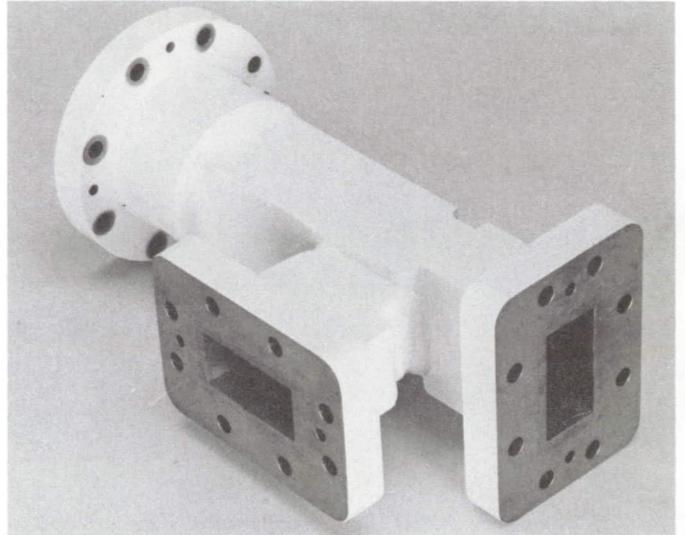


Fig. 3. Orthogonal mode transducer showing output flanges

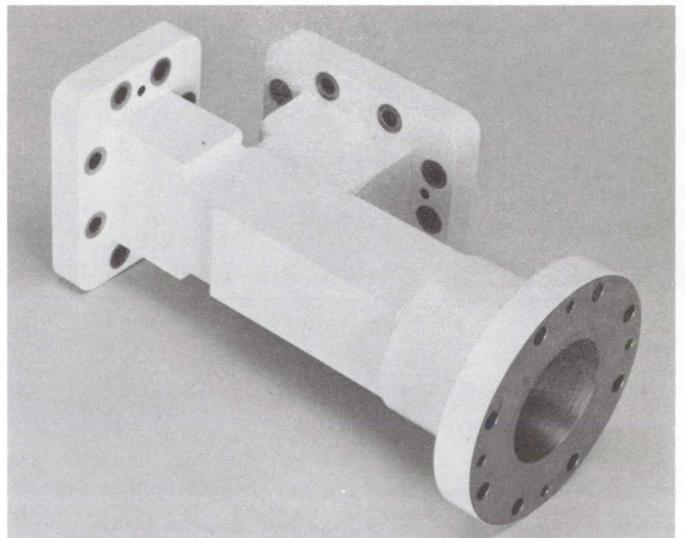


Fig. 4. Orthogonal mode transducer showing input flange

Multi-Mission Receiver (MMR)

H. Donnelly and H. Nishimura
Radio Frequency and Microwave Subsystems Section

An introduction to the Multi-Mission Receiver (MMR) is presented. The MMR contains occultation receiver channels (OCC) and very long baseline interferometry (VLBI) channels. The design considerations, performance and implementation of the OCC channels are discussed. The VLBI channels will be discussed in a subsequent report.

I. Introduction

Measurement of the signature of radio signals for obtaining radio science, radio astronomy and tracking data requires extremely stable receiver channels. The Multi-Mission Receiver (MMR) is being developed to provide this capability.

There are two different types of signal signatures that are involved in radio science and in radio astronomy and tracking.

- (1) *Occultation channels.* The first type are narrowband spacecraft transmitted carrier signals. Occultation (OCC) channels used to receive these signals measure the effect of planetary and solar media, interplanetary media, and gravitational fields by detecting minor perturbations on these received signals. Real-time reduction techniques are used in these channels to minimize the data processing costs. This is accomplished by programming the receiver local oscillator based on signal frequency prediction.
- (2) *Very long baseline interferometry channels.* The second type are wide-bandwidth signals received on the very long baseline interferometry (VLBI) channels.

To support these measurements, the MMR contains both OCC and VLBI channels (Fig. 1). Functions such as control, instrumentation and power are common to both sets of channels. The functional requirements and performance, together with a detailed description of the OCC channels, are discussed in this report. The VLBI channels will be discussed in a later report.

II. MMR Occultation Channels

A. General Description

Four occultation channels are being implemented (Fig. 2), right and left hand circularly polarized channels at both S- and X-band. The right-hand channels are equipped with very narrow, narrow and medium bands, while the left-hand channels are equipped with the medium band only. The very narrow-band channels are to be used for measurement of solar media and the narrow-band channels for measurement of planetary media. The medium bands of both the right- and left-hand channels are to be used for measurements during the Saturn ring experiment. In the future, a selection of narrow and very

narrow bandwidths might also be included in the left-hand channel. The selection of bandwidths available (Table 1) provides for all presently known requirements.

B. Design Considerations

The most significant features of the occultation channels are (1) a wide selection of operating bandwidths, (2) the use of one common oscillator to derive the first local oscillator of all channels, and (3) extreme phase stability. These features are discussed in more detail as follows:

1. Receiver bandwidth selection. The selection of the operating bandwidth is determined by such characteristics of the received signal as frequency uncertainty and spectral broadening. If a frequency uncertainty exists, the uncertainty at X-band is 11/3 that at S-band. In this situation, narrow and medium bandwidth channels are used and these filter bandwidths are designed with a S- to X-band ratio of approximately 3 to 11. If the signal being tracked has essentially no frequency uncertainty, then the filter bandwidth for both S- and X-bands can be identical, the bandwidth required being determined by spectral broadening. This is the case for the very narrow bandwidth channels.

a. Narrow and medium bandwidth channels. A typical filter bandpass response at the 10-MHz IF amplifier is shown in Fig. 3. This response translates directly down to video frequencies. The local oscillator synthesizer is programmed to maintain the signal as close to the filter bandwidth center as the accuracy of the spacecraft's predicted frequency permits. The filters have been designed so that overflow into the operating bandwidth due to image noise in the last mixer or aliasing noise due to digitizing will be at least 20 dB down. This rejection will be increased to 23 dB on future implementations.

b. Very narrow bandwidth channels. In contrast to the medium and narrow bands, the very narrow filter bandwidths are the same for S and X. However, since the X-band signal in the IF is offset from the IF center frequency 11/3 as much as the S-band, the filter center frequencies for S and X differ (Fig. 4a). The passband at video frequencies for both the S- and X-bands do not go down to dc (Fig. 4b); however, with undersampling techniques aliasing causes the passband to appear as though it goes down to dc (Fig. 4c) (Ref. 1). With this technique a sampling rate of no greater than 2 BW is required.

2. Common oscillator. The mechanization of the local oscillators is similar to the technique used in the Block III open loop receiver; however, the first IF is designed for 300 MHz to accommodate the X-band channels. The oscil-

lator source is a Dana Synthesizer. With the control assembly, the Dana Synthesizer frequency can be programmed to compensate for doppler and maintain the signal within the receiver narrow bandwidth. At present, the controller interfaces directly with the computer in the Radio Science Subsystem. At a latter date, when the MMR controller is implemented, the synthesizer control will interface through the Star Switch Controller to the Radio Science Subsystem.

The synthesizer frequency (Fig. 5) required to maintain the signal within the passband for a narrow or medium bandwidth S-band channel is

$$f_S - f_{LO_S} = 300 \text{ MHz} + \frac{BW_S}{2}$$

where

$$f_S = \text{S-band downlink frequency}$$

$$f_{LO_S} = \text{S-band local oscillator frequency}$$

$$= f_S - 300 - \frac{BW_S}{2}$$

$$= 48 f_{DANA}$$

$$\frac{BW_S}{2} = \frac{1}{2} (\text{S-band sampling bandwidth})$$

$$f_{DANA} = \text{synthesizer frequency}$$

$$= \frac{f_S - 300 \text{ MHz} - \frac{BW_S}{2}}{48} \quad (1)$$

Likewise for the X-band channel:

$$f_X - f_{LO_X} = 300 \text{ MHz} + \frac{BW_X}{2}$$

$$f_X = \text{X-band downlink frequency}$$

$$f_{LO_X} = \text{X-band local oscillator frequency}$$

$$= f_X - 300 - \frac{BW_X}{2}$$

$$= 11 (16 f_{DANA} + \frac{800}{11} \text{ MHz})$$

$$\frac{BW_X}{2} = \frac{1}{2} (\text{X-band sampling bandwidth})$$

$$f_{DANA} = \frac{f_X - 1100 \text{ MHz} - \frac{BW_X}{2}}{11 \times 16}$$

or since $f_X = 11/3 f_S$ and $BW_X = 11/3 BW_S$

$$f_{DANA} = \frac{f_S - 300 - \frac{BW_S}{2}}{48} \quad (2)$$

Equation (2) is identical to (1), demonstrating that the common Dana Synthesizer oscillator can handle both S- and X-band channels. This can also be shown for the very narrow bandwidth channel.

3. **Improved phase stability.** The most significant parameter of the received signal is the information contained in its phase. To improve the accuracy of the phase data, the receiver phase instability is reduced to a minimum by enclosing the most phase-sensitive portions of the receiver (local oscillator multipliers) in a temperature-controlled assembly. Phase stability requirements are listed in Table 1.

III. Operations

Two functional types of control are required, configuration and operation.

A. Configuration Controls

To set the receiver in the required configuration, it is necessary only to select the desired operating filter bandwidths. The filter select switch is a manual control and must be selected locally. This mode of operation will be required until the MMR controller is implemented, which will occur when the VLBI channels are added and the MMR implementation is completed.

B. Operation Controls

The following operation controls are required: (1) The adjustment of the output video signal level (which is a function of the received signal strength), and (2) the programming of the local oscillator frequency to track the incoming signal.

1. **Output level adjustment.** The video output channels of the receiver are digitized on a converter with an input voltage range of ± 5 V. Under strong signal-to-noise ratios the signal output level is set to 5 (0.707) V_{rms} or 3.54 V_{rms} (+24 dBm into 50 ohms). Likewise, with a low signal-to-noise ratio, the output is essentially noise and, allowing for 5 \times sigma noise peaks, the output level is set at 5/5 or 1 V_{rms} (+13 dBm into 50 ohms). For any ratio of signal-to-noise between these two values, the output level is set in accordance with the graph of Fig. 6. The output level adjustment is made by locally setting manually controlled IF attenuators.

2. **Local oscillator frequency programming.** The other control required to operate the OCC channels is the local oscillator frequency programming. The computer in the Radio Science Subsystem is used to program the frequency synthesizer of the local oscillator. No operator intervention at the receiver is required.

IV. Implementation

The occultation channels are being implemented in phases as required to support mission tracking requirements. Table 2 shows a summary of the implementation schedule.

Reference

1. Gregg, W. David, *Analog and Digital Communication*, John Wiley and Sons, New York, pp. 286-294.

Table 1. Occultation channels

Sampling bandwidth (BW), kHz	S-band		X-band	
	RCP	LCP	RCP	LCP
Very narrow BW	0.1	–	0.1	–
	0.5	–	0.5	–
	1.0	–	1.0	–
Narrow BW	0.82	–	3.0	–
	2.04	–	7.5	–
	4.09	–	15.0	–
	8.18	–	30.0	–
Medium BW	50	50	150	150
Phase instability, 1000 s				
Single channel	S-band X-band		< 36 deg rms < 132 deg rms	
Differential	S - 3/11 X S-band RCP-LCP X-band RCP-LCP		< 10 deg rms < 1 < 1	

Table 2. Implementation schedule

DSS	Project	Operational date S - and X-bands					
		BW - RCP channel			BW - LCP channel		
		Medium	Narrow	Very narrow	Medium	Narrow	Very narrow
63	Voyager 1 Jupiter encounter		10-78				
	Pioneer 2 Solar conjunction			7-79			
	Voyager 1 Saturn encounter	6-80			6-80		
	Galileo Faraday rotation exp					7-83	7-83
14	Voyager 2 Jupiter encounter		5-79				
	Pioneer 2 Solar conjunction			7-79			
	Galileo Faraday rotation exp	7-83			7-83	7-83	7-83
43	Voyager 2 Saturn encounter		5-81	5-81			
	Galileo Faraday rotation exp	7-83			7-83	7-83	7-83

