FK

(NASA-CR-157310) THE DEEP SPACE NETWORK Progress Report, Mar. - Apr. 1978 (Jet Propulsion Lab.) 312 p HC A14/MF A01 CSCL 22D

N78-28108 THRU N78-28134 Unclas G3/12 26889

The Deep Space Network Progress Report 42-45

March and April 1978

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The Deep Space Network Progress Report 42-45

March and April 1978

June 15, 1978

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Prepared Under Contract No NAS 7-100
National Aeronautics and Space Administration

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

MISSION SUPPORT

Ongoing Planetary/Interplanetary Flight Projects

SUPPORTING RESEARCH AND TECHNOLOGY

Tracking Ground-Based Navigation

Communications—Spacecraft/Ground

Station Control and Operations Technology

Network Control and Data Processing

NETWORK AND FACILITY ENGINEERING AND IMPLEMENTATION

Network

Deep Space Stations

OPERATIONS

Network Operations

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 provides the project management, the Lander spacecraft, and conducts mission operations, and for which JPL provides the Orbiter spacecraft; Voyager, for which JPL provides project management, spacecraft development, and conduct of mission operations; and Pioneer Venus, 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 Jupiter Orbiter Probe Mission, for which JPL provides the project management, spacecraft development and 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:

- 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.
- The DSN Telemetry System receives, decodes, records, and retransmits engineering and scientific data generated in the spacecraft to Mission Control.
- 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.
- 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.

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

- 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.
- 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 three 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:

- Control and coordination of Network support to meet commitments to Network users.
- Utilization of the Network data processing computing capability to generate all standards and limits required for Network operations.
- 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 operations 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:

- Processing of data used by Network Operations Control for control and analysis of the Network.
- Display in the Network Operations Control Area of data processed in the Network Data Processing Area.
- Interface with communications circuits for input to and output from the Network Data Processing Area.
- 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.

Viking Extended Mission Support

R. L. Gillette
Deep Space Network Operations Section

This report covers the period from 1 January through 31 February 1978. It reports on DSN support of Viking spacecraft activities during the period and continues reporting on the DSN Viking Command and Tracking support. It also continues the reports on the status of DSN Mark III Data (MDS) Subsystem Implementation Project related Viking testing.

I. Viking Operations

A. Status

The two Viking Orbiters continued to make detailed photo maps of Mars during this reporting period. Viking scientists received close-up photos of surface feature details never before seen, as the planet slowly moved toward an alignment with the Earth and Sun.

As well as making detailed photographs of Mars, the Viking Orbiters also continued to check Martian cloud patterns, temperatures, and water vapor in the atmosphere. The Viking Landers continued their investigation into the composition of the Martian soil. Both Landers pushed rocks aside and picked up subrock soil samples for chemical analysis.

Photo mapping of the Mars surface by the Viking Landers also continued during this reporting period. Photographs of the Mars surface under one of Lander 1's descent rocket motors was taken using mirrors on the scoop mechanism.

On Jan. 19, 1978 the signal received from the Viking spacecraft reached its maximum signal level as the Earth-Mars distance reached a minimum of 99.7 million kilometers with a one-way-light-time of 5 min and 26 s.

B. Spacecraft Problems

On Feb. 8 the Viking Orbiter 2 spacecraft developed a major gas leak in the roll axis stabilization system. Despite vigorous corrective activity, the leak continued for more than two days. During this time, the leak was variable ranging from 100 g per day up to as much as 680 g (1.5 lb) per day. During the two days, about 900 g (2 lb) total of attitude control gas was lost. Prior to the leak, Orbiter 2 was using between 22 and 36 g of gas per day for normal operations. During the remaining weeks in February the Orbiter 2 spacecraft continued to experience an intermittent gas leak.

C. Radio Science

Radio Science activities and experimentation continued during January and February. These activities include near simultaneous Lander/Orbiter ranging, Orbiter 1 and 2 Earth occultation coverage, bistatic radar passes, and the Gravity Wave experiment passes.

II. Network Support

Table 1 shows the Viking Extended Mission (VEM) Tracking Support for January and February 1978. Noticeable during this period is the reduction of 64-m station support (DSS 14,

43, 63), from January to February. This will be a continuing trend as other Deep Space projects such as the Voyager mission to Jupiter and Saturn require the 64-m network.

III. DSN Mark-III Data Subsystem Implementation (MDS) Testing and Status

As indicated in the last report of this series, MDS test and training has been completed at all DSN stations except DSS 11 at Goldstone and DSS 61/63 in Spain. DSS 11 was released on 15 January to begin their MDS upgrade. DSS 61/63 completed their test and training during this reporting period.

A. DSS 61/63 Test Status

Test and training activity for Viking started in January and consisted of three Operational Verification Tests (OVTs), a DSN/MCCC System Interface Test (SIT), and a Viking Ground Data System (GDS) Test.

Two out of the three Viking OVTs were considered successful. The first OVT completed only 50% of the Sequence of Events (SOE) due to problems in the Simulation Conversion Assembly (SCA). The SCA is used by the DSN to simulate a spacecraft and generate simulated spacecraft telemetry data.

The SIT test as well as the first GDS test was also declared unsuccessful due to problems related to the SCA. On January 12 a successful Viking GDS retest was conducted at DSS 61/63.

After several successful demonstration passes with the Viking spacecraft, DSS 61/63 was placed under Viking configuration control effective January 31, 1978.

Table 1. VEM Tracking Support for Jan. and Feb., 1978

DSS	January		February	
	Tracks	Hours	Tracks	Hours
11	8	72		
12			4	13
14	40	319	25	218
42	22	166	23	76
43	57	294	14	207
44			11	51
61	10	88	38	243
62	6	27	2	25
63	13	101	21	181
Total	156	1067	138	1014

Note: Number of tracks represent the summation of all Viking spacecraft tracked. Track time represents scheduled station support.

Table 2. Number of Commands Transmitted in Viking Extended Mission During Jan. and Feb. 1978

DSS	No. of Commands	
	Jan.	Feb.
11	1947	
12		1
14	4565	1079
42	1447	1305
43	1593	1732
44		255
61	992	3548
62	1	1006
63	895	128
Total	11440	9054

Deep Space Network to Viking Orbiter Telecommunications Performance During the Viking Extended Mission, November 1976 Through February 1978

F. H. J. Taylor
Telecommunications Systems Section

This article discusses the telecommunications performance during the Viking extended mission, starting with the completion of the first superior conjunction in November 1976 and ending with the jettison of the VO-2 aft bioshield early in March 1978. Continued Viking operations are planned through February 1979, concluding with the second superior conjunction. Included in this article are Viking Orbiter activities and problems, the ground system activities and problems, radio science activities, and communication link performance. The substantial involvement of the Deep Space Network and the coordination of their telecommunications planners with Viking Project telecommunications analysts in the planning and execution of complex Viking Orbiter sequences are discussed.

I. Introduction

Viking-1 was launched on Aug. 20, 1975; Viking 2 on Sept. 9, 1975. The Mars Orbit Insertions were on June 19, 1976 for VO-1 and on Aug. 7, 1976 for VO-2. Lander 1 touchdown on Mars was July 20, 1976, and Lander 2 touchdown was September 3, 1976. The end of the primary mission was defined as Nov. 15, 1976, ten days before the first Viking superior conjunction. Viking activities involving the Deep Space Network, starting with prelaunch compatibility testing and mission and operations planning and ending with the primary mission, are discussed in the three volumes of Ref. 2. The superior conjunction telecommunication link effects are discussed in Ref. 1. Deep Space Network and Orbiter telecommunications configuration and parameters are contained in Ref. 3.

This article begins with a description of a number of Viking Orbiter extended mission activities in which the telecommunications links were heavily involved. Following that is a section describing various problems, failures, and anomalies in the Orbiters that affected the links or in which the links were used to resolve the problems. Ground system activities and ground system problems comprise the next sections, following which a number of representative radio science experiments are discussed from a communication link viewpoint. The article ends with a quantitative and graphical evaluation of S-band and X-band Orbiter-Deep Space Network link performance.

II. Viking Orbiter Spacecraft Activities

A. Viking Orbiter High-Gain Antenna Calibrations

Four times during 1977 high-gain antenna calibration sequences were programmed into the Orbiter on-board computer. The sequences each caused the high-gain antenna to execute a series of steps in azimuth and elevation, tracing a square approximately 1 deg on each side. A 64-m station was allocated to track the X-band downlink and to send the

resulting signal level data to the Mission Test Computer in the MCCC. The X-band signal level data was used, together with the telemetered azimuth and elevation position data, to ascertain the actual pointing of the Orbiter antenna relative to the predicted directions. Orbiter attitude control limit cycle motion, also telemetered, was removed from the data.

The four extended-mission calibrations include:

Date, 1977	VO	Reference star	Station
May 23	2	Jupiter	DSS 43
July 15	1	Procyon	DSS 43
September 30	1	Alkaid	DSS 63
December 2	2	Arcturus	DSS 14

Generally a high-gain antenna calibration was requested for each new attitude control system reference star used, as the antenna operated in a different portion of the azimuth/elevation angle range for each star.

The Deep Space Station activity was similar for each calibration. The X-band CONSCAN was used prior to the first calibration point and following the last calibration point. CONSCAN was otherwise turned off during the sequence to minimize signal level variations, except for a 4-min period during the fifth of the 10 points. The high-gain antenna remained at each point for 10 min, approximately, to permit an adequately long sample of X-band signal level points. Generally the station antenna control was very good, with the antenna computer controlled. For example, during the May 23 DSS 43 sequence, the X-band signal level showed less than a 0.3 dB change when CONSCAN was turned on during the fifth point. These calibrations are made with the station receiver AGC bandwidth set at medium to assure that the station receiver will not go out of lock when the Orbiter antenna changes position. Signal level changes of as much as 7 dB occur from one position to the next.