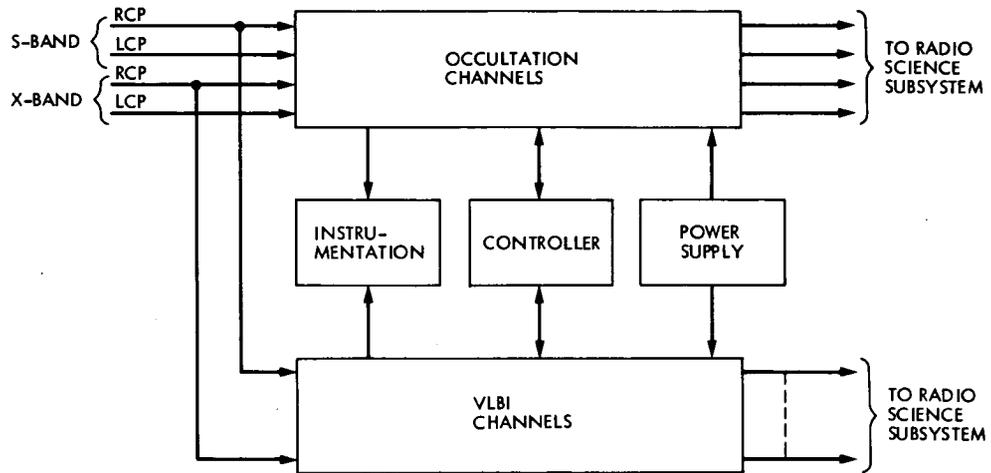


Fig. 1. Multi-Mission Receiver

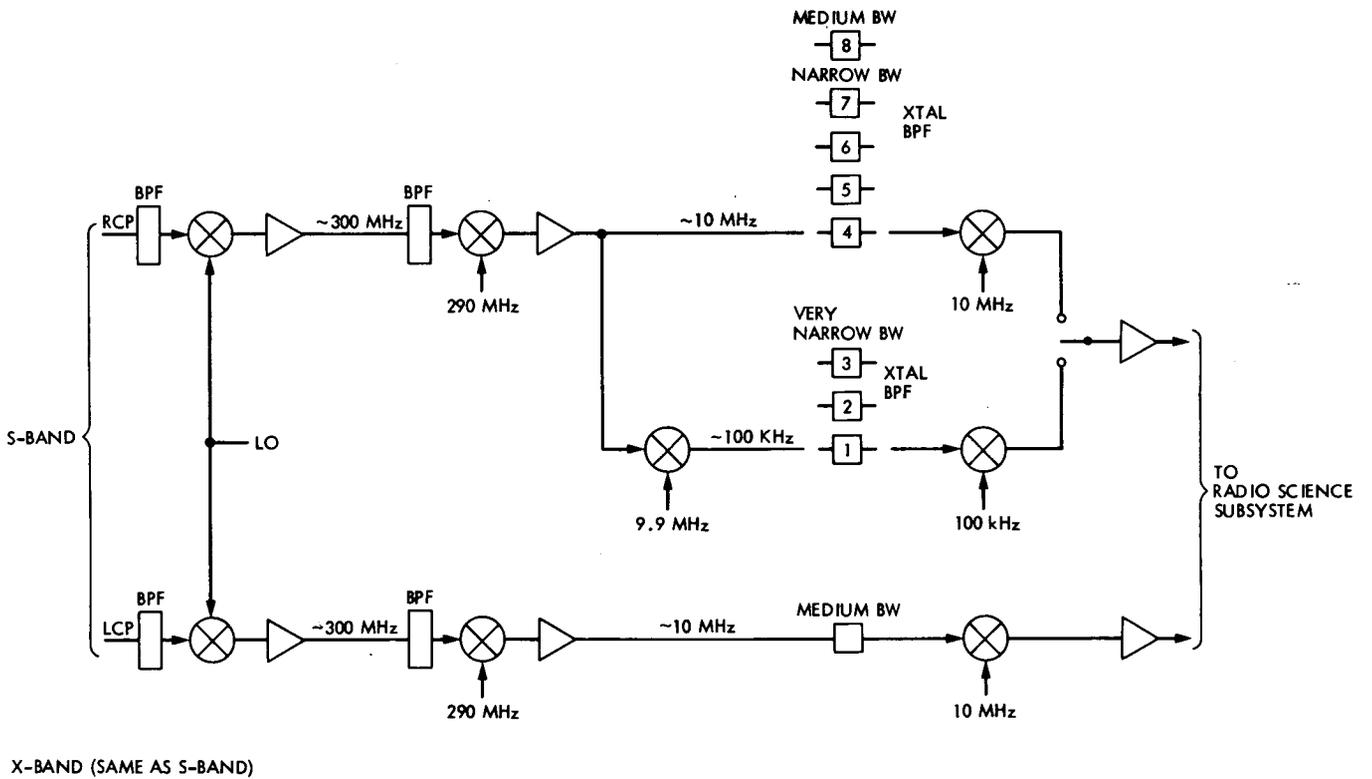
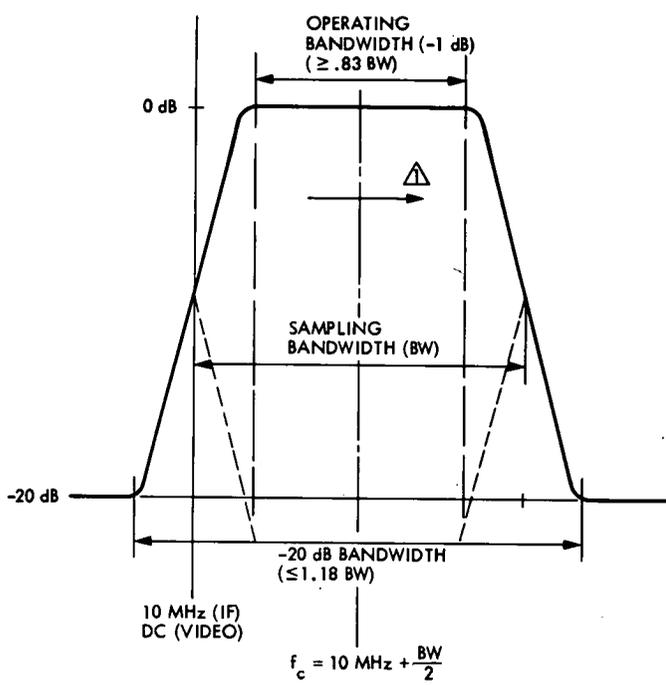
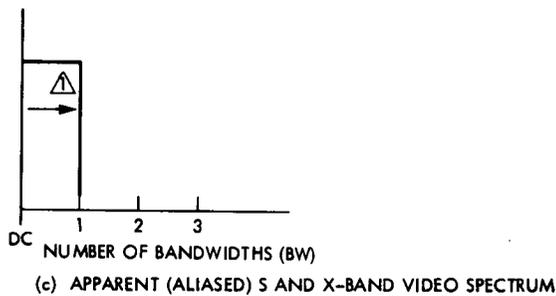
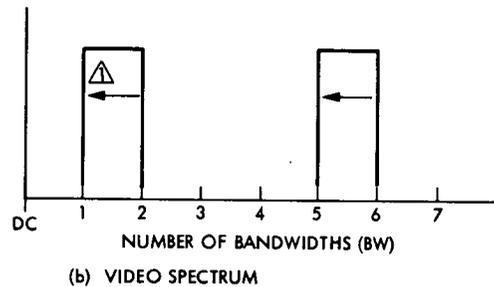
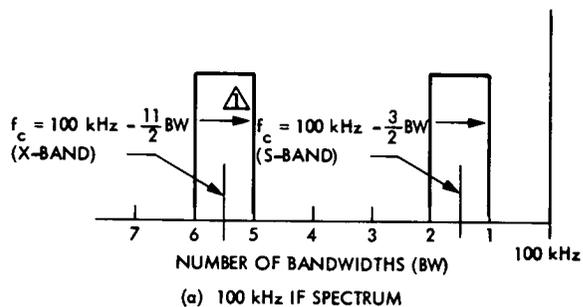


Fig. 2. Occultation channels



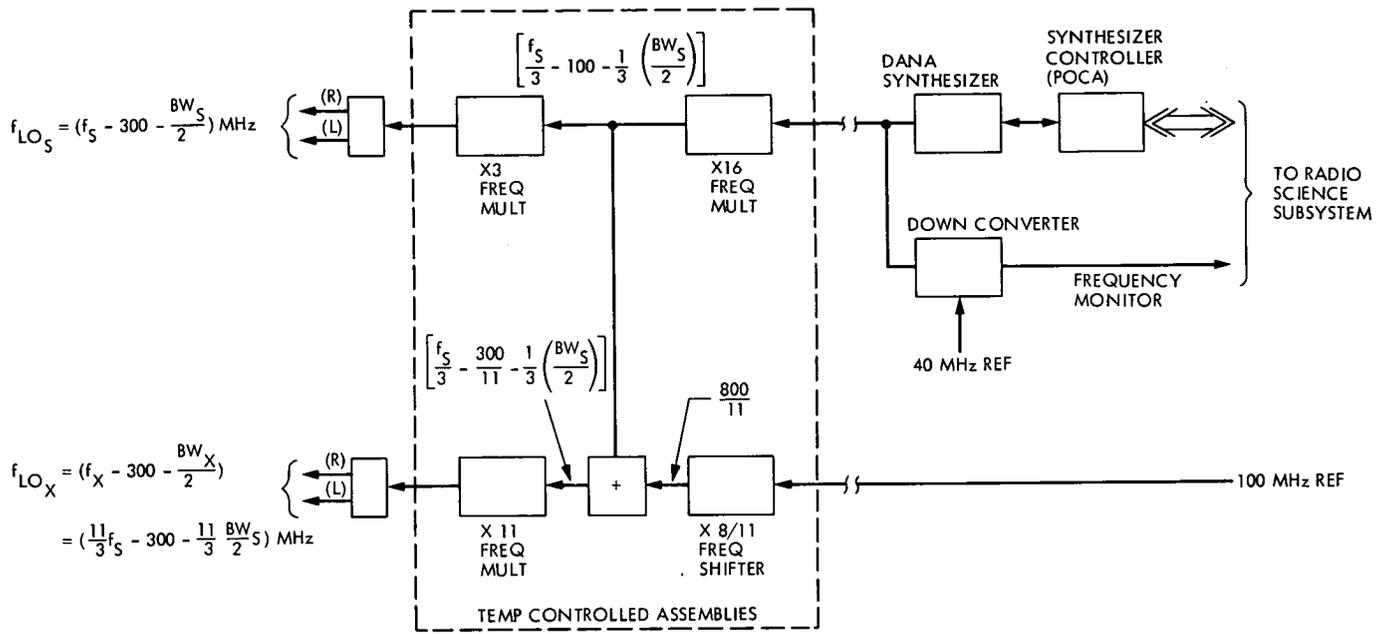
△ ARROW INDICATES DIRECTION OF OUTPUT FREQUENCY CHANGE WITH INCREASING INPUT FREQUENCY

Fig. 3. Medium and narrow band filter responses



△ ARROW INDICATES DIRECTION OF OUTPUT FREQUENCY CHANGE WITH INCREASING INPUT FREQUENCY

Fig. 4. Very narrow band filter channel spectrum: (a) 100-kHz IF spectrum, (b) video spectrum, (c) apparent (aliased) S- and X-band video spectrums



BW_S IF FILTER BANDWIDTH (S-BAND)
 BW_X IF FILTER BANDWIDTH (X-BAND)

Fig. 5. Occultation channels – first local oscillator

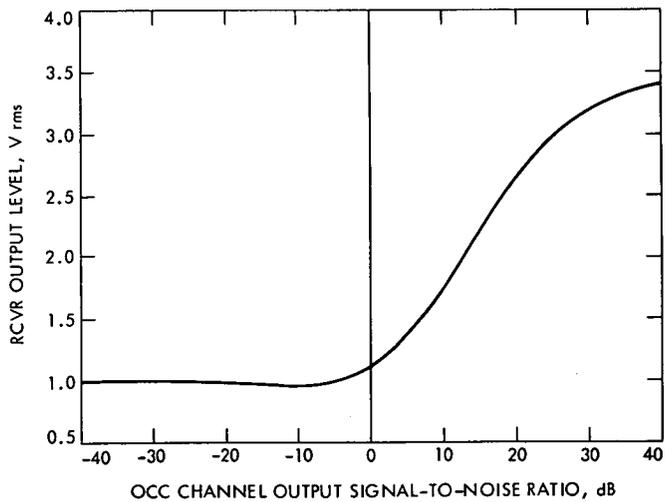


Fig. 6. OCC channel operating level

Preliminary Report on DSN System Performance Under Local Weather Effects

J. M. Urech
Deep Space Station 62

Local effects of precipitation are studied, and a simplified working model is developed. Experimental results obtained by simulation are in good agreement with the model, showing that it could be an important contribution to system degradation. If this is confirmed, some suggestions for improvement are presented. Nevertheless, definite results with actual rain are not yet available.

I. Introduction

The effects of the weather on X-band telecommunications system performance have generally been of major concern. At stations having only S-band capability, however, subject effects have proven to be very minor for normal tracking. Nevertheless, during DSS 62 continuous support of the planetary radio astronomy program (Jupiter Patrol), the weather problem became a major concern as the minor SNT fluctuations presented catastrophic results when measuring Jupiter temperature of approximately 0.5 K. In some cases the observations had to be cancelled because of rain.

As a result of this problem, it was decided to consult the DSN/Flight Project Interface Design Handbook (810-5) and Ref. 2. In both cases, statistical and probabilistic data are presented for the effects (mainly at X-band) of nonprecipitating clouds. Although this information is especially important for long-term planning, the interest of the station was mainly on the total effects, including precipitation. Although the probability of precipitation is small, some radio astronomy passes

at S-band have undoubtedly been spoiled by rain and critical X-band reception has been badly hampered.

In a further search for weather performance data, other available documents were consulted (Refs. 2-4). However, the information obtained presented a set of curves giving the path losses (dB/km) due to rain for different frequencies and rainfall rates. By using these path losses (which do not include direct antenna effects), the equivalent system noise temperature (SNT) increase may be estimated for certain rain assumptions. The estimates for equivalent SNT increases are normally lower than the values encountered in actual rainfalls.

The extra SNT during rain (not totally explained by cloud attenuation and path losses) as well as the fact that zenith SNT is exceptionally high once rain has subsided has led to the belief that there exists another important contribution related directly to rain falling on the antenna (Mylar and/or reflector). As this potential effect was not contemplated in the consulted documents (Refs. 2-4), a simple experiment was performed to clarify our assumptions.

With the DSS 62 antenna at zenith, a plastic tray was set in front of the horn in order to measure the S-band SNT as a function of the thickness of the various water layers. The initial result of the experiment was absolutely surprising – a system noise temperature of 180 K was achieved with only about 0.5 mm of water. Because of this unexpected result, the entire system was completely checked out and the experiment repeated. However, the second attempt yielded the same results and produced alternating maximum and minimum SNT values for increasing water thickness. The results were not firmly accepted and doubts existed as to whether they were correct and the validity of their interpretation. Nevertheless, it was encouraging to think that the station was on the right track toward the understanding of the precipitation problem.

II. Simplified Working Model

As a first step to clarify the high SNT contribution, studies were made on the properties of water as a thermal radiation absorber and emitter. This, however, did not help as a water layer would only contribute less than 10 K through simple absorption and emission. A clear explanation would probably require an expert in electromagnetic theory, microwaves and radiation theory. In the absence of the latter, numerous documents were consulted on the above fields in order to shed some light on the problem (Refs. 5-7). Although the effect is still not fully understood, it is believed that our simplified working model may help explain the increase in SNT.

A. Effects of Water on Mylar Covering the Horn

Water, as an imperfect dielectric will partially attenuate the electromagnetic radiation by absorption and consequently (Kirchoff law of radiation) partially emit thermal radiation. The most important fact in our case, however, is that the refraction index is fairly high for S- and X-band, thereby acting as a good reflector. If the horn is considered together with the waveguides, diplexer or polarizer (S- or X-band), etc., all the energy received within the design bandwidth will be absorbed with almost no reflection. Reciprocally, all internally generated energy will be radiated out of the horn with almost no back reflection. This may be equivalent to considering the horn as a special black body cavity radiator emitting thermal radiation at ambient temperature. Transmitted to the outside, the radiation is normally unnoticeable. However, if a water film or other reflective dielectric is present in front of the horn, part of the energy will be reflected back, some absorbed by the film, and the remainder will be transmitted outwards. If r is the fraction of incident power reflected, α is the fraction of incident power absorbed, and t is the fraction of incident power transmitted, then $r + \alpha + t = 1$.

However, the thermal radiation power e emitted by the film (if in local thermodynamic equilibrium) is, according to the Kirchoff law of radiation, identical to the absorbed radiation α . Therefore, if the horn and dielectric film are at ambient temperature T_0 , the power (noise temperature) detected by the receiver for this effect will be:

$$T = T_0 (r + e) = T_0 (r + \alpha) = T_0 (1 - t)$$

which is the same as the well-known formula generally used to determine the noise temperature contribution of any attenuation or mismatch loss.

Obviously any signal received or transmitted will be attenuated by $10 \log t$ (dB's). The power reflection and transmission coefficients for a dielectric film in air are (Ref. 8)

$$r = \frac{\rho^2 [(1 - A^2)^2 + 4A^2 \sin^2 \Phi]}{(1 - A^2 \rho^2)^2 + 4A^2 \rho^2 \sin^2 (\Phi + \chi)}$$

$$t = \frac{A^2 [(1 - \rho^2)^2 + 4\rho^2 \sin^2 \chi]}{(1 - A^2 \rho^2)^2 + 4A^2 \rho^2 \sin^2 (\Phi + \chi)}$$

where

$A = \exp(-2\pi d n/\lambda_0 K)$ amplitude transmission coefficient for single film transversing

$\phi = 2\pi d n/\lambda_0$ phase shift for single film transversing

$\rho = n - 1/n + 1$ amplitude reflection coefficient for a single reflection

$\chi = \tan^{-1}(2nK/n^2 - 1)$ phase shift of single reflection

d = film thickness

n = refraction index

λ_0 = wavelength in air

$K = tg\delta/2$

$tg\delta = \sigma/\omega\epsilon$ loss tangent or dissipation factor

σ = dielectric conductivity

ϵ = dielectric constant

The above expressions have been evaluated for the following cases of general interest:

(Fig. 1) Temperature and attenuation of a water film in S-band

(Fig. 2) Temperature and attenuation of a water film in X-band

(Fig. 3) Temperature and attenuation of snow and ice in S-band

(Fig. 4) Temperature and attenuation of snow and ice in X-band

These results should be considered indicative rather than totally accurate, because the values for the dielectric constant and dissipation factor at S- and X-band have been interpolated from the data presented for other frequencies in Table 4-19 of Ref. 2 (as the only source presently available). Nevertheless, it is considered that this questionable accuracy would not invalidate the following important conclusions:

- (1) A very thin water film on the Mylar will contribute to very high system noise temperatures, especially at X-band. This is mainly due to the high refraction index of water for these frequencies, thereby making it a good reflector, reducing also the effective wavelength ($\lambda = \lambda_0/n$).
- (2) Figures 1 and 2 are evaluated up to very thick water films, not that they can be formed by rain, but in order to permit validation of the model by feasible experiments. In both cases, the puzzling result initially obtained of alternating maximum and minimum SNT is confirmed. This occurs at multiples of $\lambda/4$ (wavelength in water) when the internal reflections of the film are in or out of phase with respect to the main reflection at the interface.
- (3) Figures 3 and 4 clearly show that SNT contributions for freshly fallen snow, hard packed snow, and ice are in any case much smaller than for liquid water. Subject contributions are a consequence of the lower refraction indexes: 1.1, 1.22 and 1.78, respectively, as compared to 8.83 and 6.5 for liquid water at S- and X-band.

B. Effects of Water on Antenna Reflector Surface

This could also be considered another potential contribution to the general degradation during rain. However, it is difficult to confirm experimentally. Therefore, if the model used in the previous case is deemed valid, the reflection losses may be estimated using a similar expression.

The power reflection coefficient for a dielectric film covering a perfect reflector is (Ref. 8), using the same notation as before,

$$r' = \frac{(\rho + A^2)^2 - 4A^2 \sin^2(\phi - \chi/2)}{(1 + \rho A^2)^2 - 4\rho A^2 \sin^2(\phi + \chi/2)}$$

Based on a temperature contribution of $T = T_0 (1 - r')$, the following values were estimated:

Liquid water		
Film thickness, mm	S-band temp, K	X-band temp, K
0.25	2.6	0.81
0.50	2.7	7.4
0.75	4.3	33.0
Snow		
10	0.005	0.42
20	0.11	0.58

As can be seen, this potential contribution will be negligible when compared with similar films on the Mylar, as has been previously shown.

It could be argued that the above models would not be applicable to real cases as a water film may not be formed. This is partially true because the water by surface tension may form dispersed drops instead of a film. Nevertheless, if the horn illumination is supposed to be uniform, and the temperature contribution model practically linear for small thickness, the effect of the dispersed drops could be estimated assuming a film thickness equivalent to the total amount of water on the surface.

III. Empirical Results

Once the simplified working model was developed, a test, similar to the one mentioned in the introduction, was carefully prepared for S-band at DSS 62. A flat plastic tray was fabricated having a slightly larger size than the horn window. With the antenna at zenith, the tray was perfectly leveled in front of the horn and barely touching the Mylar. Then each time a controlled amount of water was added, a measurement of the system noise temperature was taken with the noise adding radiometer. For the first two points obtained (equivalent to film thickness of 0.5 and 1 mm), the film was not actually formed, but the drops were purposely distributed more or less uniformly.

All SNT data points have been directly plotted in Fig. 1, and the solid line curve is the theoretical water film contribution (not the total SNT including waveguide losses, receiver, etc.). The results, in general, are in good agreement with the simplified model, thereby confirming two aspects: First, that the model is valid for the study of these types of effects; second and more important, that the water film on the Mylar may have a catastrophic effect on telecommunications performance.

A similar test setup was prepared for X-band at DSS 63. In this case, due to the geometry of antenna and X-band cone, the horn window could not be set horizontally and therefore not parallel to the levelled tray. This has some implication on the angle of reflection which may slightly jeopardize the test results. However, the SNT measurements, which are plotted in Fig. 2, are also in fairly good agreement with the model. As should be expected and can be seen from Figs. 1 and 2, the water temperature contribution of very thin films is much higher for X-band: 95 K vs 25 K for an equivalent film of one-tenth millimeter.

No similar tests have been conducted to check Figs. 3 and 4 (snow and ice) and therefore they should be taken only as indicative.

IV. Suggestions for Degradation Reduction

The possibility of an equivalent water film forming on the Mylar surface during rain will be directly related to two opposing factors:

- (1) The amount of water impinging on the surface which is a function of the rainfall rate direction and the antenna position.
- (2) The water drainage on the same surface which is also a function of the type of surface and its affinity to water, wind and rain intensity and direction, and antenna position.

Although it is very difficult to quantify all these factors, one technique which will enhance the situation is to improve the water drainage as much as possible. In this regard, many different ideas may be proposed. Some of our suggestions are presented below for further study, testing or development, based mainly on the X-band feed as the case of major concern.

- (1) The capability of Mylar to retain water drops seems to be low but there may exist another material with better performance, or a water repellent coating.

- (2) The ring connecting the Mylar to the X-band horn is quite thick. This could retain some water on the internal edge which may not be negligible when compared to the window size.
- (3) A special type of plastic shield wiper having low refraction index and dissipation factor may be designed and tested.
- (4) A vibrating window (acoustically or ultrasonically activated) might be developed.
- (5) A centrifugal double window similar to the ones used in boats might be used. This method, if feasible, would be absolutely efficient in draining water (see sketch A of Fig. 5).
- (6) A well-designed system of nozzles (jets) blowing compressed air at high-speed over the Mylar window might be employed.

As this system is relatively simple, it has been tentatively tested in a rudimentary form. A standard nozzle attached to a small plastic fixture (see sketch B of Fig. 5) was temporarily mounted near the X-band horn (outside its field of view) and connected by a hose to an air-compressor with a pressure of 90 psi. With a "rain emulator" (a plastic watering-can) a more or less uniform shower was generated over the Mylar, obtaining a system noise temperature of about 180 K. Then with the nozzle blowing the SNT came down to about 100 K. The test was repeated with a lighter shower, changing the SNT from 85 K to 55 K when blowing.

V. Conclusions

From the theoretical and experimental results of this study, the logical conclusion would be to believe that the local rain effect may be of prime importance. However, to reach a firm conclusion, further tests have to be performed during actual rain in order to measure the relative contributions of the path and local effects and determine if they are separable. For a simple qualitative test, the rudimentary system suggested in IV-6 above has been temporarily left installed, but no test data is yet available.

References

1. Greenhall, C. A., "Examination of the DSN X-Band Weather Specifications," in *The Deep Space Network Progress Report 42-45*, pp. 197-208, Jet Propulsion Laboratory, Pasadena, Calif., July 15, 1978.
2. *Reference Data for Radio Engineers*, 5th edition, ITT Handbook.
3. Fink, D. G., *Electronic Engineers' Handbook*, McGraw Hill, New York.
4. Marton, L., and Meeks, M. L., *Methods of Experimental Physics: Volume 12, Astrophysics, Part B: Radio Telescopes*.
5. Ramo, S., Whinnery, J. R., *Fields and Waves in Modern Radio*, 2nd edition, Wiley, 1953.
6. Eisberg, R., and Resnick, R., *Quantum Physics*, Chapter 1, "Thermal Radiation."
7. Menzel, D. H., selected papers on the transfer of radiation.
8. Blevis, B. C., "Losses Due to Rain on Radomes," *IEEE Trans. Ant. Prop.*, pp. 175-176, Jan. 1965.

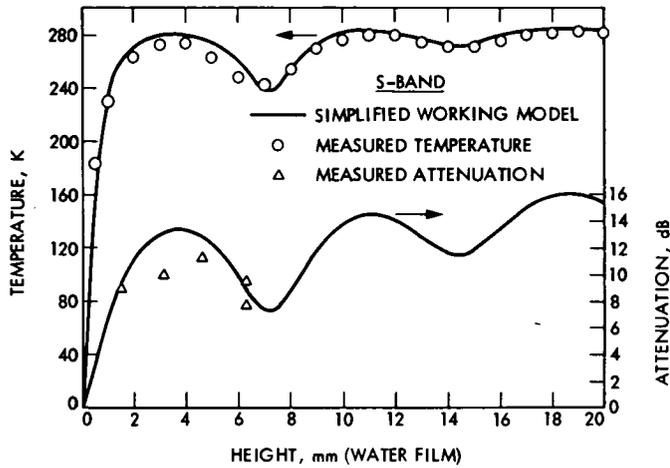


Fig. 1. Temperature and attenuation of a water film in S-band

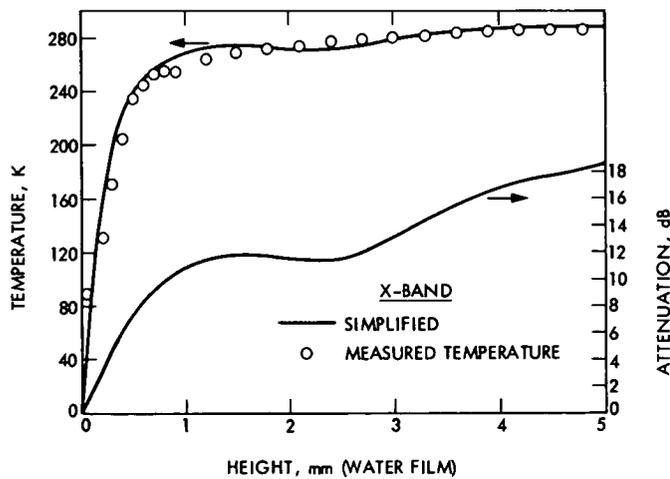


Fig. 2. Temperature and attenuation of a water film in X-band

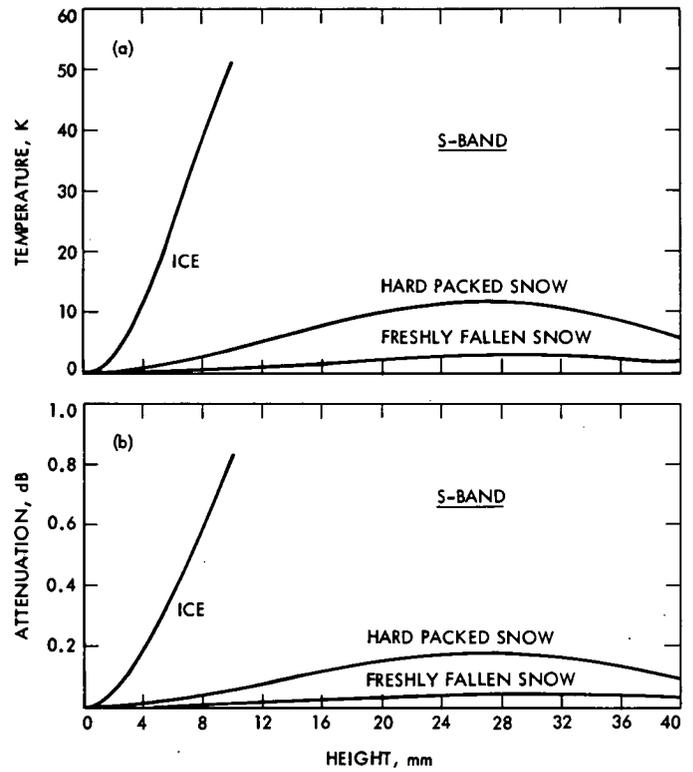


Fig. 3. Temperature and attenuation of snow and ice in S-band

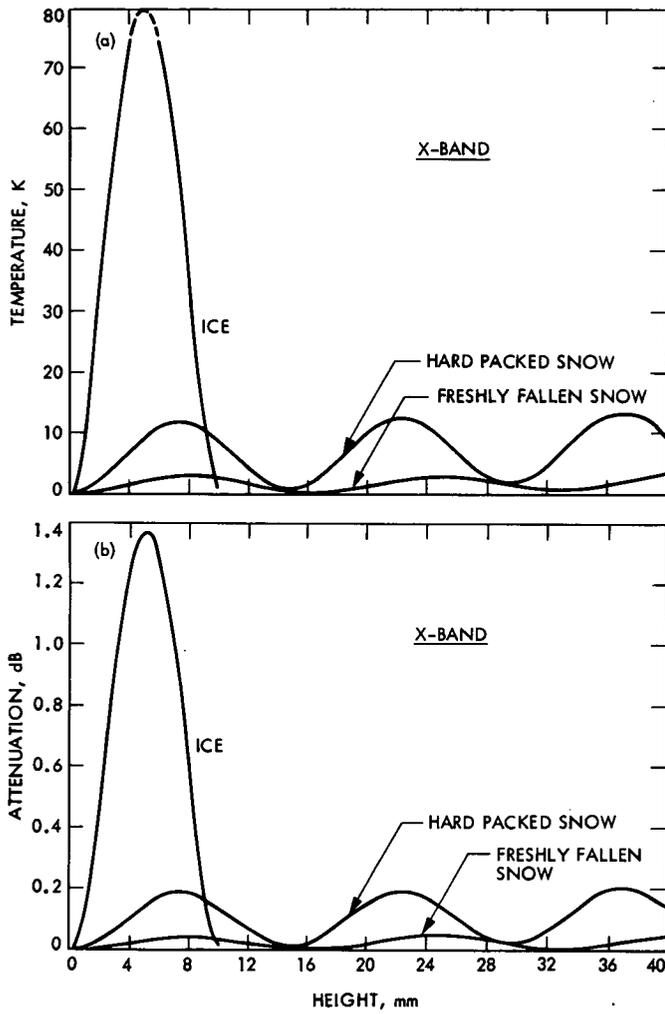


Fig. 4. Temperature and attenuation of snow and ice in X-band

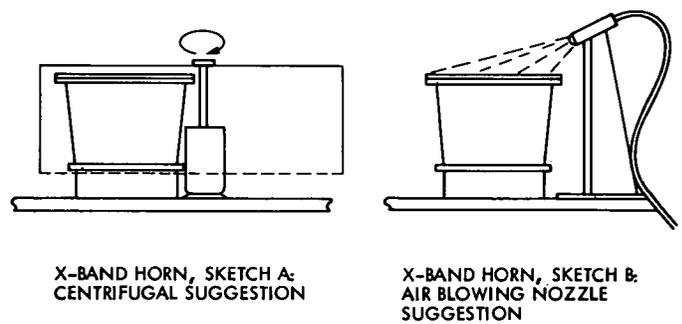


Fig. 5. Proposed test configuration

Comparison of Two Total Energy Systems for a Diesel Power Generation Plant

V. W. Chai

DSN Engineering Section

This article compares the capabilities and limitations, as well as the associated costs for two total energy systems for a diesel power generation plant. Both systems utilize waste heat from engine cooling water and waste heat from exhaust gases. Pressurized water heat recovery system is simple in nature, requires no engine modifications, but operates at lower temperature ranges. On the other hand, a two-phase ebullient system operates the engine at constant temperature, provides higher temperature water or steam to the load, but is more expensive.

I. Introduction

Total energy systems have been widely studied as a more effective means of using the depletable energy sources. Waste heat from the diesel engine represents a potential source of energy that can be further utilized for heating and/or cooling purposes or, coupled with a Rankine bottoming cycle, to produce extra power. Diesel engines for power generation have an efficiency between 30 to 35 percent, which means that 65 to 70 percent is wasted to the atmosphere.

First, about 40 percent of the total waste heat is exchanged during the combustion process in the engine, which tends to increase the cylinder wall temperature to such an extent that without cooling, the lubrication oil can evaporate resulting in engine damage. Water is usually employed to cool the engine. The temperature of the cooling water leaving the engines range from 87.7°C (190°F) to 104.4°C (220°F). However, to avoid excessive thermal stresses in the engine, it is advisable to have a temperature difference of 5.55°C (10°F) to 8.33°C (15°F) between the inlet and the outlet water temperatures. The hot cooling water from the engine is then fed into an air-cooled or a water-cooled radiator

before recirculating it back to the engine. Instead of rejecting the heat to the atmosphere, as much as 90 to 100 percent can be recovered for heating or cooling purposes. Second, the other 60 percent of the waste heat is in the form of sensible heat in the exhaust gases at the end of the expansion stroke. The temperature of the exhaust gases varies from engine to engine, ranging between 371.1°C (700°F) and 698.9°C (1200°F). The sensible heat in the exhaust gases can be extracted by means of a stack-type heat exchanger. To avoid water condensate in the exhaust ducts, the temperature of the exhaust gases leaving the stack cannot be lower than 176.7° (350°F), which means that only 60 to 70 percent of the waste heat from exhaust gases is recoverable.

DSN overseas stations in Spain and Australia operate their diesel engines continuously around the clock to provide power for the stations' needs. Therefore, it is felt necessary to make use of the waste heat from the engines to conserve the valuable and depletable diesel fuel. DSS 62 is currently using the diesel waste heat for heating and cooling. DSS 61/63 is in the process of installing a waste heat recovery system for space heating, and an effort is underway to further use the diesel waste heat to provide cooling.