The Sept. 30 sequence was the first one in which a portion of the sequence was tracked simultaneously by DSS 63 and DSS 43. In this way, a comparison could be made of the X-band signal level variations attributable to station effects only. The results of this comparison appear in Fig. 1. That figure shows that the absolute level of the received signals varied, with DSS 14 being 2.5 dB higher than DSS 63 at any given time. This is attributed to calibration differences. The DSS 63 CONSCAN caused approximately 0.5-dB peak-to-peak variations in the signal while it was on.

B. Mars Orbit Trims (MOT)

The Mars Orbit Trims accomplished during the extended mission are summarized in the next several paragraphs.

1. **January 20, 1977: VO-1 MOT-10, DSS 63.** The Orbiter high-gain antenna was used in the flipped position to enable reception of 2-kbps high-rate telemetry data during the motor burn. A GO command for motor burn was transmitted by DSS 63.

2. **February 4, 1977: VO-1 MOT-11, DSS 42.** This maneuver also required a GO command, which was transmitted from DSS 42. Unlike MOT-10, however, no downlink signal could be received during the portion of the maneuver when the Orbiter low-gain antenna was in use and the Orbiter yawed away from the Earth to motor burn attitude.

3. **February 11, 1977: VO-1 MOT-12, DSS 42 and DSS 43.** This sunline maneuver was the third in the triad of maneuvers to obtain a close flyby of the Martian moon Phobos. DSS 42 had the uplink and DSS 43 received the downlink from VO-1.

4. **March 2, 1977: VO-2 MOT-9, DSS 14 and DSS 42.** MOT-9 occurred near the VO-2 orbit periapsis, when tracking doppler rates were high. DSS 14 lost downlink receiver lock upon selection of the Orbiter low-gain antenna, as a result of the 20-dB decrease in signal level. It was later determined that the DSS 42 uplink tuning following earth occultation increased the already high doppler rate present on the downlink, further hampering the attempts of DSS 14 to reacquire. DSS 14 was also busy with direct links from the Viking Lander.

5. **March 11, 1977: VO-1 MOT-13, DSS 63, DSS 14.** The maneuver went well. DSS 63 and DSS 14 performed all actions involving their transmitters and command modulation exactly per plan, and all expected uplink capability was available. Downlink data was also received as expected, with the minor exception that DSS 63 had had a failure of their primary S-band maser and had to use the backup maser. The 2-kbps engineering data was well above threshold for the motor burn, and good S-band and X-band doppler.

6. **March 24, 1977: VO-1 MOT-14, DSS 42.** With a roll and yaw turn before motor burn, and no plans to select the Orbiter high-gain antenna, DSS 42 had no chance to receive downlink data at motor burn. The maneuver was therefore planned to occur "in the blind," but with the telemetry data to be played back about ten hours later over DSS 63. All went according to plan.

7. **April 18, 1977: VO-2 MOT-10, DSS 43.** As with MOT-9, there was also an Earth occultation during MOT-10, further complicating the sequence. No uplink station would be available for MOT-10, as DSS 43 uplink was devoted to VO-1 for orbit determination. For MOT-10, the Project requested the

use of an X-band downlink receiver to confirm proper high-gain antenna pointing and thus correct attitude at motor burn. In addition, plans were made for possible use of the DSS 43 uplink for VO-2 if problems should occur during the maneuver sequence. The careful planning and coordination paid off: the switching of Receiver No. 4 back and forth between VO-1 and VO-2 allowed both observation of the VO-2 motor burn and a nearly continuous tracking of VO-1 doppler for orbit determination.

8. May 15, 1977, VO-1 MOT-15, DSS 43. The Project navigation team checked the orbit determination solutions carefully for the May 27 near encounter with Phobos. The orbit of Phobos had been determined with considerable precision from Mariner '71 photographs and more recent Viking Orbiter photos. However, the VO-1 periapsis was occurring over the Tharsis region of Mars, which is one of the major regions of gravity anomaly on Mars. The orbit of the Orbiter was affected each time it passed over the region; it was difficult to project distantly into the future the exact orbit. At the time the planning for a possible maneuver would have to begin the expected miss distance to Phobos was less than 12 km. This maneuver absolutely eliminated any possibility of collision with Phobos, yet left the orbit close enough to Phobos for the desired high-resolution photography.

9. July 1, 1977, VO-1 MOT-16, DSS 14 and DSS 43. Following the successful Phobos flybys, the Viking Project wished to place VO-1 into a 24-h orbit for radio science activities, with the periapsis centered over the DSS 14 and DSS 43 view period overlap. A motor burn of just 3 s would accomplish this, following a roll-yaw-roll turn set. The Earth occultation occurred just after the motor burn, with downlink data being lost until the end of the occultation just before the end of the yaw unwind. DSS 43 acquired the uplink at motor burn attitude, During the maneuver, DSS 43 was not able to lock up promptly on the 2-kbps data prior to the turns. The problem was caused by difficulty with the ground antenna polarizer, and they were unable to switch from linear polarization (used for the previous spacecraft track) to circular polarization (for Viking).

10. September 26, 1977, VO-2 MOT-11, DSS 63. MOT-11 was the first of three VO-2 maneuvers which together (a) phased the VO-2 orbit for a near-Deimos encounter to occur on October 15; (b) lowered the periapsis altitude to 300 km; and (c) changed the orbit period to 24 h. In a sense, these maneuvers were parallel in purpose to MOT-10, -11, and -12 for VO-1. DSS 63 was the uplink and downlink station for VO-2 MOT-11. However, DSS 63 inexplicably failed to lock the downlink receiver to the S-band signal after the yaw turn. Worse, the station still could not lock the S-band downlink during the subsequent 16-min period when the Orbiter was on the high-gain antenna. DSS 63 was unable to lock the X-band receiver during motor burn also, probably because their major concentration was on the S-band problem. The maneuver telemetry data was played back later, and this showed that uplink lock was maintained throughout the maneuver, and uplink events were executed properly. An investigation into the S-band problem showed that had the station used the automatic Block IV receiver sweep system, downlink lock probably could have been achieved. The navigation team report was that the accuracy of MOT-11 was excellent, with a resulting orbit period just 5 s longer than planned, and a periapsis altitude just 0.1 km different from that planned.

11. October 9, 1977, VO-2 MOT-12, DSS-14. MOT-12 was planned to occur far from periapsis, so the doppler rate should not be a problem. Downlink data during the motor burn period was essential, because a NO GO command would be transmitted if the spacecraft performance was not normal. The Viking Project Manager briefed the DSS 14 crew who would be participating in this maneuver as to the purpose and the importance of the maneuver. Proper execution would result in a near-approach to Deimos on October 15. The advance planning and coordination again paid off. All uplink and all downlink activities were accomplished properly and on time. The trajectory following the maneuver was nearly as planned, with a predicted mean miss distance to Deimos of about 24 km. A contingency Deimos avoidance maneuver which would have occurred on October 14 was not needed.

12. October 23, 1977, VO-2 MOT-13, DSS 14 and DSS 43. MOT-13 was the last propulsive maneuver to be done on either Viking Orbiter. The 13-2/3 s motor burn caused a change in doppler frequency of 150 Hz, as had been predicted. This maneuver was particularly critical in that a severe overburn would have caused the Orbiter to impact Mars. DSS 14 transmitted a GO command when the Project analysts had determined that all subsystems were operating normally. The maneuver lowered the orbit periapsis to 300 km.

Figure 2 is a plot showing how the uplink and downlink signal quantities for VO-2 MOT-13 change during a typical maneuver. The five link quantities are plotted on a common time base. They are: uplink signal level, S-band downlink level, X-band downlink level, low-rate channel signal-to-noise ratio, and high-rate channel signal-to-noise ratio. The signal levels are in decibels below 1 mW (dBm); the signal-to-noise ratios are in decibels. Six key events are noted on the plot; these are (1) roll turn start, when the Orbiter is already on the low-gain antenna; (2) yaw turn start, when the signal levels decrease as the Orbiter low-gain antenna boresight moves farther from Earth; (3) yaw turn end, and the selection of the high-gain antenna, when a 20-dB increase occurs in the S-band signal levels and the X-band signal can be received; (4) yaw unwind

start after motor burn, when the low-gain antenna is again selected, and the X-band signal disappears as the high-gain antenna boresight moves from Earth; (5) yaw unwind end, when the S-band levels are again at the same value as they were at the yaw turn start; (6) roll unwind end, which is the end of the maneuver sequence, and the high-gain antenna is again selected.

C. VO-2 Aft-Bioshield Jettison

The VO-1 aft-bioshield was separated shortly after Lander separation in 1976. The shock impact on the VO-1 radio subsystem was a momentary change in the uplink signal level, indicating a momentary out-of-lock even with an uplink present. Also, a transient occurred in the traveling wave tube helix current, which required a half hour to recover. Because of these effects, and a failure in a gyro power supply on VO-2 at Lander separation, the VO-2 aft bioshield was not separated during the Viking primary mission.