II. Relationship of Engine Load and the Available Waste Heat

To utilize the waste heat from diesel engines for heating and/or cooling, it is important to investigate the engine operating loads and the energy required for heating and/or cooling, since they determine the sufficiency or deficiency of the engine waste heat. By employing the waste heat for heating and cooling, the engine will be operating at a reduced load level since it no longer provides the electricity needed for the present heating and cooling devices such as electric boilers and vapor compression refrigeration units. In other words, the waste heat available will be less. On the other hand, the coefficient of performance (COP) of the absorption chiller and the efficiency of the heat exchangers should be taken into consideration to determine the heat needed to meet the heating and/or cooling load. For instance, a diesel engine operating at 500 kW_e is providing heating and cooling loads of 150 kW_t and 100 kW_t, respectively. Utilizing the waste heat for heating and cooling, the engine load should be reduced by 200 kW_e assuming an efficiency of 0.9 for the electric boiler and the COP of 3.0 for the vapor compression chiller. Further, taking the COP of the absorption chiller at 0.6 and the heat exchanger efficiency at 0.9, the heat needed will be enhanced to approximately 350 kW_t, and obviously, the waste heat from jacket water alone for an engine operating at 300 kW_e load is not sufficient and the waste heat from exhaust gases will have to be used to compensate for the heat needed. Figure 1 shows the relationship between the engine load and the sufficiency of jacket waste heat for a one-day period. From hour 12 to hour 22, the waste heat from jacket water alone can no longer provide enough heat for both heating and cooling; the heat available from exhaust gases or back-up systems should be further utilized to make up for the deficiency.

In the event that the waste heat from exhaust gases is used to supplement the balance of heat required, several heat recovery systems can be used and their capabilities and limitations, as well as their associated costs, should be carefully investigated.

III. Heat Recovery Systems

Basically, there are two types of heat recovery systems that utilize both the engine jacket water and the exhaust gases: 1) a pressurized water system, and 2) an ebullient or 2-phase steam/water system. The pressurized hot-water system operates the engines at a temperature range of 82.2°C (180°F) to 92.9°C (210°F). One-third of the fuel energy input to the engine will dissipate into the jacket water, which would cause a temperature difference between the water coming into the engine and the water leaving the engines of

not more than 8.33°C (15°F). To achieve this, a temperature-controlled water flow rate should be maintained. The water leaving the engine is then fed to a gas-heat recovery unit to extract the waste heat from the exhaust gases. The gas-heat recovery units are typical gas-to-water heat exchangers with exhaust gases running in the hot side, and the engine jacket water in the cold side. Reference 1 lists the common temperatures of exhaust gases for various models of caterpillar engines. Since the temperature of the exhaust gases leaving the heat recovery units should not be lower than 176.7°C (350°F), the amount of recoverable waste heat can easily be estimated. Moreover, the temperature of the water leaving the heat recovery units also can be determined. The temperature increase for the water across the heat recovery units will generally not exceed 5.55°C (10°F). As a result, a maximum water temperature of approximately 100°C (212°F) can be achieved only by using this scheme. For engines operating at temperatures close to 100°C (212°F), pressurized water systems should be employed throughout the cooling water loop to avoid flashing in the engines. It is also advisable to use in the external loop a heat exchanger to deliver the heat to the load to form a closed loop for the engine cooling water. A booster pump will also be installed since the existing engine pump will not be enough to circulate the water. At times when cooling is not needed, the heat can be removed using the existing radiators or converted to chilled water and stored in a chilled water storage tank. The pressurized hot-water system does not require any modifications to the existing engines. However, the maximum obtainable hot-water temperature (100°C) is lower than the absorption chiller manufacturer's specification (116°C), which means an oversized chiller would be used at a lower capacity rather than the nominal size, even though it operates with the same coefficient of performance (COP). Table 1 lists the absorption chillers with the nominal sizes at 115.6°C (240°F) and the degraded capacities at 90.5°C (195°F). Figure 2 depicts the schematic flow diagram for the pressurized hot-water system.

The 2-phase steam/water ebullient system removes the heat from the engines by heat of vaporization. For an ebullient system, the engines are operating at a constant high temperature ranging between 112.8°C (235°F) to 121.1°C (250°F). The cooling water is fed to the engines at the operating temperature. As the water passes through the cylinders, the heat rejected will vaporize a portion of the water into wet steam. The wet steam is fed to heat recovery units which serve both as mufflers and as steam separators. Using the heat from the exhaust gases will further convert the wet steam to dry saturated or superheated steam, which feeds the absorption chiller. Theoretically, the cooling water into the engine should have the same temperature as the water leaving the engine, but a 1.1°C (2°F) to 1.6°C (3°F) increase in temperature is normally experienced. Because of the smaller temperature

gradient across the engine, the engine is subject to smaller thermal stresses, which give the engines a longer life. Further, the cooling water circulation is by natural convection and a pump is not necessary. The water-steam flow rate is about 1/70 of that of the pressurized system due to the large latent heat of water. However, this system requires a separate cooling loop for oil coolers, which uses water at 29.4°C (85°F). The cooling water loop should maintain a pressure of 69 kPa (10 psi) to 103.5 kPa (15 psi) to avoid flashing inside the engines; the temperature of the steam collected will be in the neighborhood of 115.6°C (240°F) to 121.1°C (250°F), which satisfies the chiller manufacturer's specification. For engines running at such a high temperature, the seals, plumbing system, and even the cylinder head will have to be replaced for high-temperature applications. Thus, it is appropriate to convert the engines to the ebullient cooling system during a major overhaul period. When excess steam is not needed, the steam will be condensed in a bypass condensing unit and returned to the heat recovery units. Figure 3 shows the schematic diagram of a high-temperature ebullient system. Water treatment for the engine cooling water is necessary for an ebullient system, and it is critical to engine operation. It is desirable to have chemical testing of the water every other day to keep the water at a certain pH level.

IV. Cost Comparison

A pressurized hot-water system does not require any engine modification. However, the system needs a heat exchanger, a booster pump, and an oversized absorption chiller. On the other hand, an ebullient system requires no heat exchanger, works with nominal size absorption chiller, but requires engine modification, an excess steam condensing unit, a holding tank for the condensate, and a more expensive heat recovery unit. It is necessary to have a thorough breakdown of the costs for each system to select the most cost-effective system.

To demonstrate the cost comparison of the two systems, assume a diesel power generation plant that operates four diesel engines and utilizes the waste heat from the engines to feed an absorbing chiller of 700-kWt capacity.

Table 2 lists the cost for each component for both systems. The installation cost and the maintenance cost for the two systems are assumed the same. The main differences in cost between the two systems fall into three categories: 1) engine modification, 2) heat recovery units, and 3) the absorption chillers. For ebullient systems, engines have to be modified for high-temperature application. The cost of the modification is taken as \$7,000 per engine including the replacement of the cylinder head, seals and necessary changes in the plumbing system. The pressurized hot-water system uses a gas-water heat exchanger for the heat recovery unit, while the ebullient system employs a more expensive heat recovery unit acting as both a heat exchanger and a steam separator. Further, the pressurized hot-water system requires an oversized chiller to provide the same amount of cooling as the ebullient system. The material cost for hot-water system and the ebullient system are approximately \$103K and \$150K, respectively.

V. Conclusion and Remarks

A pressurized hot-water system and the high-temperature ebullient system are the two commonly used heat recovery systems associated with diesel total energy systems. Both systems recover waste heat from the engine jacket water as well as waste heat from the exhaust gases. The pressurized hot water system is simple, requires no engine modification, but needs an oversized absorption chiller. On the contrary, high-temperature ebullient systems require engine modifications, more expensive heat recovery units and a water treatment system. The installation and maintenance costs for both systems are very much the same, but the material cost for the ebullient systems are 1.5 times higher than that for the pressurized hot-water system. Furthermore, for future increase in station cooling consumption, the pressurized hot-water system has the capability of providing the additional cooling with a supplementary source such as solar energy, to enhance the hot water temperature to, in turn, increase the capacity of the absorption chiller, independent of the engine waste heat.

Reference

1. *Air Conditioning Handbook*. Caterpillar Tractor Company, Peoria, Illinois, 1970.

Table 1. Degradation of chiller capacities and corresponding costs

Nominal size at 115.6°C (240°F), kWt	Operation capacity at 90.5°C (195°F), kWt	Estimated cost, \$K
2180	879	70.0
1976	791	64.5
1835	738	61.0
1723	703	59.0
1396	562	48.5
1097	440	42.0
819	334	36.0

Table 2. Cost comparison for pressurized hot-water system and high-temperature ebullient system

Component	Pressurized hot-water system, \$K	High-temperature ebullient system, \$K
Engine modification	—	@ 7 × 4 = 28
Heat recovery units and Insulation and controls	@10 × 4 = 40	@20 × 4 = 80
Heat exchanger or condensing unit	4	10
Absorption chiller and accessories	59	36
Total Cost	103	154

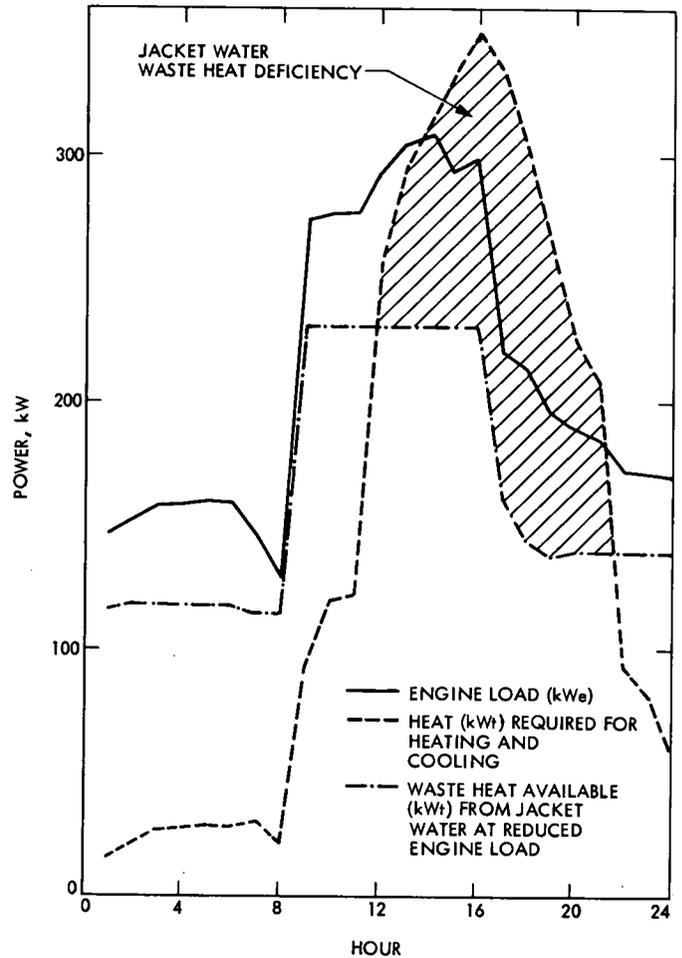


Fig. 1. Example of waste heat profile for one day

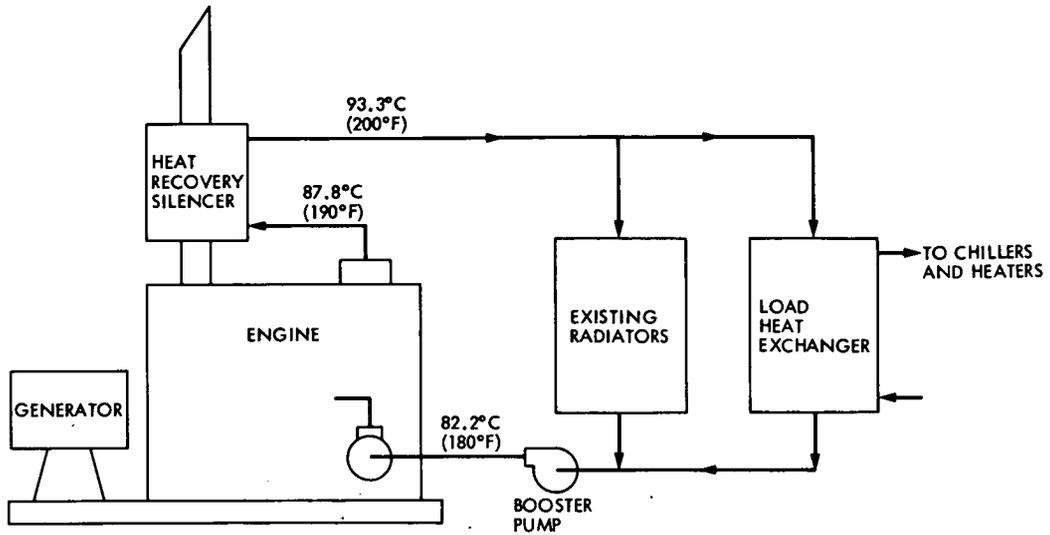


Fig. 2. Pressurized hot-water system

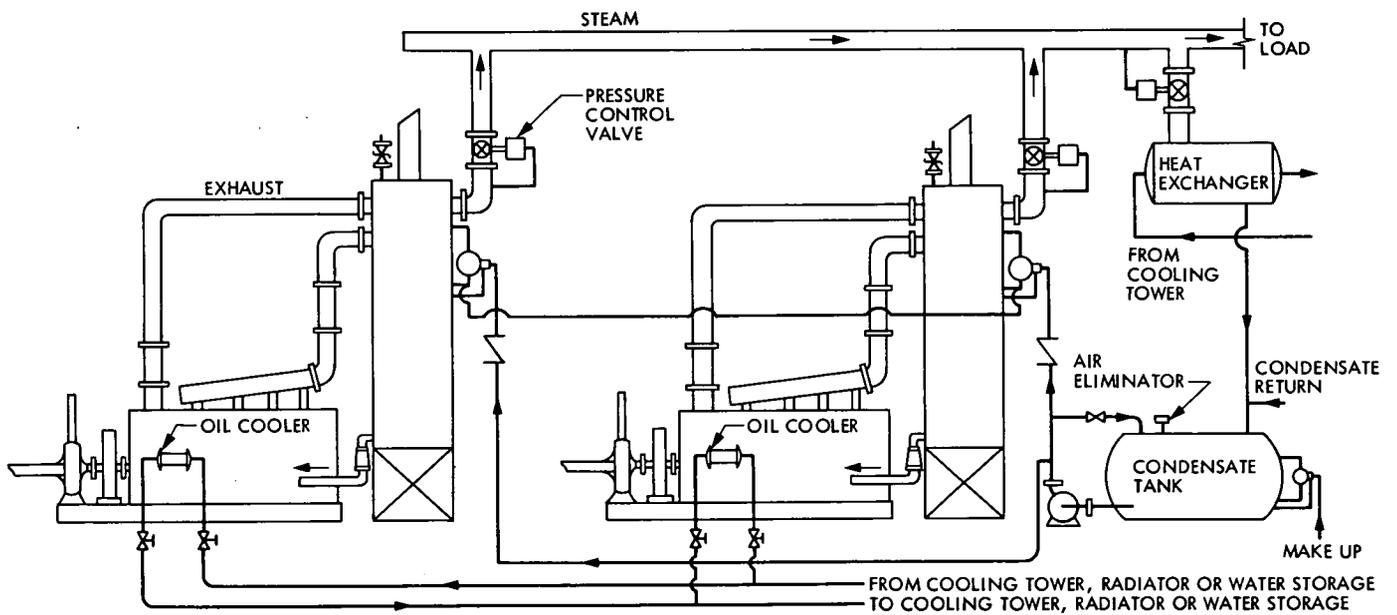


Fig. 3. High-temperature 2-phase water/steam ebullient system

Minimizing the RMS Surface Distortions from Gravity Loadings of the 34-m HA-DEC Antenna for Deep Space Missions

M. S. Katow
DSN Engineering Section

The computer analysis of the 34-m HA-DEC antenna by the IDEAS program provided the rms distortions of the surface panels support points for full gravity loadings in the three directions of the basic coordinate system of the computer model. The rms distortions for the gravity vector not in line with any of the three basic directions were solved and contour plotted starting from three surface panels setting declination angle. By inspections of the plots, it was concluded that the setting or rigging angle of -15 degrees declination minimized the rms distortions for sky coverage of plus or minus 22 declination angles to 10 degrees of ground mask.

I. Introduction

A HA-DEC antenna has an axis of rotation parallel to the Earth's axis of rotation combined with a second rotation axis that is normal to the first axis. Sky coverage is thus obtained by the despining action of the first (polar) axis with the elevation or declination angle pointing provided by the rotation about the second (declination) axis. The polar axis rotation is measured in hour angle (HA) and the declination axis rotation in plus or minus declination angle (DEC) from the equatorial plane of polar axis rotation.

At an Earth's location, the rotational motions of the antenna changes the direction of the gravity loading vector with respect to its structural symmetric plans and axis. This action results in variations of the reflective surface's distortion values, after best fit of a paraboloid.

II. Calculations

By assembling a full or complete structural computer model of the 34-m reflector structure, the design and analysis were done in the IDEAS program (Ref. 1). The analysis results (Table 1) were the rms (root-mean-squared) distortions of the surface panels' attach points with the applications of the full or 1.0-gravity loadings in the three directions of the cartesian coordinate system of the model. The cross-sectional areas of the reflector's truss members were changed in the IDEAS program through its iterative steps of designing cycles to minimize the rms distortions.

Using the method of analysis described in (Ref. 2), the rms distortion values of the 34-m HA-DEC antennas were plotted through the hour angle-declination angle ranges for several surface panels setting or rigging declination angle at zero hour

angle. Setting angles considered were at zenith look, at zero degrees declination, and at -15 degrees declination angles.

III. Results

The planetary missions normally confine the spacecraft to within plus or minus 22 degrees of declination angle values when the maximum gain capabilities of the ground antennas are required. By setting the surface panels at plus or minus 15 degrees declination angle, with the sign dependent on the hemisphere location, the rms distortion was minimized for full mission sky coverage above 10 degrees from ground mask.

This minimization is illustrated by Figs. 2, 3, and 4 showing the rms distortions for DSS 12 with the surface panels setting angles of zenith look, zero degrees declination angle, and -15 degrees declination angle. Figures 5 and 6 illustrate the rms distortions with the setting at -15 degrees declination for DSS 61 and +15 degrees declination for DSS 42.

Figure 1 describes the sign conventions used at the northern and southern hemisphere stations.

The latitude values are shown in Table 1 with the full or 1.0 gravity load values of the rms distortions for the three components loading directions as output by the IDEAS program.

IV. Summary

- (1) The rms distortions quoted are only for the surface panels' support points on the reflector structure.
- (2) For the selected observing ranges of plus or minus 22 degrees declination angles and 10 degrees elevation ground mask, the maximum rms distortion is 0.4 mm for the surface panels set at plus or minus 15 degrees declination, as required for the hemispheric requirement.

References

1. Levy, R., "Iterative Design of Antenna Structures," in *The Deep Space Network Progress Report*, Technical Report 32-1526, Vol. XII, pp. 100-111, December 15, 1972. Jet Propulsion Laboratory, Pasadena, California.
2. Katow, M. S. and Levy, R., "Computation of Gravity RMS for HA-DEC Antennas," in *The Deep Space Network Progress Report 42-27*, for March and April 1975, pp. 138-147. Jet Propulsion Laboratory, Pasadena, California.

Table 1. RMS distortions for full-gravity loads

Seq	DSS	Latitude, deg	RMS distortion for 1.0 gravity for component direction		
			X, mm	Y, mm	Z, mm
1	12 (Goldstone)	+35.20805	0.307	0.528	0.648
2	42 (Australia)	-35.21922	0.307	0.528	0.648
3	61 (Spain)	+40.23885	0.307	0.528	0.648

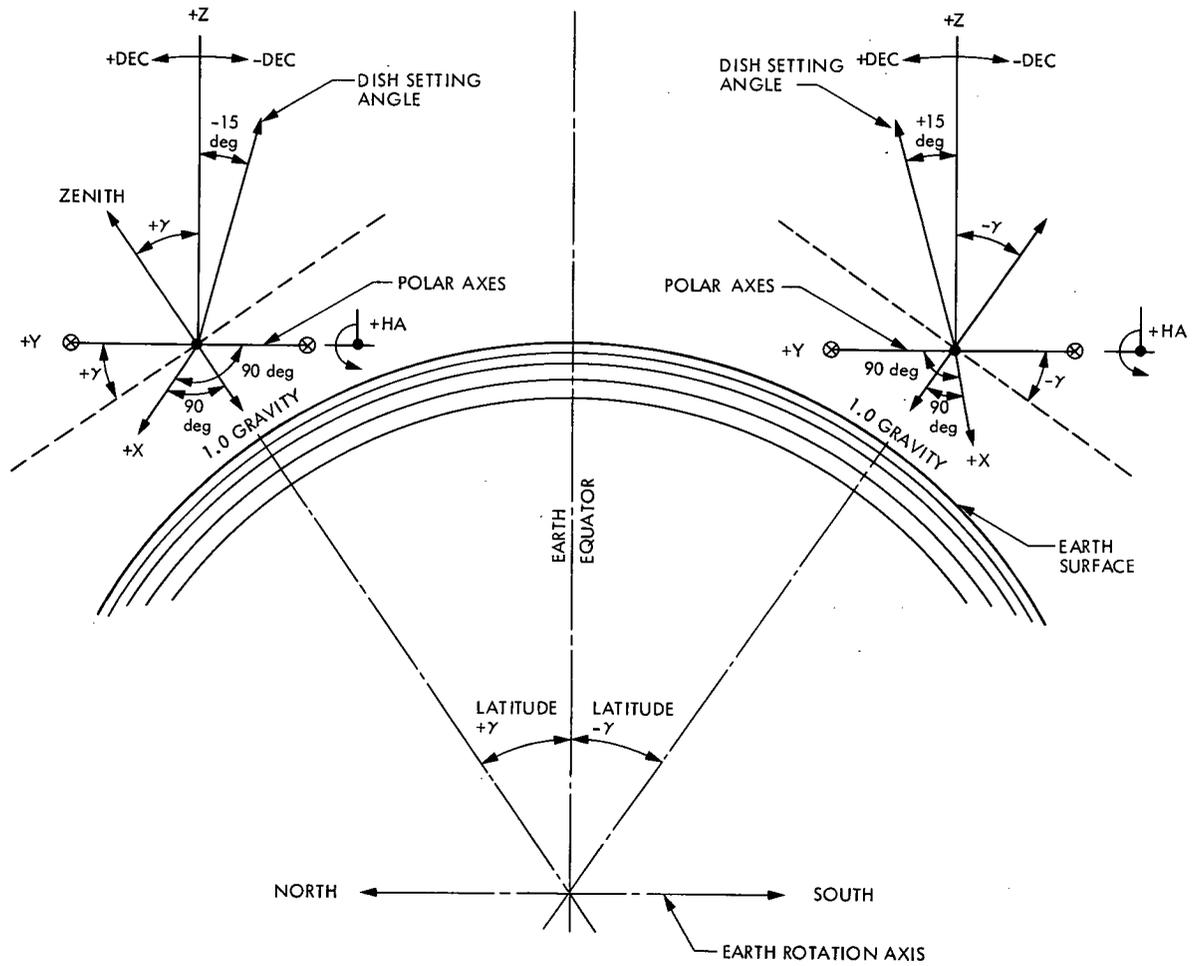


Fig. 1. Sign conventions – north and south hemisphere

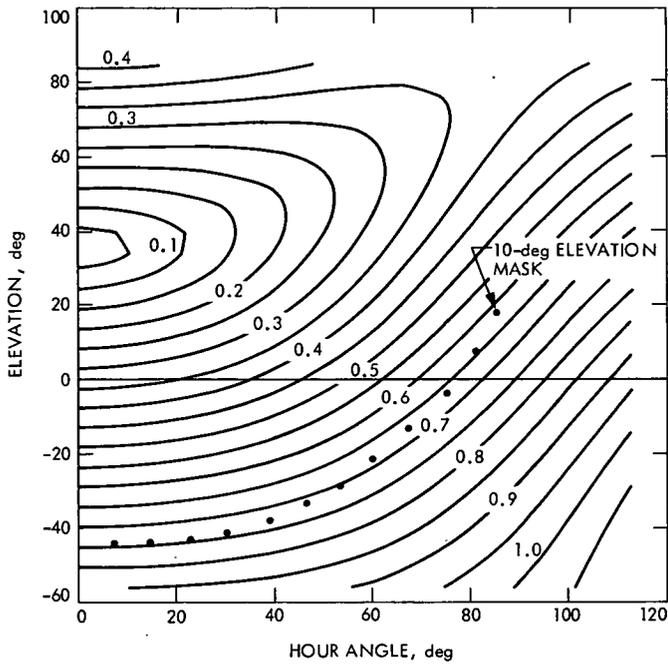


Fig. 2. DSS 12 surface panels set at zenith look; panels supporting points; rms distortion contour map, mm

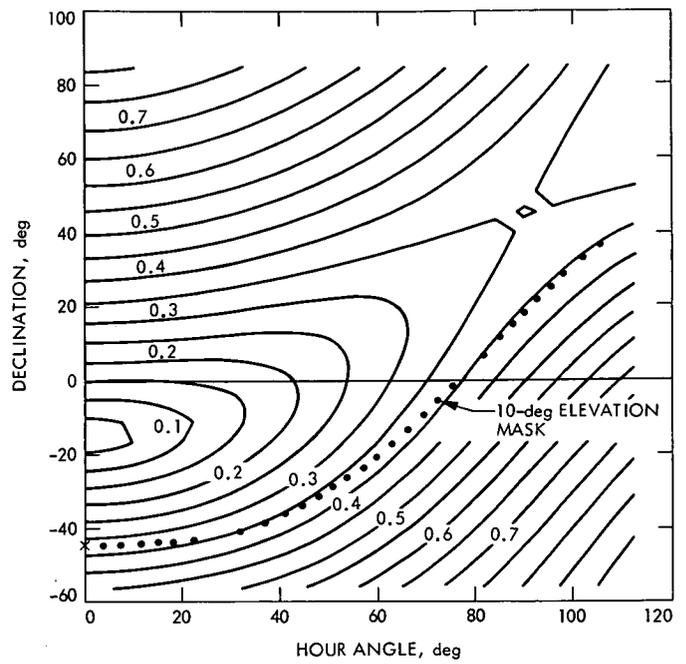


Fig. 4. DSS 12 surface panels set at -15 deg declination; panels supporting points; rms distortion contour map, mm

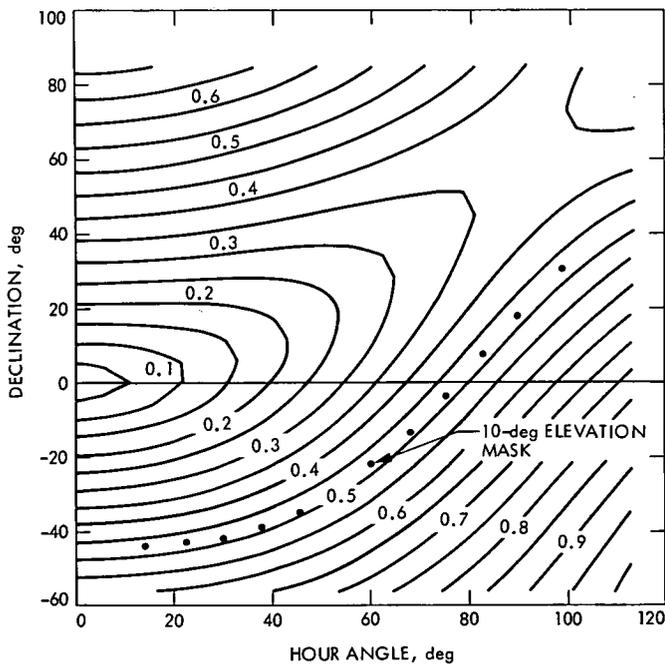


Fig. 3. DSS 12 surface panels set at 0 deg declination: panels supporting points; rms distortion contour map, mm

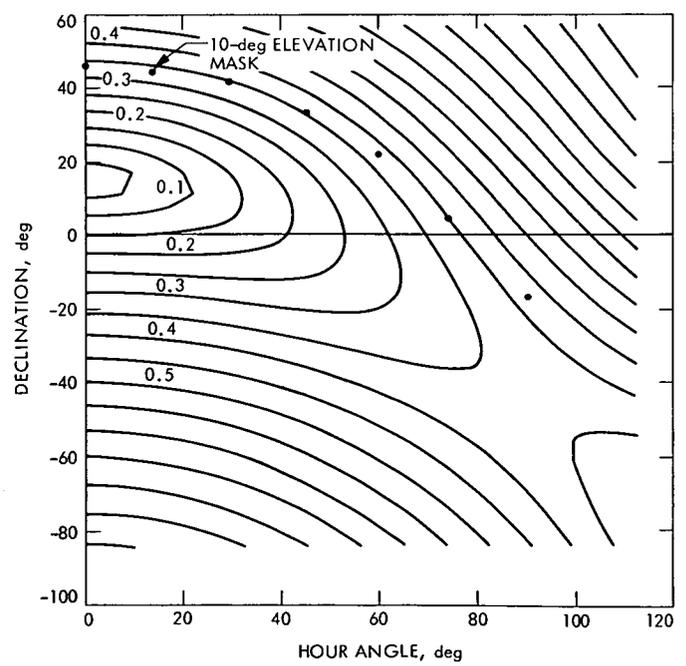


Fig. 5. DSS 42 surface panels set at -15 deg declination; panels supporting points; rms distortion contour map, mm

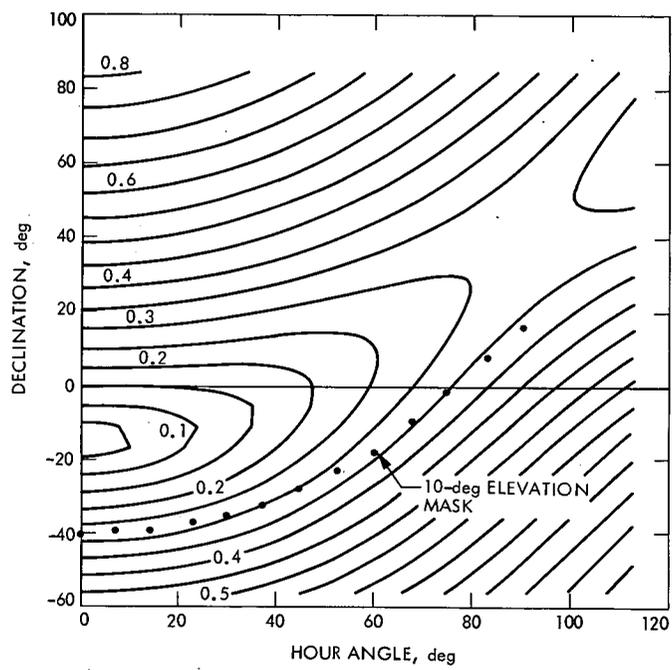


Fig. 6. DSS 61 surface panels set at -15 deg declination; panels supporting points; rms distortion contour map, mm

Interim Radio Spectrum Surveillance Station

D. R. Hersey

Telecommunications Systems Section

Radio frequency interference at NASA's deep space stations has become a serious problem. A radio spectrum surveillance capability at these stations is needed to determine the sources of interference so that preventative measures can be taken. The first phase of a program to develop this capability was the development of a low-cost surveillance station now in operation at the Goldstone Deep Space Communication Complex near Barstow, California. This interim surveillance station is described and findings from the use of this equipment are presented.