The decision was made to separate the bioshield on March 3, 1978. DSS 63 was the primary tracking station, but with DSS 61 also assigned as a backup if needed for the GO command for the sequence. The data was returned in real-time, using the 2-kbps high-rate channel, and the data was also recorded for later playback. The separation sequence went entirely without problems. The first indication of a successful separation was an abrupt change in the spacecraft two-way doppler from DSS 63. Prior to the separation, the X-band signal level had been steady at −141.4 dBm at DSS 63, the first sample after separation was at a value 7.5 dB lower, indicative that the X-band boresight had moved about 1.2 deg from Earth. Pointing was restored by the attitude control system within 45 s of separation, as indicated by restoration of the X-band signal level to its previous value.

The VO-2 aft bioshield separation is the last major spacecraft event planned on either Orbiter. With the bioshield off, the Orbiter will use slightly less gas in maintenance of its normal attitude and in maneuvers; also the science instruments will enjoy a greater field of view.

D. Viking Extended Mission and Viking Continuation Missions

The operations associated with the extended mission are planned to end June 1, 1978; the continuation mission will commence then and continue until Feb. 28, 1978. This latter date will allow observation of the second Viking Superior Conjunction, which is of great interest to the radio science experiment.

The extended and continuation missions required a strict limit on Orbiter attitude-control gas usage per week, in order to stretch the supply to the end of the continuation mission. Both Orbiters were in 24-h orbits, and both Orbiters had periapsis altitudes of 300 km by the end of 1977.

The number of Deep Space Network station passes was reduced somewhat during the extended mission and was planned to be reduced further for the continuation mission. One 64-m station pass per day and twelve 26-m passes per week for metric tracking would be provided. Large amounts of science playback, particularly the visual imaging subsystem (television) data at 16 kbps was accomplished late in 1977 and early in 1978 when communication distances were smaller, thus partially compensating for the reduced station allocation. In the continuation mission, communication ranges will again become large as during the primary mission, the 16-kbps rate will not be possible, and the television data will be limited to 15 frames per orbit.

E. Orbiter Equipment Operating Times and Operating Cycles

As of March 3, 1978, all Orbiter radio equipment is in fine working condition, and none of the redundant subsystem elements have had to be used. The radio equipment includes redundant receivers, exciters, traveling wave tubes, telemetry modulation units and command detectors.

Through March 3, 1978, the telecommunications subsystems had accumulated the following number of hours of operation

Subsystem	VO-1 hours	VO-2 hours
X-band transmitter	20,000	20,000
Relay radio and telemetry	13,500	12,500
Radio frequency subsystem	22,000	22,000

As of March 3, 1978, there have been 927 days since the launch of VO-1 and 907 days since the launch of VO-2.

III. Viking Orbiter Telecommunication-Related Problems

A. Ranging Modulation Interference to One-Way Orbiter

This effect is caused by uplink ranging modulation intended for one of the two Viking Orbiters reaching the receiver of the second Orbiter. It is, of course, only possible when both Orbiters are within the beamwidth of the Deep Space Station transmitting antenna, as was the case throughout the extended mission with both Orbiters in orbit around Mars. The relative channel spacings and power levels for the DSN-Orbiter link

also caused the effect to be present only when the interfered-with Orbiter was one-way, when the relative doppler offset between the two Orbiters corresponded to a sufficiently strong ranging sideband, and when the particular ranging sequence being transmitted produced a sufficiently strong sideband. Since all of these factors had to be present at the same time, ranging modulation interference was a relatively rare occurrence, happening perhaps a dozen or so times during 1977, each time the interference causing one or several momentary disruptions in tracking each lasting less than one minute. The stronger the uplink, the more probable the occurrence. Therefore, ranging interference occurred only when a 64-m station was transmitting the ranging modulation, and it happened more frequently as the Earth-Mars communications range decreased during 1977. A reduction in station transmitter power or in ranging modulation index could often eliminate the interference on a given occasion. To put things into perspective, VO-1 is on channel 9, with an uplink carrier frequency of 2111.6 MHz; VO-2 is on channel 20, with a frequency of 2115.4 MHz.

The first instance of ranging interference during the extended mission occurred on December 9, 1976. DSS 43 was two-way with VO-1 at a transmitter power of 50 kW. When ranging acquisitions occurred, each acquisition caused the VO-2 downlink to switch from the auxiliary oscillator frequency source to the voltage-controlled oscillator source, as the VO-2 radio was "fooled" into going two-way momentarily and then dropping back to the normal one-way mode. These transitions in downlink frequency then caused the DSS-43 downlink receivers to drop lock momentarily. Later in December, this problem also occurred over DSS 14 and DSS 63 but seemed associated only with the use of the high-power transmitters. Once, during a DSS 43 ranging pass, dropping the transmitter power from 50 to 25 kW stopped the effect.

On Jan. 18, 1977, when VO-1 was one-way, and DSS 43 was ranging with VO-2, the VO-1 downlink glitched a total of eight times over about a half-hour interval. The ranging was a special test for the benefit of the Voyager spacecraft, and the interference stopped when DSS 43 handed over to the 26-m DSS 42 for another portion of the test All VO-1 radio subsystem telemetry parameters were normal.

Figure 3 shows that both S-band downlink receivers tracking VO-1 at DSS 63 suffered short out-of-locks at the same time, as denoted by the vertical glitches occurring during the relatively stable downlink signal level plot. These downlink receiver out-of-locks, in turn, caused equipment downstream of the receivers also to go out of lock, resulting in disturbances in the signal-to-noise ratio of both the 33-1/3 bps low-rate and the 16 kbps high-rate data. In one case, an outage of low-rate engineering data caused a bit error on the telemetered uplink signal level, shown as the vertical spike in the top plot.

B. Loss of Orbiter Roll-Reference

When the Orbiter attitude-control system roll-reference is lost, the spacecraft enters a safing routine which, among other things, switches the telecommunications links from the high-gain antenna to the low-gain antenna. The tracking station, of course, observes the unplanned decrease in downlink received signal level which is about 20 dB in magnitude. In addition, if the Orbiter is in a dual-subcarrier mode, the high-rate and low-rate telemetry channels most likely will be below their thresholds when the switch to the low-gain antenna occurs.

As part of the process of restoring the Orbiter to normal operation, it is very important to determine the amount that it has rolled from the reference star. During some roll-reference losses in December 1976 and January 1977, the X-band telecommunication link was used for this purpose. The beamwidth of the Orbiter's high-gain antenna is such that the X-band level is down 10 dB at a distance of 1.3 deg from the boresight. Therefore, the X-band level received at a tracking station can be a valuable indicator of the pointing error.

C. VO-1 Uplink Signal Level Anomaly

Shortly after the August 20, 1975 launch of Viking Orbiter 1, the telemetered uplink received signal level was as much as 3.5 dB lower than predicted by the Project. The difference between the observed value and the predicted value is called a residual. The uplink signal level anomaly has continued since launch, varying in size from time to time. Figure 4 is an overview of the variation in the residual. The discrepancy between observed and predicted values decreased within about a month after launch as the Orbiter low-gain antenna operated at smaller and smaller Earth cone angles. Thus, it seems that a part of the discrepancy was due to low-gain antenna pattern uncertainties at large angles from boresight. When the switch to Orbiter high-gain antenna occurred about 100 days after launch, there was no noticeable change in the residual, suggesting that our knowledge of the gain of both the low-gain and the high-gain antenna was good and that any hardware cause of the discrepancy did not lie in the antennas or their separate cabling.

To isolate the cause of the VO-1 uplink variations, a number of special tests involving the Deep Space Network were scheduled. These include (a) a Radio Frequency Subsystem (RFS) Threshold Test, (b) several Command Detector Unit (CDU) Signal-to-Noise Ratio Estimator (SNORE) tests, and (c) a series of routine CDU SNORE "checks." These will be discussed subsequently.

1. **VO-1 RFS threshold test.** The test was accomplished on Aug. 3, 1977, over DSS 14 for the downlink and DSS 12 for the uplink. The test was preceded by a Project and Deep Space Network coordination meeting, at which the sequence of events for the test was made final. It was later sent to the stations involved for their information. The test was planned such that the received carrier power at the Orbiter would be decreased in 1-dB steps, by use of increasing amounts of ranging modulation carrier suppression from the DSS 12 Planetary Ranging Assembly. Three separate runs were included in the overall sequence. The one successful run of this test proved that there has been little, if any, shift in the radio's threshold during flight; neither has there been a change in the threshold of the command detector unit. However, the threshold test is not a sensitive indicator of normal operating range measurements, which are at levels of 20 to 50 dB higher than "threshold" for Viking. The difficulties in executing this test, given the planning and coordination that went before it, make it difficult to see how "valid" RFS threshold tests can be made in-flight without using special Deep Space Station configurations which make correlation with the normal station configurations difficult.

2. **VO-1 CDU SNORE tests.** The first formal command detector unit signal-to-noise ratio estimator (CDU SNORE) tests were approved for VO-1 for September 15 and for VO-2 for September 16, both over DSS-61. After these tests were completed, it did not seem that the settability or accuracy of carrier suppression with the station's auxiliary modulation source could be made good enough for the very precise results needed.

The CDU SNORE results for both Orbiters suggest residuals in the −2 dB range, compared with predictions made on the basis of prelaunch tests.

3. **CDU SNORE "checks."** The checks are accomplished by the uplink station leaving the command system in the IDLE 2 state for one hour following normal Orbiter command loads, with the transmitter power reduced to a predetermined level of less than 1 kW. The checks were instituted about Aug. 1, 1977, because it was desired to have a continuing check on the performance of the Orbiter command systems. Throughout the period the CDU SNORE checks were conducted routinely, a considerable amount of scatter appeared from test to test for no known cause. Observation of the uplink signal level change from 20 kW down to the power level for the test suggests that the stations cannot set a desired 500 or 100 W level accurately, at least without a calibration.

D. Attitude Control Gas Leaks

Both VO-1 and VO-2 have had "gas leaks" involving the dry nitrogen attitude-control gas. Once stopped, the leaks were further controlled by switching to the alternate half-gas system on each Orbiter, involving a different set of valves. More recently, VO-2 had a major gas leak, at rates of up to 6.7 kg (1.5 lb) of gas per day. This leak was eventually controlled but no redundant half-gas systems remained. Careful management of the amount of gas intentionally used for normal operation and science experiments is necessary during the extended and continuation missions.

The Viking Project declared a spacecraft emergency on November 1, 1977 in order to obtain use of DSS 62. A gas jet leak had been detected on VO-2, in the yaw axis, and the station was needed in order for the Project to watch the Orbiter's behavior closely and to send commands to characterize the leak and possibly to eliminate it. The leak rate was low, but variable. Commands were transmitted to perform 0.72-deg yaw turns to clear the leak in the valve. With this satisfactorily accomplished, the reaction control assembly No. 2 was turned OFF.

A small gas leak occurred in VO-1, roll axis, on Nov. 18. Voyager Project was willing to release DSS 62 without Viking Project's having to declare a spacecraft emergency, and the station was quickly made available to Viking. The gas leak had cleared itself by the time Viking data came on line from DSS 62. This gas leak was believed to be caused by a particle that cleared itself. By far the most serious gas leak of the extended mission occurred on VO-2 beginning February 8, 1978. The leak continued for more than 2 days after that, with a total loss of 1.18 kg (2.6 lb) of gas. Based on previous experience with the VO-1 solar occultations and gas leaks, it is believed by the Orbiter guidance and control analysts that the large VO-2 leak was probably induced by the temperature variations and that the relative impurity of the helium supply compared to the nitrogen may have contributed. Several VO-2 nonpropulsive maneuvers were canceled after the gas leak to allow the Orbiter to get through the solar occultation season without additional strain. The solar occultations ended late in February 1978.

E. VO-2 Central Computer and Sequencer Timing Problem

Each Orbiter computer contains two processors. Certain tandem events require that the relative timing between the two processors not exceed a preset value, normally 160 ms. One such tandem event is the "flip," in which the master/slave relationship between the processors changes. Flips are used when the Orbiter sequence contains so many events that they cannot all be executed by a single processor.

On Sept. 19, 1977, at the time scheduled for a flip from Processor B to Processor A on VO-2, the ongoing sequence

abruptly halted and the spacecraft was put into a safe state, transmitting 8-1/3 bps single-subcarrier via the low-gain antenna. The immediate indication of trouble was lack of downlink data from the tracking station DSS 44. This 26-m station could barely hold receiver lock at an indicated downlink carrier level of −167 dBm, and the 8-1/3 bps signal-to-noise ratio was about 1 dB. Viking Project declared a spacecraft emergency, in order to obtain use of DSS 43 in Canberra and then DSS 63 in Madrid. DSS 43, in fact, was still in final Deep Space Network system tests after conversion to the Mark III Data System, and DSS 63 was allocated to Helios.

DSS 43 was brought on-line in a little more than one hour after the spacecraft emergency declaration. After Orbiter analysts assessed the telemetry data, they transmitted commands from DSS 44 which still had the VO-2 uplink. These commands restored downlink transmission on the high-gain antenna, and put the Orbiter into the 33-1/3 bps data mode. The Orbiter remained in this condition, with the remainder of that week's onboard science sequence canceled while testing of the central computer and sequencer subsystem continued. VO-2 returned to its science mission on Sept. 22. The spacecraft emergency was lifted on Sept. 20.

IV. Ground System Activities

A. Upgrade of Deep Space Stations for Post-Viking Era

During the Viking Extended Mission, the Deep Space Stations were taken out of service for approximately three months each to install Mark III Data System hardware and software. The conversion was accomplished first at Goldstone (DSS 12 and DSS 14), then at Canberra (DSS 42 and DSS 43), and then at Madrid (DSS 61 and DSS 63). The remaining 26-m stations were also included, but Viking operations were primarily affected by loss of the 64-m stations, over which the majority of the science data playback occurred.

1. **DSS 14.** This Goldstone 64-m station tracked Viking informally on June 23, 1977, as a checkout of Mark III equipment. Demonstration tests took place in July, including reception and processing of 8-kbps and 16-kbps high-rate block-coded data. The Mark III Data System 64-m stations have capability for only four telemetry streams simultaneously, instead of the six that Viking was used to previously. As a result, during one of the demonstration tests when both Orbiters were in the dual-subcarrier mode, and there was a simultaneous Lander direct link, it was necessary to have DSS 14 track VL-1 and VO-1 and to have the simultaneously available DSS 63 track VO-2. DSS 14 was also used for the VO-1 RFS Threshold Test during its checkout period.

2. **DSS 43.** The last Viking pass by either of the conjoint Madrid stations, DSS 61 and DSS 63, took place during the VO-2 Deimos near-flyby on Oct. 15. The stations were then taken down for three months. About this same time the conjoint Canberra stations, DSS 42 and DSS 43, completed their requalification tests with no problems, and were again made available for Viking. Competition for the available stations became severe in July and August, considering operational readiness tests for the upcoming Voyager launches as well as routine Helios and Pioneer tracking. By this time, both VO-1 and VO-2 could play back data at 16 kbps over a 64-m station. The high-rate capability was limited for transmission of this data from the overseas stations to the space flight operations facility because of the 27.8 kbps capacity of the wideband data lines. Operationally, the 16 kbps from one of the Orbiters would be transmitted over the wideband line in real time, while the 16 kbps from the other would be recorded on a digital original data record, and then replayed from the station post-pass.

The first use by Viking of DSS 43 was under the spacecraft emergency requirement on Sept. 19, when the VO-2 central computer and sequencer developed its timing problem. The station performed admirably under these circumstances, obtaining for Viking telemetry data when it was really needed. The more formal demonstration passes, ground data system tests, began the second week in October. The Project telecommunications analyst noted several times during these tests that the ranging carrier suppression was set to an erroneous value. The station was checking out a new ranging module. The stations were scheduled to be available formally to Viking on Oct. 15, the same date that the Madrid stations would be taken down.

3. **DSS 63.** The first systems integration test for DSS 61 and DSS 63 was scheduled for Jan. 5, 1978. However, as in the case of the first use of DSS 43 in September, the first Viking track by DSS 63 was on a semi-emergency basis on Jan. 3. The situation arose because earlier on that same date, DSS 43 had not been able to bring their antenna to point at the scheduled start of track. The playback of Lander relay data from VO-2 was lost as a result. DSS 63 was made available, and a command load transmitted to VO-2 to play back the missed data again during the DSS 63 pass. DSS 63 not only received the 8-kbps data with a good signal-to-noise ratio, but they also transmitted the 25 commands that were necessary to position the Orbiter recorder and to initiate the playback.

B. Improvement in Analog Original Data Record System

By mid-November 1977 the 64-m stations had obtained and installed the FR 2000 analog recorders intended as replacements to the old FR 1400's. The first demonstration pass for

the new equipment was at DSS 14 on Nov. 18. A similar demonstration pass took place at DSS 43 on Nov. 25. The tests were successful, and the stations were then authorized to use the FR 2000's for all future Viking analog recording and playback. Each station has a single FR 2000. Therefore, for long periods of analog recording, there yet remain unavoidable gaps in the data; these occur during the times the operators are changing the tapes. For particularly critical sequences, the times that the tapes are changed may be coordinated by the Viking real-time Flight Control Chief.

C. Uplink Transmitter Power

For the superior conjunction period, November 1976 through January 1977, the 64-m stations overseas used 50-kW uplink power and DSS-14 at Goldstone used 100 kW. The higher powers were necessary for the radio science relativity experiment, considering the increased noise from the Sun and the maximum Earth-Mars distance at the time. After superior conjunction, the 20-kW power level was again used throughout 1977. On Feb. 8, 1978, the Viking Mission Director authorized the Deep Space Network to use 10 kW for all Viking uplinks for the remainder of the extended mission.

D. Planning for Orbiter Playback Bit Rates

Throughout the Viking primary mission, the playback bit rate profile for each Orbiter was generated manually by Project telecommunications people.

During the last three months of 1976, Orbiter telecommunications people were developing a program, the "Telecommunications Playback Opportunities Profile," (TPBOP) which would specify the bit rate change times automatically. The program was in the demonstration phase during January 1977, and put on-line for production in February 1977. The Viking Extended Mission sequences were, in fact, more complex than those of the primary mission in terms of the data playback. This was because there were fewer Deep Space Stations allocated, meaning (a) the 64-m station coverage was no longer continuous, so many more bit rate changes occurred each day; and (b) uplinks were no longer continuous, so there were more one-way to two-way transitions to contend with. As a result of the reduced DSN allocations, those allocations remaining were used to the maximum, with the result for less margin for error in planning. Earth occultations also occurred during the extended mission on both Orbiters, and these further complicated the tape management process. However, with the direct computer file interface between the TPBOP software and the tape management software, sequence planning was generally smooth throughout the extended mission. Most errors occurred as the result of station allocation changes made after the original tape management had already been completed.