I. Introduction

In recent years there has been a significant increase in the number of occurrences of radio frequency interference (RFI) at the NASA deep space tracking stations. The interference in some cases degraded telemetry received from deep space spacecraft and in a few cases has caused the receivers to lose lock on the spacecraft carrier signal. Attempts to identify the sources of the interfering signals have been severely handicapped because the stations have no means of determining the direction of the interfering source and very limited capability to view the interfering signal spectrum. Further, since the prime function of the station is to provide spacecraft data, the major activities of the station operators at times of signal degradation are to verify that a malfunction of station equipment has not occurred; to reacquire the spacecraft signal, if necessary; and to improve the quality of the spacecraft data.

To facilitate the identification of RFI sources so that preventative measures can be taken, a radio spectrum surveillance station has been developed and installed at the Gold-

stone Deep Space Communications Complex near Barstow, California. Because of the immediate need for RFI surveillance, a low-cost system using mostly off-the-shelf equipment was developed to provide surveillance capability. Later, a more sensitive and sophisticated system will be implemented. Its development will be influenced by the observations of this low-cost system.

The purpose of this presentation is to describe the interim surveillance system and to present findings from field use of this equipment.

II. Background

Before describing the surveillance station, a brief discussion of the interference protection criteria for the deep space stations will point out the difficulty in providing them with interference protection. For brevity, the discussion is limited to the 64-m antenna stations.

Also some motivating factors for implementing the surveillance station and a general overview of the equipment cost and schedule are presented.

A. Interference Criteria

There are three major interference criteria for the deep space stations. These criteria were established to prevent receiver saturation and degradation of carrier tracking and telemetry performance.

Power levels of -120 dBW or stronger in the band of 2290 to 2300 MHz will produce saturation of the DSN receiver. This condition causes an increase in the receiver-system noise temperature because of gain compression of the maser and can cause spurious signals to be introduced in the receiver because of operation in a nonlinear region of one or more stages of the receiver. Assuming that a signal source producing a -120-dBW signal level at the receiver is 2 deg off the main beam axis of the station's antenna, the power flux density (pfd) at the antenna corresponding to this power level is -116 dBW/m², using an off-axis antenna gain of:

$$G = 32-25 \log \phi \quad 1 < \phi < 48$$

$$G = -10 \quad \phi \geq 48$$

where ϕ is the angle from the main beam axis in degrees and G is in dBi (Ref. 1). For the source having angles equal to or greater than 48 deg from the main beam axis, a pfd of -81 dBW/m² is required to saturate the receiver.

The second criterion is a continuous wave (CW) signal no greater than -212 dBW in any 10-Hz band between 2290 and 2300 MHz. This amplitude of interfering signal will induce phase modulation on the desired carrier signal of approximately 10-deg amplitude, when the frequency separation between the two signals is comparable to, or less than, the carrier tracking loop bandwidth (assumed to be 10 Hz at receiver threshold). Using target locations of 2 deg and 48 deg off the antenna main beam axis, as above, the maximum allowable power flux densities are -208 dBW/m² and -173 dBW/m².

The third criterion is a power spectral density for noise-like interference no greater than -222 dBW/Hz. A -222-dBW/Hz amplitude interfering signal will cause approximately 1-dB degradation in telemetry performance. Both this and the second criterion are based upon C.C.I.R. Recommendation 365-3 (Ref. 2) which assumes a 16-K system noise temperature and an antenna elevation angle of 30 deg. Corresponding power flux densities at 2 and 48 deg off pointing angles are -218 dBW/m²-Hz and -183 dBW/m²-Hz.

B. Motivating Factors

At the time of considering the feasibility of a surveillance station at the Goldstone facility, there was evidence that aircraft in the vicinity of the stations produced the most detrimental radio frequency interference. Calculations show that an aircraft at a 160-km range and radiating 1-mW CW towards a deep space station can produce -145 dBW/m² at the station's antenna. Comparing this value with pfd values calculated from the interference protection criteria, it can be seen that emissions from aircraft can (and do) cause very serious interference.

Of particular concern is the possibility that communications with a spacecraft will be interrupted during critical operations such as a planetary encounter or trajectory change maneuver. At these times, it is very desirable to perform radio frequency surveillance of the spacecraft downlink frequency band. At the time of the decision to implement the surveillance station, the two upcoming encounters were the Pioneer-Venus spacecraft with Venus and the Voyager spacecraft with Jupiter.

In July 1977, it was decided that, in the 18 months remaining before the Pioneer-Venus encounter, a surveillance station suitable for detecting aircraft emitters could be developed for reasonable cost. This station would provide protection against the major source of interference and yield radio environment information valuable for the development of future surveillance stations.

C. Schedule and Cost

The work on the development of the station began in July 1977, and postinstallation testing was completed in November 1978. Hardware costs were approximately \$50,000, and slightly over 3 manyears of effort were expended. Much of the equipment such as power supplies, the equipment rack, and digital recorder were on hand and not purchased.

The surveillance station did support both Pioneer-Venus spacecraft encounters and the Voyager 1 encounter.

III. General Description

The surveillance station is located on a mountain peak central to the three deep space stations and is normally unattended. In the unattended configuration, it monitors the frequency band of 2290 to 2300 MHz while scanning through 360 deg of azimuth in approximately 1 min. The receiver sensitivity allows detection of power flux densities of -137 dBW/m² or greater at the antenna of the surveillance station. Upon detecting a signal, a photograph is made showing the

signal spectrum, spectrum analyzer control settings, time, and antenna azimuth. At the same time, the equipment provides a notification signal via landline to the Goldstone Radio Spectrum Coordinator.

The surveillance station is normally operated continuously and serves a very useful function in aiding deep space station operators to determine if a radio frequency interferer is external or internal to their station. Upon occurrence of interference, the station operators may call the Radio Spectrum Coordinator to determine if a signal was detected by the surveillance station. If so, the signal source is known to be external to the deep space station. If it was not detected by the surveillance station, however, it can not be concluded that the source is internal to the deep space station for it may be an external signal too weak to be detected by the surveillance station.

The surveillance station can also be operated manually. In this mode the operator has full control of the spectrum analyzer settings, the antenna pointing direction, antenna rotation rate, and operation of the camera and digital recorder. The maximum frequency range of the equipment in this mode is 2250 to 2320 MHz and, by use of the narrowest resolution bandwidth (10 Hz) of the spectrum analyzer, sensitivities as high as -177 dBW can be achieved.

The manual mode is used primarily for real-time identification of interferers. Because the operator can determine the received signal characteristics and the direction of the source almost immediately, other agencies can be contacted while the source is still in the vicinity of Goldstone, thereby greatly facilitating the identification of the source.

A third mode of operation, in which the spectrum analyzer is removed from the surveillance station and used with a hand-held antenna, is also provided. This mode is used for investigating radio frequency sources at locations other than the surveillance station site, for example, searching for radio emissions internal to one of the deep space stations.

IV. Functional Description

Figure 1 shows a functional block diagram of the surveillance station. The heart of the system is an analog spectrum analyzer (Hewlett Packard 8568A). Requisite features of this instrument are very high frequency stability, annotation of the display with important analyzer control settings and provision to display user-supplied information. The user-supplied information in this application is time, antenna-rotation rate, and antenna azimuth. The frequency range of the analyzer is 100 Hz to 1.5 GHz, and down conversion of the S-band received signals is necessary.

The down conversion is performed by the Microwave Assembly which contains the mixer and local oscillator circuits. To maintain high frequency stability, the local oscillator signal is generated by coherently multiplying a highly stable reference signal supplied by the spectrum analyzer.

The surveillance antenna is a standard gain horn having a maximum gain of 17.2 dBi. Its half power beamwidths are 28 deg in azimuth (H plane) and 26 deg in elevation. The mounting bracket for the antenna allows the antenna elevation angle to be wrench-adjustable from -10 to $+70$ deg. Since the primary goal of this equipment is to detect the presence of aircraft emitters and sources at ground level, the antenna elevation is currently set to $+5$ deg.

To achieve a useful receiver sensitivity at moderate cost, two uncooled, low-noise amplifiers [gallium arsenide field-effect transistor (GAs FET)] are used as a preamplifier to amplify the antenna output signal. The noise figure of the preamplifier is 1.2 dB. A bandpass filter is used before the preamplifier to protect it from degradation by transmissions from a nearby deep space station and to restrict the number of mixer product terms out of the first mixer. A CW signal having a power level of -153 dBW (corresponding to a pfd of -138 dBW/m²) will produce a barely discernible signal on the spectrum analyzer (12-dB signal-to-noise ratio) when the station is configured for unattended operation. The resolution noise bandwidth in this mode is 4400 Hz.

Automatic recording of the received signal spectra is performed by a 35-mm camera triggered by a signal-to-noise ratio detector. Detecting the signal-to-noise ratio instead of the absolute signal level eliminates the problem of receiver gain variation with temperature (approximately 2 dB).

The presence of a signal-to-noise ratio which exceeds a present threshold value, in addition to actuating the camera, also:

- (1) Produces a printed record of the time and azimuth at the end of the sweep during which the signal was detected.
- (2) Notifies the Radio Spectrum Coordinator of the detection.
- (3) Increments counters at both the surveillance station and at the Radio Spectrum Coordinator's position.

Figure 2 is a photograph of the spectrum analyzer display taken by the recording camera. Time, antenna azimuth, and antenna rotation rate displayed at the top of the display are updated at the end of each spectrum analyzer sweep. With these parameters and the analyzer sweep time (displayed at the

end of the bottom line), the antenna azimuth at which the signal was received and the time of occurrence of any signal seen on the analyzer trace can be determined by interpolation. The Display Controller controls the timing of the equipment sequences, performs the antenna rotation rate calculation, and performs the data formatting necessary for communicating with the spectrum analyzer.

The antenna motor is a DC motor with the field and armature separately excited. A switch is used in series with the armature to start, stop, or reverse the antenna rotation, and the motor speed is controlled by adjusting the armature power supply.

An S-Band signal source located 7.4 km from the surveillance station is provided for assessment of equipment performance. The source may be turned on from either the Radio Spectrum Coordinator's position or the surveillance station thereby providing a test signal whose amplitude, direction, and frequency are accurately known.

For transportable operation, the spectrum analyzer, the microwave assembly, and a hand-held antenna are used. The hand-held antenna assembly consists of a 12 dBi gain horn and a preamplifier having a noise figure of 3.1 dB. The preamplifier power supply is contained in the microwave assembly.

The performance parameters of the equipment are listed in Table 1.

Photographs of the major station elements are shown in Figs. 3 through 5.

V. Operational Findings

Since installation of the equipment in November 1978, there have been several hundred photographs taken by the equipment of radio spectra having frequency components in the deep-space-station downlink receive frequency band (2290 to 2300 MHz). Although cataloging of these photographs is just beginning and many of the radio frequency sources have not yet been identified, some important facts have been learned.

The results to date indicate that the number of detections correlate with military exercises in the vicinity of the Goldstone Deep Space Station Complex with the greatest number of detections associated with the use of FAA airspace R2508. During periods of critical deep space spacecraft operations, such as planetary encounters or spacecraft emergencies, military agencies are requested to impose radio silence on frequencies near the deep-space receive band and, during these periods, no sources have been detected. Further, it has

been found that the weekends and evening hours, the hours not usually used for military exercises at the aircraft gunnery ranges, have the fewest number of detections. It should be pointed out, however, that, because of the large amount of effort required to identify potential interference sources, only sources which are considered as having a high RFI potential have been investigated. Of those investigated, approximately 65 percent have been identified. The identified sources are all military aircraft.

As an example of the benefit of the surveillance station at the Goldstone Deep Space Station Complex, it was learned by use of this equipment that a usual practice on board certain Air Force aircraft is to test some electronic equipment for a few minutes shortly after takeoff. The emissions from this equipment were first detected by the surveillance station at a range of approximately 170 km. From the surveillance-station photographs, the Air Force by very conscientious efforts identified the aircraft involved and modified operational procedure to eliminate this potential interference source.

VI. Room for Improvement

Although the unattended operating mode has satisfied the original requirements levied upon it to provide records showing the spectra and azimuth of potential interferers, there is a clear need to incorporate additional features which will facilitate the identification of the interferers.

Identifying radio frequency sources is much like fighting a fire, the sooner you get to it the better are your chances. To solicit help from other agencies in identifying a radio frequency source, the Radio Spectrum Coordinator must have time, azimuth, and a description of the signal spectrum. With the present design, time is available to him, if he notes the time upon hearing the audible alarm. If he is not present at the time of detection, time is printed at the surveillance station (approximately 20 minutes away). Azimuth is not supplied to the Radio Spectrum Coordinator's position, but this too is printed at the surveillance station.

The most serious delay is that he must wait for development of the camera film to see the received signal spectrum. This can require from 6 hours to several days, depending on the quantity of film to be processed and to which facility it is sent for processing.

A real-time display containing time, azimuth, and spectrum at the Radio Spectrum Coordinator's position is the obvious solution. Since this position is not staffed on a 24-hour basis, the display device should be capable of making a permanent record.

Another deficiency is that the fixed spectrum-analyzer control settings used in the unattended mode do not always provide an optimum display of the received signal spectrum. In several instances, the interfering signal carrier was outside the 2290- to 2300-MHz range used and only sidebands falling within this range were photographed. Increasing the operating range results in lower frequency resolution and violates other requirements of the station.

This problem can best be solved by providing adaptive control of the spectrum analyzer either by the Radio Spectrum Coordinator, by software routines in the Display Controller, or by both.

VII. Conclusion

The surveillance station described is providing highly useful experience in surveillance system requirements and techniques. It has identified several particular interferers and is yielding spectral signature data which, after cataloging, will serve as a library for rapid identification of frequently seen interference.

It has helped deep-space station operators to determine whether or not RFI sources are external to their station. It has aided the designers of a more sophisticated surveillance station, employing fast Fourier transform techniques, in developing processing algorithms for their system.

References

1. International Radio Consultative Committee (C.C.I.R.) Report 391-2, "Radiation Diagrams of Antennae for Earth Stations in the Fixed Satellite Service for Use in Interference Studies," *Fixed Service Using Communications Satellites*, Vol. IV, XIII Plenary Assembly, Geneva, 1974, Published by the International Telecommunication Union, Geneva, 1975.
2. International Radio Consultative Committee (C.C.I.R.) Recommendation 365-3, "Frequencies, Bandwidths and Interference Criteria for Manned and Unmanned Deep Space Research," *Space Research and Radio Astronomy*, Vol. II, XIVth Plenary Assembly, Kyoto, Japan, 1978, Published by the International Telecommunication Union, Geneva, 1979.