E. Use of 26-m Stations for Orbiter High-Rate Data Return

As competition from other Projects for the 64-m stations increased in mid-1977, the Viking Project became interested in the predicted capability of the 26-m stations to receive 2-kbps high-rate data. The Project telecommunications assessment was that the communications range would become small enough by August 1977 for the 26-m stations to support 2-kbps visual imaging data. The Telecommunications Playback Opportunities Profile was upgraded to include data transition times for 26-m stations. The first such playback was to be Aug. 16, 1977.

There are station limitations which preclude the reception of more than one simultaneous 2-kbps data stream at some of the stations. DSS 11, for example, had a single symbol synchronizer because the second one was being used for support of Voyager launch activities at Cape Canaveral. In addition, the 26-m stations could support only two data streams (high rate or low rate) simultaneously. In this case, if an Orbiter were in the dual-subcarrier mode (2 kbps and 8-1/3 bps, for example) and the second Orbiter were in the single-subcarrier mode (33-1/3 bps, for example), the on-duty Flight Control Chief would have to decide which of the three streams to sacrifice. Eventually, all data streams were given priorities so that such decisions did not have to be made under pressure in most cases. Selection of the data stream to be dropped would be made on the basis of which Orbiter was busier at the time; the busier Orbiter's low-rate engineering stream would be returned together with the 2-kbps high-rate playback stream.

V. Ground System Problems

A. Interference to Orbiter Low-Rate Channel by High-Rate Data Content

The Orbiter high-rate subcarrier frequency is 240 kHz; the low-rate subcarrier frequency is 24 kHz. It was noted during prelaunch testing and during interplanetary cruise that certain high-rate data patterns would cause a decrease in the low-rate channel signal-to-noise ratio. The cause of the problem is sidebands of the data spectrum around the 240-kHz subcarrier falling on top of the 24-kHz subcarrier. This causes difficulty with the station's subcarrier demodulator assembly, and was verified by tests at Compatibility Test Area 21 (CTA 21). These tests showed that 16-kbps uncoded data (present during the Viking Lander Capsule checkout in interplanetary cruise) could create enough energy near 24 kHz to knock the subcarrier demodulator assembly out of lock. Some data patterns, such as "all zeroes," of 8 kbps or 16 kbps block-coded data could cause lesser degradation.

Figure 5 shows that the 8-kbps and 16-kbps block-coded data did cause such degradation during the Viking Extended Mission. With downlink carrier signal levels in the −140 dBm range, certain data patterns caused degradation of between 2 and 6 dB in the 33-1/3 bps low-rate telemetry signal-to-noise. Considering that the average signal-to-noise ratio of 16 to 17 dB was well above the threshold signal-to-noise ratio of 5.2 dB, these degradations did not cause any difficulty in reception of the data.

The maximum interference occurred at the maximum signal level. In several specific cases, such as in Fig. 5, the interference was associated with specific events such as Orbiter tape recorder reversals, or periods of "all data zeroes" during the time the tape recorder was not running but the dual-subcarrier mode was on. The condition of "all data zeroes" refers to the input to the spacecraft block coder, and not to the symbol stream from the block coder to the high-rate modulator.

B. High Doppler Rate Near VO-2 Orbit Periapsis

Late in November 1977, several passes occurred in which 26-m stations could not maintain lock on the (two-way) downlink carrier when the Orbiter was on the low-gain antenna. This condition would persist for about one hour total, the time centered on the periapsis. The downlink signal level was about −157 dBm, predicted, at this time, and the Orbiter was in the single-subcarrier mode. The highest two-way doppler rate on Nov. 21, 1977 was 19 Hz/s. That day, DSS 11 downlink receiver went out of lock at 08:24 GMT, and did not regain lock until 08:41 GMT, with periapsis at 08:34 GMT.

C. Missed Uplink Acquisitions After VO-1 Earth Occultation

The most difficult uplink acquisitions of the extended mission were those to VO-1 just after exit from Earth occultation. Prior to enter occultation, the outgoing station was, by procedure, to have the uplink tuned so that the Orbiter radio would be at best-lock frequency. The Orbiter would enter Earth occultation two-way, and would exit one-way. Orbital doppler rates were high near-Earth occultations, which occurred near periapsis on VO-1.

Difficult uplink acquisitions are not limited to any one station or site, nor to any one of the Viking vehicles. Figure 6 shows the variation of the Orbiter receiver static phase error during two different types of acquisitions. In the normal track, the track synthesizer frequency (TSF) is set so that the static phase error goes equally in both directions from a telemetry data number of 64 (receiver best lock) during the station pass. Part (a) of the Fig..6 shows a one-way to two-way acquisition by DSS 11, when the orbit periapsis occurred well after start of track. Part (b) shows a DSS 61 to DSS 14 handover, in which the handover was required to occur near-orbit periapsis. Both these acquisitions happened to be with VO-2.

D. High-Rate Data Losses

The only time during the extended mission playback data losses became a concern to the Project was early in September 1977. At that time, a significant fraction of the total infrared science played back from the Orbiters was lost because of station problems.

Most of the data losses had occurred over DSS 63, which happened to be the only 64-m station tracking Viking at the time. Most of the data outages were of 4-kbps playback data near the start of the station pass. The root cause of the outages was the station crew had too many things to do all at the same time. The conjoint stations DSS 61 and DSS 63 were operated with a single four-person crew normally, and the crew was shorthanded at the time. The station operators were required to establish an up-link to an Orbiter and to get two low-rate and two high-rate data streams started within about 15 min of the start of track.

Participants at the meeting agreed that the most immediate relief could be obtained by having the high-rate playbacks at the beginning of the pass start at different times rather than simultaneously for the two Orbiters. In the Playback Opportunities Profile, the VO-1 playback would continue to be allocated to begin 10 min after the station start of track. The VO-2 playback would not be allowed to start until 20 min after station start of track. This expedient worked and no further losses were experienced.

There were no significant problems associated with imaging playbacks during the extended mission, other than when hardware failures prevented the reception of the data.

E. Degradation to Orbiter 16-kbps Signal-to-Noise Ratio on Back-up S-band Maser

During the Viking primary mission, a "pad" of 1.2 dB was added to the required threshold for science data playback to account for the more likely degradation factors. Among these factors was the occasional use of a 64-m station's backup S-band maser when the primary maser would fail. The backup maser has a slightly higher system noise temperature and, on average, one can expect a 1 to 1.5 dB degradation in signal-to-noise ratio when the backup maser is used.

During the week of May 15, DSS 43 operated several days on its backup maser. During these passes, the system noise temperature at S-band was about 34 K instead of the 22 deg that is expected for the primary maser. Even with the Orbiter ranging channel OFF, the VO-1 16-kbps signal-to-noise ratio

averaged 0.5 dB and the VO-2 16-kbps signal-to-noise ratio averaged 1.0 dB. The station's performance was degraded by between 2 dB and 3 dB, compared with the expected 1 to 1.5 dB. Prior to Viking launch, during the systems integration tests, the DSS 43 backup maser system noise temperature had been 27 K, and it had gone to 34 K some time after those tests. The primary maser was back in operation on May 22, and the 16-kbps performance was normal after that. (The threshold for 16-kbps visual imaging data, without the 1.2 dB pad is 1.8 dB; therefore, the observed values while on the backup maser were well below the required threshold.)

F. "Catastrophic" Increases in Station System Operating Temperature

On Jan. 7, 1977, it snowed heavily at DSS 14 and DSS 11 at Goldstone DSCC. The DSS 14 system noise temperature went to 450 K, compared to a normal of about 20 K at S-band. The resulting 14-dB degradation caused the loss of high-rate playback from both VO-1 and VO-2. Nothing could be done about the loss, except to wait for the next 64-m station, DSS 43, to rise. DSS 11 system noise temperature increased to 125 K, from a normal of about 33 deg K, due to the same snowstorm.

Another weather-related type of data loss is excessive wind which requires stowage of the ground antenna. During one Viking Extended Mission pass, DSS 63 was required to stow the antenna for three hours, causing the loss of 16-kbps data from both Orbiters.

G. Radio Frequency Interference from External Sources

Radio frequency interference was predicted during the extended mission from the Earth-orbiting GEOS satellite and (at Goldstone DSCC) from military exercises in the area. A large amount of Viking and Deep Space Network analysis had gone into potential interference to Viking by GEOS, which has a transmitter in the deep space downlink band; the GEOS frequency is 2299.5 MHz, compared to 2293.2 MHz for VO-1 and 2297.2 MHz for VO-2. GEOS was planned to be launched into geosynchronous orbit, very much closer to the Deep Space Stations than is Mars; its omnidirectional antenna has a beamwidth that covers the Earth. Thus, anytime GEOS appeared in the same angular direction from a Deep Space Station as Mars, interference could be expected. The Deep Space Network generated a computer program to calculate the times and severity of the interference.

This interference was predicted starting the last week in May 1977 and continuing through the end of June. The observed GEOS signal level at DSS 43 was a carrier at about −130 dBm. The first interference was noted during VO-2 playback of 16-kbps imaging data on May 29 and 30. The signature of the problem was that the 16-kbps data would go out of sync in the station's subcarrier demodulator assembly. The telemetry and command processor channel would then be reinitialized by the station operator after the subcarrier demodulator assembly lock was regained, several minutes later, and the signal-to-noise ratio would then be back at predicted levels. The degradation was not smooth in time, as a result, but rather was discontinuous due to subcarrier demodulator assembly drop-lock.

The Deep Space Network analysis suggested:

(1) Severe telemetry bit signal-to-noise ratio and bit error rate degradation when the carrier interfering signal is near or on the subcarrier's first and third harmonic spectrum of the desired signal;

(2) No noticeable degradation when the interfering carrier is near the even harmonics of the subcarrier of the desired signal;

(3) Severe tracking doppler phase jitter degradation effects, as well as telemetry bit signal-to-noise ratio and bit error rate degradation, when the interfering carrier is within 1 kHz of the desired downlink carrier signal.

In comparison with the −130 dBm level of the GEOS signal late in May 1977, the Viking Orbiter signal level was −141 dBm for single-subcarrier and −146 dBm for dual-subcarrier.

H. Scheduling Problems

Viking Project scheduling is carried out in coordination with the Deep Space Network scheduling office. About mid-1977, several instances of lost data were traced to last-minute changes in station allocation which had not been made known to the Orbiter people involved in management of Orbiter tape recorder playback. Typically, the loss would be of a few minutes' duration and would be caused by the change in time for the uplink station allocation.

On Sept. 13, 1977, one minute of 8-kbps data from the playback of a Lander-to-Orbiter relay pass was lost. The Lander analysts were bothered by this seemingly short outage because that minute of data contained engineering data spanning 24 h for some instruments. Analysis of the problem indicated that it started with the loss of a DSS 14 allocation for Voyager, and the reassignment of DSS 63 to Viking. In addition, the uplink allocation of DSS 61 was moved from the originally scheduled time. So, the playback of the relay data was moved into the DSS 63 view period and then was disrupted by the DSS 61 transmitter coming on at a time different than planned.

In October 1977 the stations continued to be squeezed in terms of time required for precalibration, compared with the start of track time scheduled. The Viking Project flight plan summaries began to show priority of data, so that the real-time Flight Control Chief could determine which things could be neglected during a particular precalibration, if required.

In November 1977 numerous late changes in scheduled tracking time occurred. Some data outages occurred because tape management had placed playback where no station allocation still existed. To alleviate this problem, Viking Project scheduling produced weekly handwritten memos indicating changed allocations. Also, DSN scheduling began to put revisions to the 7-day schedule on the Project teletype machine for use in realtime. The Viking Radio-Science Team also published periodic memos detailing the timing for their upcoming experiments, including the VO-1 and VO-2 Earth occultations, the gravity field coverage, the near-simultaneous Lander/Orbiter ranging, and bistatic radar. The scheduling difficulties decreased after these steps were taken.

In mid-October 1977 there had been several DSS 11 missed uplink acquisitions following Earth occultation of VO-1. It was found that the times in the sequence of events were about 1-1/2 minutes later than the actual times deduced from DSS 11 receiver out-of-lock (enter occultation) and in-lock (exit occultation). To bring times used by all elements of the Project and the Deep Space Network into agreement, the Viking Radio Science Team took the approach of manually tweaking the times calculated by navigation, for inclusion in the sequence of events.

VI. Viking Radio Science Activities

The Viking Radio Science Team performed a large number of experiments during the Viking primary and extended missions. Representative activities are described in the following paragraph; however, not all of the experiments are discussed.

A. Superior Conjunction Relativity Experiment

Reference 1 describes the planning for and the observations made during the Viking superior conjunction in October to December 1977. Only the Orbiter activities following the day of closest approach to the Sun, Nov. 25, are mentioned in this section.

The first Orbiter low-rate telemetry received after this day was on Nov. 26, during the DSS 14 pass. On Dec. 6 the Sun-Earth-probe angle had increased to 3 deg from its minimum of 0.25 deg on November 25. The first post-conjunction Orbiter commanding was done on Dec. 6, from DSS 14 to VO-1 and from DSS 11 to VO-2. By Dec. 9 VO-1 began transmitting continuously in the single-subcarrier mode at 33-1/3 bps.

On Dec. 13 VO-1 operated in the dual-subcarrier mode, at 2 kbps and 8-1/3 bps, for the first time since the conjunction. The link performance was satisfactory, and the signal-to-noise ratios about the same as the corresponding time before solar conjunction. VO-2 went through a similar "wakeup" sequence the next day. By Dec. 20 the Sun-Earth-probe angle had increased to 7 deg and the S-band and X-band links had essentially returned to their preconjunction condition.

B. Quasar-VLBI Experiment

For this experiment, the tracking stations alternately point their antennas at an Orbiter and at a quasi-stellar source (quasar). The Quasar-Very Long Baseline Interferometry experiment occurred during interplanetary cruise, during the primary mission, and several times during the extended mission. DSS 14 and DSS 42 were involved in such an experiment on Feb. 10, 1977.

Late in 1977 there were several quasar-VLBI experiments conducted using radioastronomy facilities at Haystack, Mass. and at Owens Valley, Calif. The necessary coordination between the experiment and the normal Orbiter mission planning was crucial for these, since outside agencies were involved. On Dec. 14, 1977, while the quasar-VLBI data was being accumulated at these two sites, the X-band down-link from VO-1 suddenly disappeared. The cause for the loss of the signal was that at that time the Orbiter began a planned roll maneuver, causing the high-gain antenna to offpoint from the Earth. Previously there had been observations on Nov. 29 and 30 without incident; later, observations occurred without problem on Dec. 15 and on Dec. 29 and 30.

The quasar-VLBI experiment has been likened to a "star-plate" analogy. The Orbiter X-band signal and X-band radiation from the quasars (which are considered as fixed stars) are used to obtain highly accurate angular measurements between the Orbiter and the quasar, so as to tie together the coordinates of the Viking orbits, Mars, and the fixed stellar reference frame.

C. Bistatic Radar

Late in February 1977, approval was given to the bistatic radar experiment. This experiment would consist of having the Orbiter point its high-gain antenna toward the surface of Mars, thus bouncing the X-band and the S-band downlinks off the surface and back to the Earth where they could be received at 64-m stations, using open-loop receiving equipment. The high-gain antenna would either be pointed toward a single point on the surface during the pass (fixed-point observation) or such

that the angle of incidence equalled the angle of reflection (specular pass). It was not expected that normal closed-loop downlink carrier or telemetry data could be received because of the frequency smearing resulting from the surface reflection. The bistatic radar experiment was a "guest experiment" with the principal investigator from Stanford University.

The experiment objective was to determine the surface roughness and the electrical and other properties of the surface of Mars in selected regions not accessible by Earth-based radar, by means of spectral analyses of the reflected and direct signals from the surface.

A new bistatic radar telemetry format was created and loaded into the Orbiters' flight data subsystems by the end of October. It emphasized the articulation subsystem channels so that a detailed telemetry record of the high-gain pointing directions could be obtained. The Deep Space Network was involved in the coordination of this experiment with the Project and the experimenters. A special instruction message was transmitted to DSS 14 and DSS 43, the two stations that would be receiving bistatic data. The Orbiter sequences included contingency periods of low-gain operation following the bistatic passes, in the event that the high-gain antenna did not return properly to Earth pointing.

Good reflected data was observed throughout the Nov. 8 bistatic pass at DSS 14. The open-loop recordings were flown back to Stanford University almost immediately for processing.

By Nov. 11 the experiment leader had done sufficient processing of the data to make a first estimate of the average slope of the surface and also of the dielectric constant of the surface. During December and January, bistatic passes were accomplished from both Orbiters and at both DSS 14 and DSS 43. Review of the Orbiter S-band high-gain antenna patterns show that there are sidelobes in the -30 dB to -35 dB relative gain out to at least 30 deg at S-band, thereby further substantiating the plausibility of the 64-m closed-loop lock-up on the downlink carrier. Lock-up of the S-band and X-band downlink carriers and the 8-1/2 bits/s telemetry data occurred during several bistatic radar passes.

The bistatic radar experiment required that system noise temperature measurements be made not only during precalibration but also immediately after the completion of the experiment. Such measurements required that tracking of the Orbiters be discontinued for approximately 15 min while the measurements were made. By January 1978, the Radio Science Team and the Deep Space Network had determined that they could use the maser noise diode on the S-band Polarization Diversity maser and the maser gas tube in the MOD III maser to make noise temperature measurements in lieu of the standard post-calibration ambient noise temperature measurements which required breaking track.

D. Near-Simultaneous Lander and Orbiter Ranging and Doppler Data

A number of radio science experiments are enhanced by the unique Viking situation of having one vehicle on the surface of the planet and a second in orbit around the planet. The Orbiter S-band and X-band doppler and ranging data can be used to calibrate the Lander data for charged particle and other perturbing effects. On Oct. 6, 1977 the traveling wave tube amplifier for the direct S-band link from Viking Lander 2 failed. The other VL-2 traveling wave tube had apparently failed a year previously when it did not come on for a planned direct link. No further attempts to operate a direct link using Lander No. 2 were made after Oct. 6, 1977, and all of the near-simultaneous radio science work was subsequently carried out using Lander 1, together with either VO-1 or VO-2.

Some of the near-simultaneous passes were complex to carry out. During one week in March 1977 there were a total of six near-simultaneous passes over DSS 14 and DSS 63. Each of these was a split pass, in which the station uplink was sent first to the Viking Orbiter, then to the Viking Lander, and finally again to the Viking Orbiter, which meant three separate uplink acquisitions during the pass.

During the first nine months of 1977 there were 59 good ranging passes for Viking Lander 1, which totals to 68% of those scheduled for this Lander. There were 25 good ranging passes to VL-2 out of 38 scheduled for a 66% success ratio. A good pass is one in which at least one good ranging point was acquired.

Particular emphasis was placed on the near-simultaneous ranging passes between November 1977 and March 1978, when communication range was a minimum. Strong signals were needed in order to obtain excellent orbits for Mars and Earth for the superior conjunction relativity experiment which will take place in early 1979. In the process of near-simultaneous doppler and ranging, there are approximately 40 tracking parameters to be estimated simultaneously. The signatures of some of these are similar to those of others, and some are of long period. The number of measurements required, and their spacing, led to an estimate on the part of the Radio Science Team that about one Lander ranging pass per week would be needed during the November-March period. This estimate assumed that the ratio of good to total ranging passes would continue to be approximately 2/3.

E. Solar Wind Experiment

This experiment took place during the primary mission on May 16, 1976 and during the extended mission on April 7,

1977. On the latter day DSS 14 and DSS 43 were each requested to accumulate 10/s doppler data during their overlap period. There was no effect on the normal telemetry acquisition from the Orbiters. The purpose of the experiment was to compare the S-band and X-band doppler perturbations caused by the solar wind on the downlinks from VO-2 to DSS 14 and to DSS 43. The data quality was judged by the Radio Science Team to be excellent from both stations.

F. VO-1 and VO-2 Earth Occultations

At various times during the Viking extended and continuation missions, Mars interposes itself between VO-1 or VO-2 and the Earth, thus creating an Earth occultation. The Radio Science Team members use the occultation data to produce pressure, temperature, and density profiles as well as to estimate the radius at the occultation points over a planet-wide range of latitudes and longitudes. The radii data are particularly valuable in the generation of a comprehensive determination of the shape and topography of Mars. As for the pressure, temperature, and atmospheric density profiles, these are being used to detect seasonal variations in Mars' atmosphere.

VO-1 Earth occultations occurred in the first months of the Viking Extended Mission, and by Nov. 19, 1977 both VO-1 and VO-2 Earth occultations were occurring every day. The VO-1 occultations occurred within the interval spanned by the much longer VO-2 Earth occultations. The VO-2 occultations were nearer apoapsis. The tracking stations saw the following sequence: (a) VO-2 enter occultation, (b) VO-1 enter occultation, (c) VO-1 exit, and (d) VO-2 exit. The station activity profile became quite complicated at times since the Radio Science Team wanted open-loop recordings of the signal during the enter and exit occultation times for both Orbiters. The bistatic radar experiments were also occurring during the same station passes at DSS 14 and DSS 43, and bistatic radar also required use of the open-loop recording equipment. In general, a station would not be required to use this equipment for both Earth occultation and bistatic radar on a given pass. The stations were also provided schedules for timing of the open-loop recorders, so as to pick up the crucial entrance and exit times for each Orbiter. Several Project-Deep Space Network coordination meetings occurred to resolve questions of predicted Earth occultation times, station transmitter tuning strategies, and procedures. VO-1 Earth occultations will continue for the remainder of the mission. VO-2 will again be occulted in April and May 1978 and from October 1978 through end of mission in February 1979.

Late in December 1977, during several of the VO-2 Earth occultations, the decrease in uplink and downlink signal levels at enter-occultation was slow enough that the changes could easily be seen in the spacecraft telemetry and the ground system monitor data. Figure 7 is a plot of the uplink signal strength, the S-band downlink signal strength, and the X-band downlink signal strength, all on a common time scale. Figure 7 is for the VO-2 Earth occultation on Dec. 27, 1977, as observed at DSS 14. All the signal level changes had occurred in a little more than two minutes. All three link quantities varied in the same manner and at about the same rate. In a time of 1-1/2 minutes, all quantities had decreased about 3 dB. Each then recovered 1 dB in about 10 s, then decreased rapidly after that.

VII. Deep Space Network-Viking Orbiter Communication Link Performance

A. Predicted Signal Levels

Predicted signal levels are expressed in decibels below 1 mW (dBm) for the uplink carrier level and the downlink signal level. The signal-to-noise ratios are expressed in decibels. Figures 8 through 11 show the predicted signal levels for the most common uplink and downlink modes on the Viking Extended Mission. The conversion factors used by the Project to define the signal levels for other modes are also provided. Each of Figs. 8 through 11 is a plot as a function of time.

Figure 8 shows predicted uplink signal levels. The three curves shown are predicted Orbiter receiver carrier level in dBm. They are for: (a) 64-m station, transmitting to the Orbiter high-gain antenna; (b) 26-m station, transmitting to the Orbiter high-gain antenna; and (c) 64-m station, transmitting to the Orbiter low-gain antenna. As would be expected, the shape of the two upper curves is identical, since the only variable is station antenna diameter, and the uplink signal level is a function only of the Earth-spacecraft distance. The bottom curve has a different shape because the uplink signal level is a function not only of the Earth-spacecraft distance but also of the angle from the low-gain antenna boresight (cone angle).

Figure 9 shows predicted downlink signal levels. The four curves shown are for: (a) Orbiter high-gain antenna, S-band downlink, dual-subcarrier mode, transmitting to a 64-m station; (b) Orbiter high-gain antenna, S-band downlink, single-subcarrier mode, transmitting to a 26-m station; (c) Orbiter high-gain antenna, X-band downlink, transmitting to a 64-m station; and (d) Orbiter low-gain antenna, S-band downlink, single-subcarrier, transmitting to a 64-m station. The similar shape of the three upper curves is due to the fact that the high-gain antenna is used, and only the Earth-spacecraft range influences the signal level as a function of time. The bottom curve, for low-gain, is dependent on the Earth cone angle as well as the Earth-spacecraft range.

All the curves in Fig. 9 assume that the Orbiter ranging channel is on. If the channel is off, the S-band carrier levels increase by 0.9 dB, and the X-band carrier level increases by 1.4 dB. The Project-assumed difference between 64-m and 26-m S-band downlink carrier levels is 8.5 dB. The difference between the carrier level for the single-subcarrier mode and the dual-subcarrier mode is 5.2 dB, the dual-subcarrier mode producing the weaker carrier.

Figure 10 shows predicted signal-to-noise ratios. Additional signal-to-noise ratios are shown in Fig. 11. In Fig. 10 the top curve is for the 8-1/3 bps single-subcarrier mode, on the Orbiter low-gain antenna, transmitting to a 64-m station. The bottom three curves are all for the high-rate channel in the dual-subcarrier mode, on the Orbiter high-gain antenna. From top to bottom, they are: (a) 8-kbps data rate, block-coded, transmitting to a 64-m station; (b) 16-kbps data rate, block-coded, transmitting to a 64-m station; and (c) 2-kbps data rate, block-coded, transmitting to a 26-m station. All of the curves in Figs. 14 and 15 assume that the Orbiter ranging channel is on. If it is off, the signal-to-noise ratios increase by 0.9 dB.

Figure 11 shows additional predicted signal-to-noise ratios. The upper curve is for the 33-1/3 bps data rate, single-subcarrier mode, Orbiter high-gain antenna, transmitting to a 26-m station. This is the so-called cruise mode, used extensively when playback of science data was not required. The bottom curve is the predicted signal-to-noise ratio for the 33-1/3 bps low-rate channel in the dual-subcarrier mode, Orbiter high-gain antenna, transmitting to a 64-m station. Both these curves assume the Orbiter ranging channel is on; the signal-to-noise ratios increase by 0.9 dB if the ranging channel is off.

All signal-to-noise ratio curves assume a station elevation angle of 25 deg, the standard value used in general Viking mission planning. Very approximately, the signal-to-noise ratio would be almost 1 dB higher near zenith than at 25 deg; it would be nearly 2 dB lower near the 6-deg horizon than at 25 deg. In detailed Viking mission planning, as well as link residual analysis, the elevation angle effects were taken into account. Figures 10 and 11 also show the threshold values and saturation values for the signal-to-noise ratios. Block-coded high-rate data had a threshold of +1.8 dB for the visual imaging playback data, +4.2 dB for infrared data playback, and +4.7 dB for Lander-to-Orbiter relay data playback. For link planning, a pad of 1.2 dB would be added to each of these numbers. Figures 10 and 11 show a threshold of 3.0 dB, which is the "padded threshold" for the imaging data.

B. Link Performance Trends During Viking Extended Mission

The Viking telecommunication analysts assessed performance trends by means of link residuals. In this way, the apparent changes caused by changing communication distances, antenna selection, data mode or rate, could be normalized out. A link residual is defined as the difference, in dB, between the observed value of the parameter and the predicted value. A positive residual indicates that the observed signal is stronger than the predicted one. The predicted values come from the Telecommunications Prediction and Analysis Program (TPAP).

The link quantities assessed included: (a) uplink signal level, (b) S-band downlink signal level, (c) low-rate signal-to-noise ratio, both for single-subcarrier and for dual subcarrier modes, (d) high-rate signal-to-noise ratio, and (e) X-band downlink signal level.

Figures 12 through 16 are summary plots of several of the primary data modes for the Viking Extended Mission. Each figure shows the VO-1 quantities at the top and the VO-2 quantities at the bottom. The plots use the mean from the monthly residual statistics, so there is one point plotted per month. Each plot covers the time span from August or September 1976 until January 1978. Data for October, November, and December 1976 are omitted, because of the severe degradation caused by superior conjunction effects (Reference 1).

Figure 12 shows VO-1 and VO-2 uplink signal level residuals. The circled points are for uplinks from 26-m stations, and the points indicated by triangles are for 64-m uplinks. The plot shows that a change occurred in the telemetered uplink signal level for VO-1 approximately January 1977 (or possibly during superior conjunction). Since that time, the month-to-month variations are believed to be due to differing mixes of stations used for the uplinks. Over the extended mission the VO-1 uplink signal level residual, for a given station, is about 5 to 6 dB lower (more negative) than the VO-2 residual. No changes have occurred in uplink performance for either Orbiter since January 1977.