Table 1. Performance parameters

Parameter	Operating Mode			Conditions
	Unattended	Manual	Transportable	
Frequency coverage, MHz	2290-2300	2250-2320	2250-2320	
Antenna rotation rate, RPM	≈1	0-4	N/A	
Receiver sensitivity, dBW	≤-153	≤-177	≤-156	4.4-kHz resolution noise B.W. 14-Hz resolution noise B.W. 400-Hz resolution noise B.W.
Detection sensitivity, dBW/m ²	≤-137	≤-164	≤-142	90% confidence with above resolution bandwidths
Source azimuth accuracy, deg	±20 at -135 dBW/m ²	±7 at -140 dBW/m ²	±12 at 138 dBW/m ²	95% confidence stationary target
Frequency readout accuracy, MHz	±0.2	±0.06	±0.06	
Amplitude accuracy, dB	±5	±4	±4	
Time of detection accuracy, s	±1	±1	N/A	
Permanent record capacity (records)	>720	>720	N/A	100-ft camera roll
False-alarm rate	≤1 per day	N/A	N/A	SNR threshold setting 12.5 dB

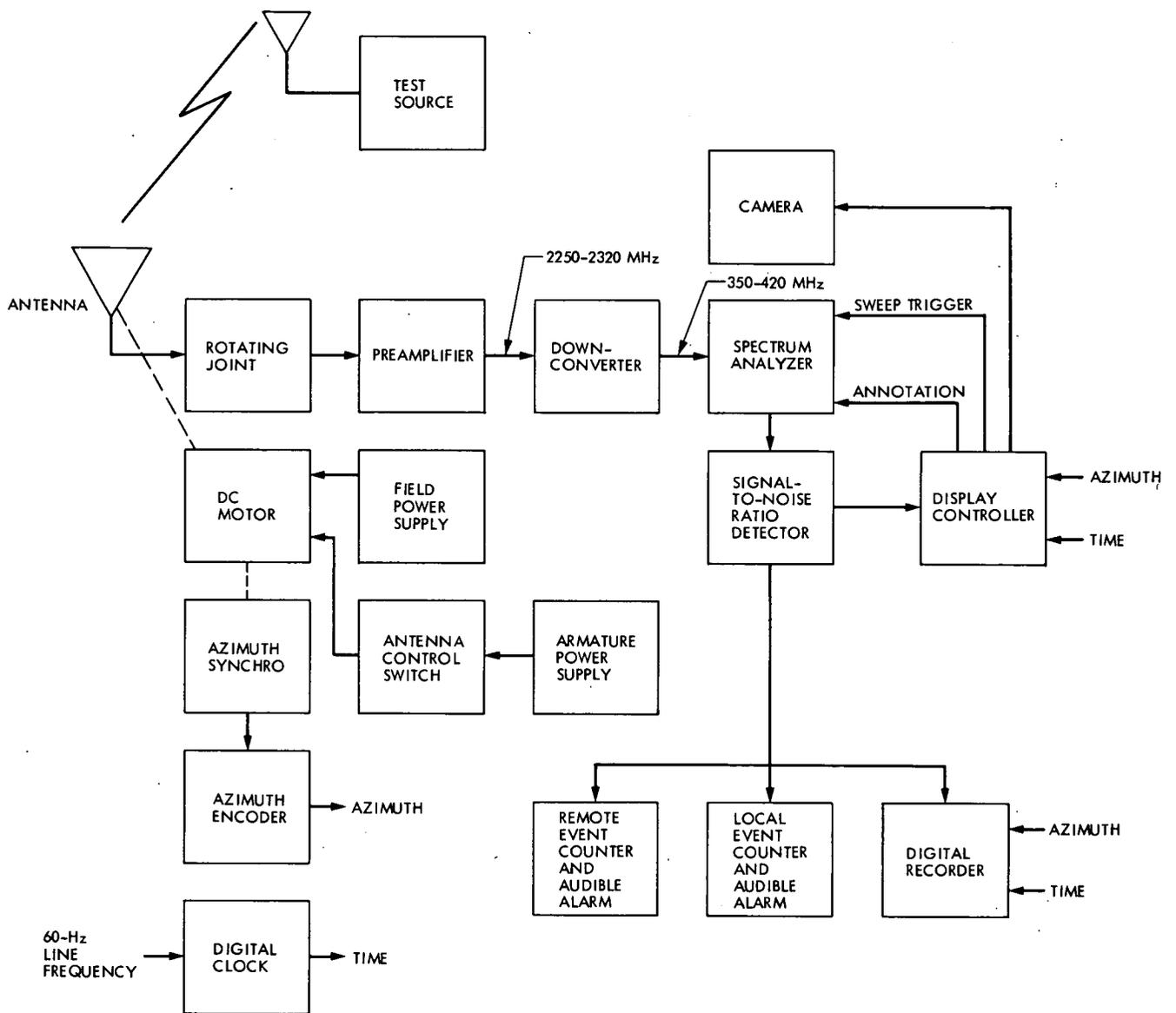


Fig. 1. Block diagram

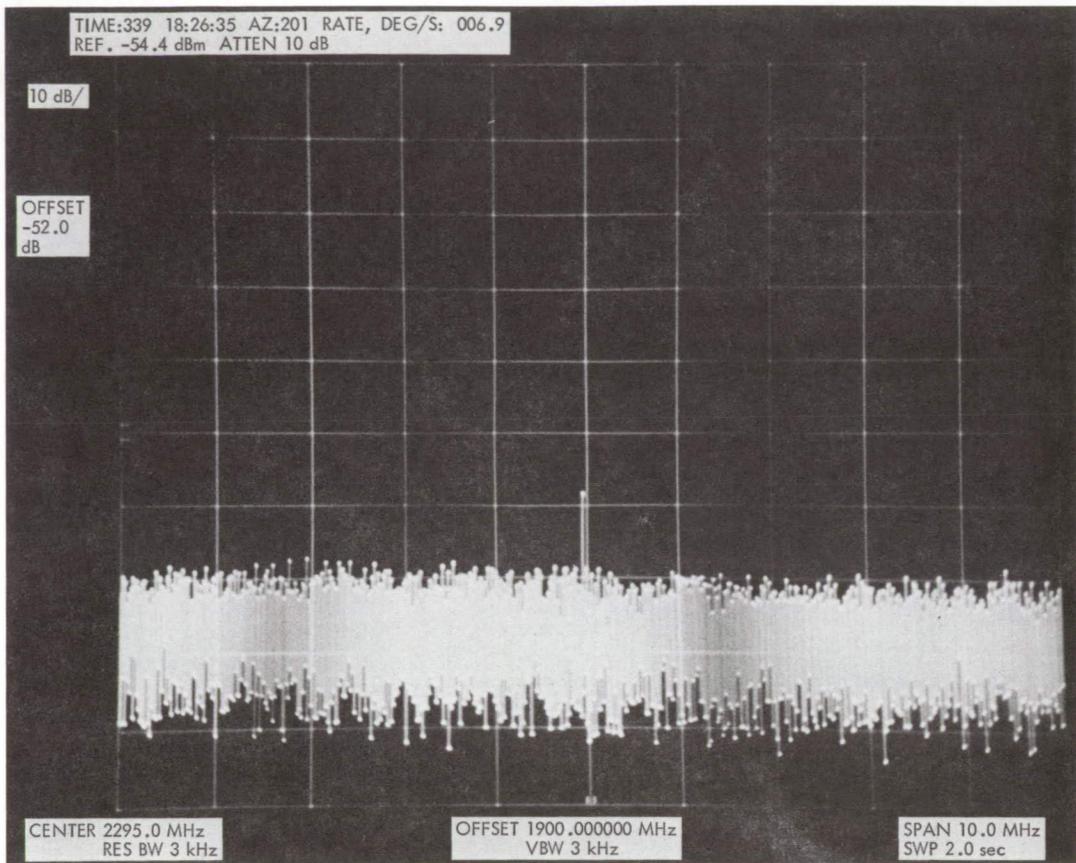


Fig. 2. Recorded spectrum

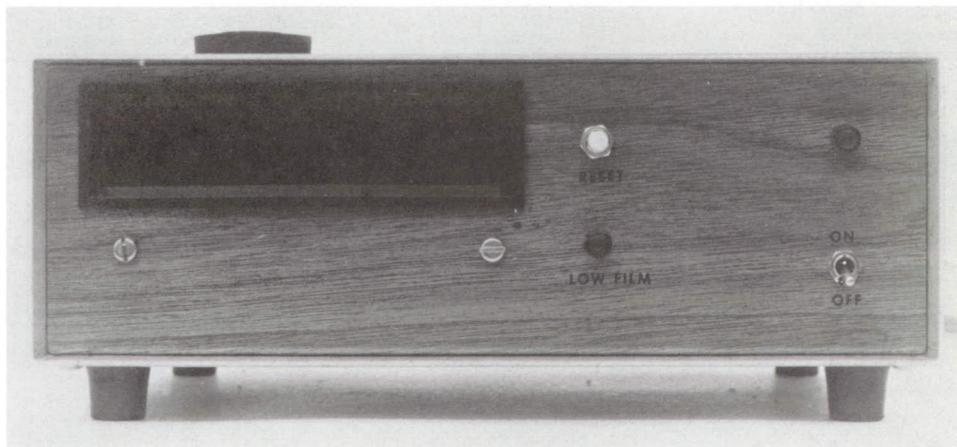


Fig. 3. Radio spectrum coordinator's display

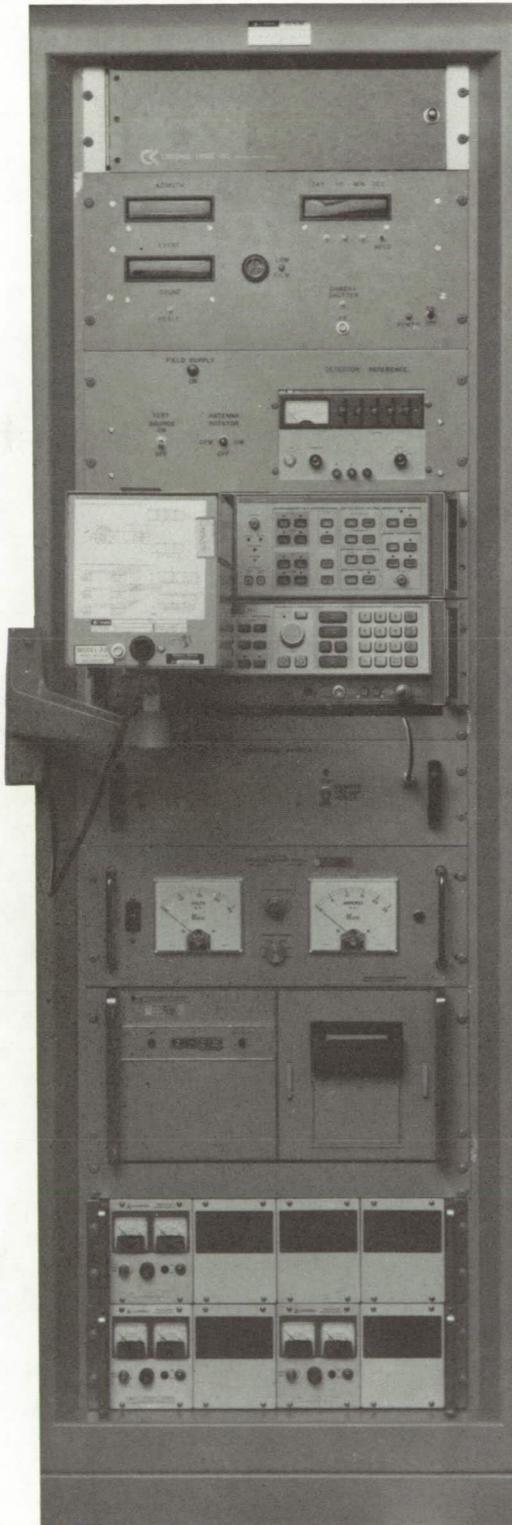


Fig. 4. Operator's console

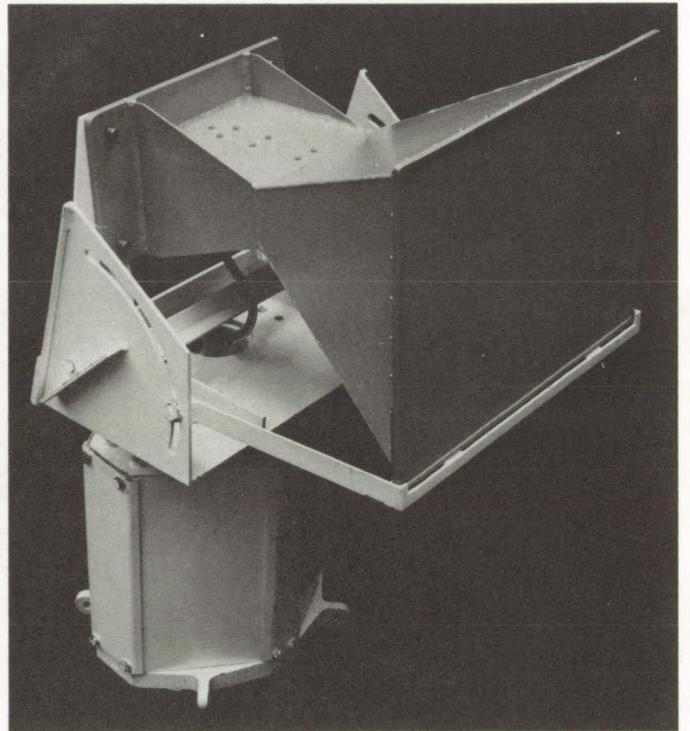


Fig. 5. Surveillance antenna

X-Band Atmospheric Noise Temperature Data and Statistics at Goldstone, DSS 13, 1977-1978

S. D. Slobin, K. B. Wallace, M. M. Franco, and E. M. Andres
Radio Frequency and Microwave Subsystems Section

O. V. Hester
Telecommunications Systems Section

X-band noise temperature data have been taken at Goldstone DSS 13 continuously since August 1975. Presented here are sample data and cumulative distributions of atmospheric noise temperature increase above the quiescent baseline for the calendar years 1977 and 1978. Comparison is made with the existing Deep Space Network noise temperature statistics.

I. Introduction

Reference 1 presents statistics of zenith noise temperature increase above the quiescent (undisturbed) baseline at Goldstone for the period August 1975 through January 1977. Included also are descriptions of elevation angle modelling and determination of atmospheric attenuation values from the given noise temperature increases. This article summarizes the X-band noise temperature statistics for the complete calendar years 1977 and 1978.

II. System Description

X-band atmospheric noise temperature data at DSS 13 (see Fig. 1a) are taken using an X-band noise-adding radiometer and horn antenna (no reflector) with a beamwidth of 15 degrees. Numerous other instruments at the site provide measurements of rainfall, temperature, relative humidity, atmospheric pressure, and solar radiation. Data points are

recorded every 2 minutes. The X-band horn is pointed at an elevation angle of 30 degrees and an azimuth of 325-degrees. A complete description of the X-band radiometer system is given in Ref. 2.

The rain gauge at DSS 13 is an MRI (Meteorological Research, Inc.) tipping bucket unit. A funnel collects rain, which is channelled to the tipping bucket. The bucket tips only after it has collected the equivalent of 1/100-inch (0.25 mm) of rain. Each tip empties the bucket and registers as a step on the cumulative rainfall data (see Fig. 1b). A very light rain could continue for an extended period and only be registered as a single step (after 0.01-inch had fallen) in the rain gauge data. Thus, the rain gauge gives inaccurate rain rate values for very light rain. The cumulative amount will be correct, however, to within 0.01-inches. Heavy rains (>10 mm/hr) will cause the bucket to empty more than once during one 2-minute data period. There is little loss of accuracy at these large rates.

III. Noise Temperature and Rainfall Data

Figure 1 shows a five-day segment of raw antenna noise temperature and cumulative rainfall data taken during January 1978. This is typical of a rainy period of the year. The characteristic "spiked" rain signature corresponds almost perfectly to the cumulative rainfall data shown directly below. The intensity of rain, or rain rate (mm/hr), can be determined from the slope of the cumulative rainfall curve. During the period 18-19 hours on day 16 the rain rate is nearly constant (~ 9 mm/hr), yet the noise temperature exhibits erratic changes. Possible causes of this could be alternate "beading" and "running-off" of raindrops on the antenna horn aperture, cover, or the radiometer "seeing" rain which is not actually falling into the rain gauge. Further investigation of rain and its effects on microwave systems will be carried out. Correlation of rain and noise temperature effects and the development of a weather/microwave model will be the subject of a future DSN Progress Report article. Figure 2 shows the correspondence of rain rate (slope of curve in Fig. 1b) and increased antenna noise temperature for day 16.

As described in Ref. 1, the noise temperature measured at an elevation angle other than zenith is referred to zenith for purposes of statistical consistency. Note in Fig. 1 that the baseline system antenna temperature is approximately 9 kelvins for 30-deg elevation angle. This includes the minimum clear sky contributions from water vapor and oxygen and the noise temperature of the ground, cosmic background, and waveguide. Since the measurement of interest is the *zenith increase* above baseline due to atmospheric variables, the baseline is subtracted from the antenna temperature and the remainder divided by $1/\sin(\text{elevation angle})$ as a "weather-mass" correction. As an example, Fig. 3 shows the zenith equivalence (above baseline) of the data in Fig. 1 from Day 16, 08 hours to Day 17, 08 hours. Times in both figures are referenced to local midnight (PST). The 67 kelvin peak at 30 deg elevation in Fig. 1 becomes a 29 kelvin peak above baseline at zenith [$(67-9) \div 2 = 29$]. Nine-kelvin antenna temperatures in Fig. 1 become equivalent zenith zero-kelvin increases above quiescent baseline in Fig. 3.