Figure 13 shows S-band downlink signal level residuals. From month to month VO-1 and VO-2 behave similarly with regard to S-band downlink levels. Although there are sometimes substantial changes from month to month, these changes track between VO-1 and VO-2. Deep Space Station calibrations may be a factor. The 64-m downlinks are indicated by circled points, and the 26-m downlinks by points with triangles. In general, both VO-1 and VO-2 consistently show

downlink signal level residuals in the −0.5-dB range; the residuals have remained constant over the extended mission, on average.

Figure 14 shows low-rate channel signal-to-noise ratio residuals. These residuals remained constant for both Orbiters during the first half of 1977, then began a gradual decline through the second half of that year and into January 1978. January 1978 was the time of minimum Earth-Mars distance. Saturation and interference effects are the causes of the decline. By January 1978, the predicted value for single-subcarrier signal-to-noise ratio was greater than 24 dB at the 26-m stations. Also, the predicted dual-subcarrier signal-to-noise ratio was greater than 18 dB at the 64-m stations. The single-subcarrier level is higher than the typical saturated signal-to-noise ratio at the stations. The dual-subcarrier signal-to-noise ratio is visibly disturbed by interference from the 8-kbps and 16-kbps block-coded data content. During the first half of 1977, the average performance of VO-1 low-rate signal-to-noise ratio was 0.5 dB higher than that of VO-2. This difference is caused by approximately 0.5 dB higher signal output from VO-2 than VO-1, and the result is seen in the high-rate signal-to-noise ratio also. (During the Viking primary and extended missions, data rate planning took into account the 0.5 dB difference between Orbiters.) By January 1978, the residuals for low-rate signal-to-noise ratio averaged more than −1.5 dB for both VO-1 and VO-2. Of course, with absolute levels in the high teens and low twenties of dB, there was absolutely no trouble with data quality. The downward trend in Fig. 18 began to reverse itself in February and March, 1978.

Figure 15 shows high-rate channel signal-to-noise ratio residuals. The high-rate channel performance was of the most interest of any link quantity to the Viking Project, because all of the science return from the platform instruments and the relay link was via this link. The high-rate channel performance remained the most constant of any of the links also, which meant that link planning based on this stability was accurate. The 8-kbps data mode was present throughout the extended mission. The 16-kbps mode did not begin until March 1977, when the predicted signal-to-noise ratio came above the 3.0-dB padded threshold required for imaging data. (In fact, for the first several months, the imaging data was returned with a pad of only 0.7 dB instead of the normal 1.2 dB, meaning that the predicted minimum signal-to-noise ratio was 2.5 dB.) The 0.5-dB difference between VO-1 and VO-2 is clear in Fig. 15. The typical 8-kbps signal-to-noise ratio residual for VO-1 was −0.25 dB, and that for VO-2 was +0.25 dB. For each Orbiter the 16-kbps signal-to-noise ratio was about 0.25 dB more negative than the 8-kbps residual. Beginning in October 1977, the 26-m stations were able to receive 2-kbps dual-subcarrier data, and these residuals are indicated on Fig. 15 by the squares. The month-to-month scatter in the 2-kbps residuals is larger than that for the 64-m high-rate reception. This is believed to be due to the much smaller number of data samples for the 26-m high-rate.

Figure 16 shows VO-1 and VO-2 X-band downlink signal levels. The X-band signal levels showed an extreme amount of variability from station-to-station and from pass-to-pass. X-band signal level calibration at the stations may have been the cause. Figure 16 shows that generally, from the one month to the next, trends in VO-2 were mirrored in VO-1. The VO-1 residuals averaged about 1.5 dB more negative than those of VO-2 during the latter half of 1977.

C. Changing Signal Levels in 1977

Viking Orbiter signal levels changed by more than 11 dB from the end of 1976 until the beginning of 1978, as shown in Figs. 8 to 11. Figure 17 shows the two trajectory quantities responsible for these changes. The Earth-Mars communication range varied from a maximum in excess of 38×10^7 km late in 1976 to a minimum of less than 10^7 km on Jan. 19, 1978. The Earth-cone angle for the low-gain antenna varied from 0.25 deg on Nov. 25, 1976 (superior conjunction) to a maximum in excess of 40 deg in late 1977.

During the second half of 1977 the Viking telecommunications analysts produced, on a periodic basis, a one-page summary of the expected signal levels. Table 1 is a reproduction of one of these predictions, the one for the peak signal day on Jan. 19, 1978. The predictions were widely used by on-duty Viking project personnel as well as members of the Deep Space Network for rough-estimate planning and for a quick check of the signal levels being observed. On January 19 the one-way light-time had decreased to 5 min and 26 s. Its maximum value was in excess of 21 min in November 1976.

The first major change in operating mode from the primary mission was the use of 16 kbps for imaging data. This occurred in March 1977. On March 28 the signal-to-noise ratio and the bit error rate for the VO-1 and VO-2 16-kbps playback first exceeded threshold requirements. On that date, over DSS 43, the VO-1 signal-to-noise ratio was 1.9 dB measured, and the bit error rate was 1.4×10^{-2} counted (for that indicated signal-to-noise ratio, the predicted bit error rate would have been 1.7×10^{-2}). VO-2 had a signal-to-noise ratio of 3.2 dB measured, and a bit error rate of 2.9×10^{-3} counted (3.9×10^{-3} would be predicted). Thus the confidence level of prediction of the signal-to-noise ratio near the absolute link threshold was raised and also in the correlation between indicated signal-to-noise ratio and the actual bit error rate.

Early in April the Viking Flight Control Chiefs were given authorization to send ground commands to turn the Orbiter

ranging channel off during 16 kbps, if they observed that the indicated 16-kbps signal-to-noise ratio was not sufficiently high. However, the ranging channels were to be left on at all other times unless a tracking station were to announce a major problem, such as failure of the primary S-band maser, prior to the scheduled start of track. The 16-kbps playbacks were planned with 0.7 dB pad through April, which was sufficient to account for observed station performance variations late in March and early in April. For some of the early playbacks, the bit error tolerance in the mission test computer was set to 5 to account for the higher bit error rate expected.

Throughout the period March to September 1977, the Orbiter Science Group requested a waiver to the Viking mission rule requiring the use of 3.0-dB padded threshold for playback of imaging data. This was to enable the use of longer playback opportunities considering the limited 64-m station availability. Beginning with Viking planning cycle 43, effective September 29 for VO-1 and September 30 for VO-2, the full padded 3.0 dB requirement was restored.

Another major change in Viking operations was the use of the 33-1/3 bps data rate during the nonpropulsive maneuvers when the low-gain antenna was in use. The first such use of the 33-1/3 bps rate over a 64-m station was on Sept. 1, 1977, over DSS 63. The signal-to-noise ratio achieved was 5 dB, as predicted. Within a month after that, this data rate was permitted anytime during a 64-m station pass. By early October 1977, the predicted signal-to-noise ratio for low-gain single-subcarrier operation over a 26-m station was 2 dB signal-to-noise ratio for the 8-1/3 bps data. This was far below the 5 dB threshold; however, the Viking Project urgently wanted to use the data mode, because many nonpropulsive maneuvers occurred during 26-m tracks when no 64-m station was available. The choice was between poor-quality data, below the threshold, or no data at all. Even at 2-dB signal-to-noise ratio, high-deck telemetry measurements that contain a large amount of redundancy could provide usable data for determining limit-cycle motion and uplink signal level. In early October, DSS 11 was able to lock the Receiver 2, with a 3-Hz loop filter, and to provide telemetry data at 2+ dB signal level. DSS 61 also managed to lock their receiver and provide 8-1/3 bps data at similar signal-to-noise ratio, when the Orbiter was in the one-way mode. By mid-November, the 26-m 8-1/3 bps single-subcarrier telemetry data reached the normal 5.2 dB threshold. On Dec. 27, 1977, DSS 11 received data from VO-1 on the low-gain antenna, at 33-1/3 bps, and the signal-to-noise ratio was 5 dB, as predicted. The signal-to-noise ratio for this low-gain antenna mode on Dec. 27, 1977 was 12 dB higher than it would have been late in November 1976.

On peak signal day, Jan. 19, 1978, the distance from Mars to Earth was 97.7×10^6 km. The station view periods were changing rapidly at this time, as shown in Fig. 18. At the end of 1977, the tracking set-time changed almost 2 h over a one-month period at DSS 43, working out to about 4 min per day. The indicated VO-2 uplink signal level reached −95 dBm in spacecraft telemetry (64-m station, no uplink modulation), the highest value since interplanetary cruise. The very strong downlink signals, −136 dBm carrier level for dual-subcarrier mode at 64-m stations, also caused the high-rate channel data interference in the demodulation of the low-rate subcarrier to become more prominent. Runs of zeroes into the Orbiter block-coder would result in downward glitches of 5 and 6 dB in the indicated low-rate channel signal-to-noise ratio, reducing it from 16 dB average to about 11 dB minimum.

References

1. Taylor, J., "Deep Space Network to Viking Orbiter Telecommunication Link Effects During 1976 Superior Conjunction", in *The Deep Space Network Progress Report 42-40,* May and June 1977, pp. 57-70.

2. Technical Memorandum 33-783, *Tracking and Data System Support for the