Figure 1 also shows typical clear weather (Day 15, 22 hours) and cloudy weather (Day 14, 06 hours) noise temperature measurements. The small noise temperature "humps" occurring at 10-14 hours were probably due to noise diode thermal effects and have since been eliminated and do not appear in the noise temperature statistics.

IV. Noise Temperature Statistics

Statistics have been generated for equivalent zenith X-band atmospheric noise temperature increase above quiescent base-

line for various year-quarters and day-quarters at Goldstone. "Equivalent" means that the data are taken at elevation angles other than zenith and are converted to zenith values as described above. Statistics for *system* noise temperature can be obtained by adding the baseline system noise temperature for a particular antenna and elevation angle. Thus, for DSS 14, the 64-meter antenna at Goldstone, approximately 25 K (at zenith) must be added to the atmospheric increase to obtain total system noise temperature. For example, if the *atmospheric* noise temperature increase at *zenith* is 5 K or less, 90% of the time, the total *system* noise temperature at *zenith* would be 30 K or less, 90% of the time. Different antenna elevation angles contribute different ground and clear sky atmospheric contributions and thus the baseline system noise temperature will be a function of elevation angle. A further discussion of elevation angle modelling and attenuation calculation is given in Ref. 1.

Figure 4 shows the cumulative distribution of zenith noise temperature increase for a very wet period in 1978 (first year-quarter, January-March) and for a very dry period in 1977 (second year-quarter, April-June). These curves have 1 K resolution. As an example for reading the curves, for the wet quarter, 95% of the time the zenith noise temperature increase was less than 5 K above the baseline. The wet/dry comparison shows that about 3 times as much data lie above 5 K in the wet quarter as in the dry quarter. Some amount of time the radiometer records values less than the accepted baseline value. This is due primarily to radiometer drift about the "zero-K" point. The baseline could alternatively be defined as that value below which only 5% or 1% of the data lie.

Figure 5 shows a comparison of distribution functions for total-year Goldstone data in 1977 and 1978. In 1978, about 70% more data lie above 5 K than in 1977. Indeed, 1978 was a much wetter year than 1977. Also shown in Fig. 5 is the total-year Goldstone noise temperature distribution function from the DSN Flight Project Interface Design Handbook. The data used to generate the Fig. 5 curve are pre-1970 and it is not known how these data compare descriptively (wet or dry) to the 1977 (very dry) and 1978 (relatively wet) data, although the curve indicates substantially worse weather than in either 1977 or 1978. As has been noted before (Ref. 3) the data appear to be "pessimistic," i.e., they indicate that the weather effects are far worse than they appear to be based on actual measurements.

Since a presentation of distributions for every year/day-quarter for 1977 and 1978 would be prohibitively lengthy, only tables of year-quarter and year-total cumulative distributions are presented. Tables 1 and 2 present the 1977 and 1978 Goldstone cumulative distributions, respectively, up to the 10 K and/or 99% level.

References

1. Slobin, S. D., et al., "X-Band Atmospheric Noise Temperature Statistics at Goldstone DSS-13, 1975-1976," in *The DSN Progress Report 42-38*, Jet Propulsion Laboratory, Pasadena, Calif., Apr. 15, 1977, pp. 70-76.
2. Reid, M. S., et al., "An X-Band Radiometer for the Microwave Weather Project," in *The DSN Progress Report 42-29*, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1975, pp. 54-59.
3. Greenhall, C. A., "Examination of the DSN X-Band Weather Specifications," in *The DSN Progress Report 42-45*, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1978, pp. 197-208.

Table 1. Cumulative distribution of zenith atmospheric noise temperature increases above Quiescent Baseline at Goldstone, 1977

TZX ^a	1st year quarter	2nd year quarter	3rd year quarter	4th year quarter	Year total
0	.487	.464	.061	.017	.300
1	.995	.955	.985	.924	.956
2	.999	.980	.992	.951	.976
3	.999	.984	.992	.959	.980
4	1.000	.985	.992	.965	.982
5	1.000	.986	.992	.970	.984
6	1.000	.988	.992	.975	.987
7	1.000	.989	.992	.978	.988
8	1.000	.990	.993	.982	.990
9	1.000	.991	.993	.985	.991
10	1.000	.991	.993	.988	.992

^aTZX = X-band zenith atmospheric noise temperature increase above quiescent baseline, K; e.g., 2nd year quarter, TZX = 3, 98.4% of data below 3 K increase

Table 2. Cumulative distribution of zenith atmospheric noise temperature increases above quiescent baseline at Goldstone, 1978

TZX ^a	1st year quarter	2nd year quarter	3rd year quarter	4th year quarter	Year total
0	.207	.053	.694	.512	.216
1	.735	.332	.783	.964	.712
2	.879	.947	.945	.980	.931
3	.914	.981	.981	.986	.959
4	.933	.983	.984	.988	.968
5	.949	.986	.987	.990	.975
6	.962	.988	.989	.992	.980
7	.971	.990	.992	.993	.985
8	.979	.992	.994	.993	.988
9	.984	.994	.995	.994	.991
10	.987	.995	.996	.995	.992

^aTZX = X-band zenith atmospheric noise temperature increase above quiescent baseline, K; e.g., year total, TZX = 5, 97.5% of data below 5K increase.

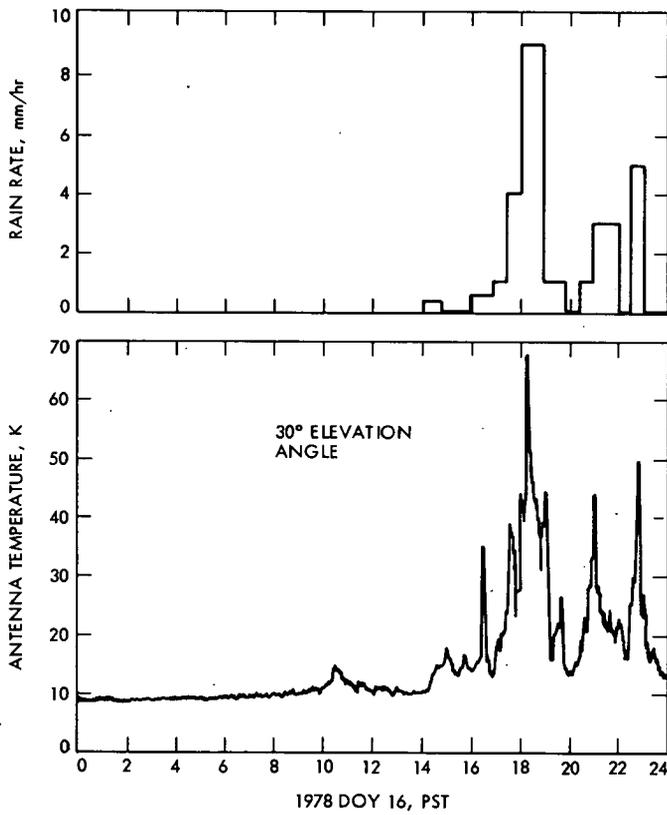


Fig. 2. X-band antenna temperature at Goldstone DSS 13 and corresponding rain rate, January 1978

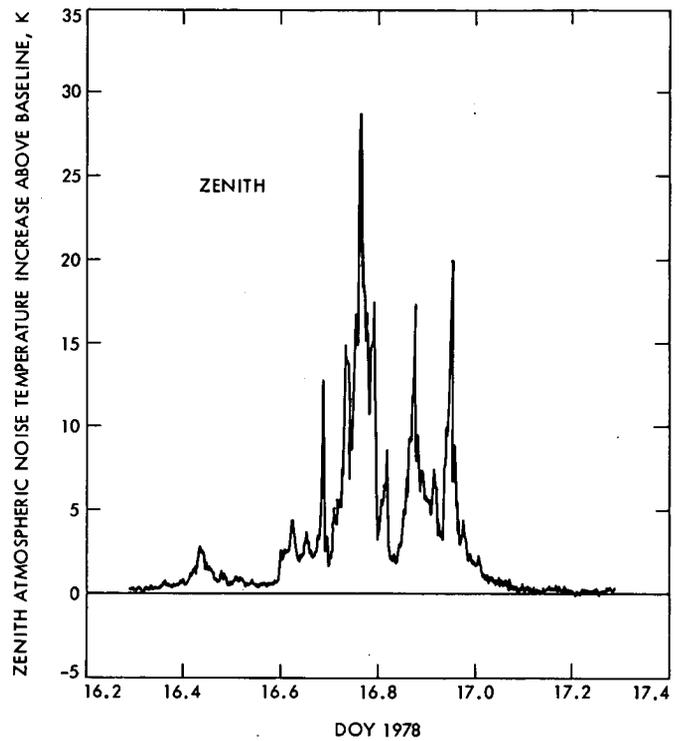


Fig. 3. Zenith equivalent X-band atmospheric noise temperature increase above baseline, K, corresponding to Figs. 1 and 2, January 1978

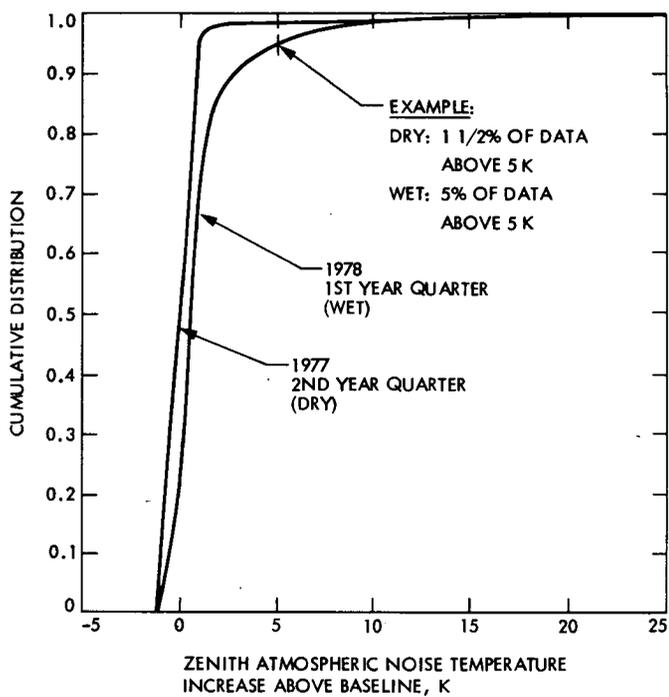


Fig. 4. Wet-quarter/dry-quarter atmospheric noise temperature statistics at Goldstone DSS 13

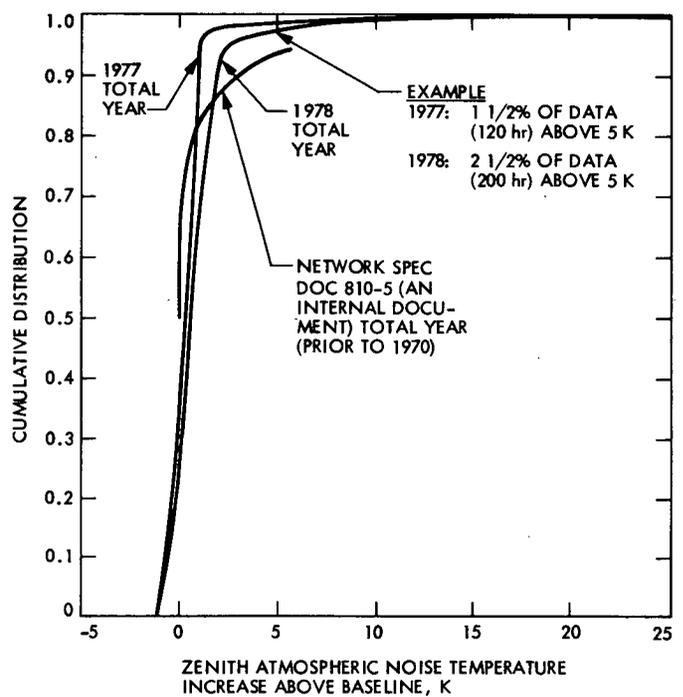


Fig. 5. 1977/1978 atmospheric noise temperature statistics at Goldstone DSS 13 and existing 810-5 distribution

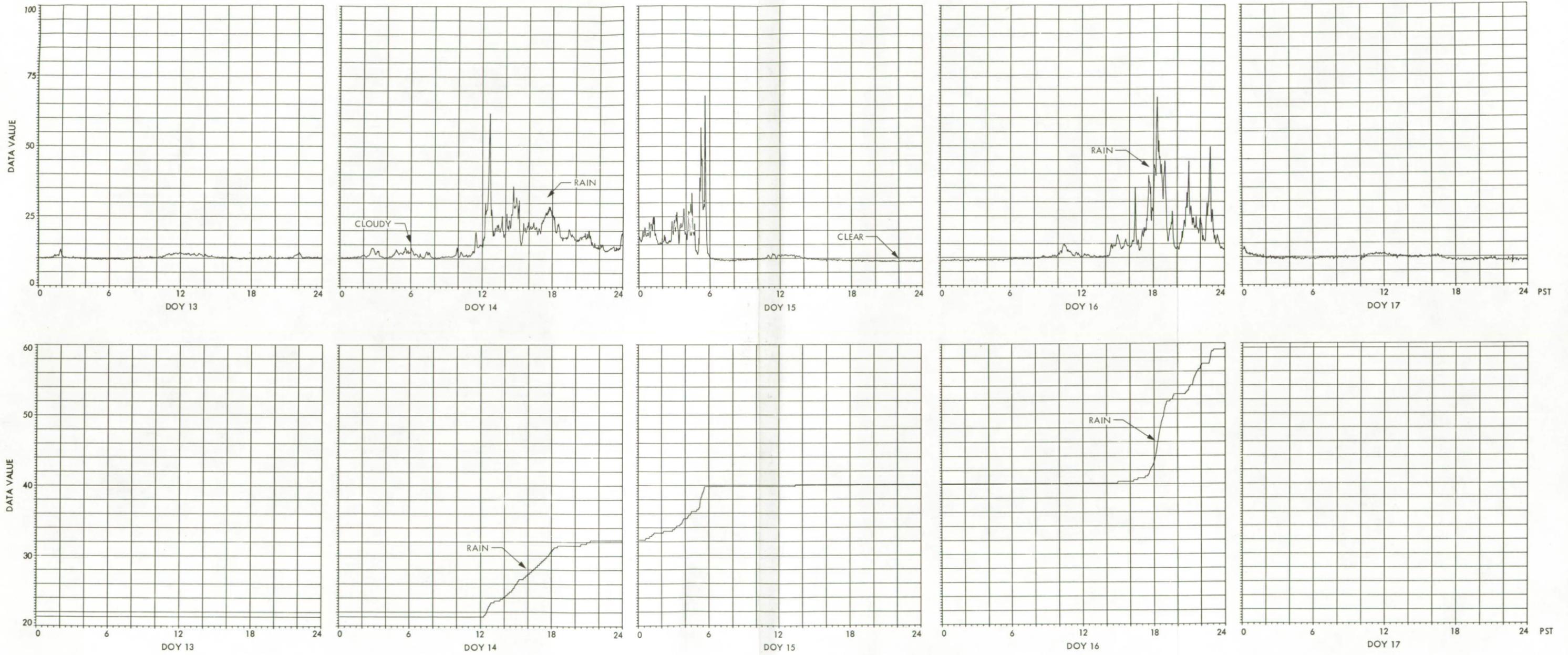


Fig. 1. Raw noise temperature and rainfall data from Goldstone, January 1978, (upper row) X-band antenna noise temperature, K, 30-degree elevation angle, (lower row) cumulative rainfall, mm



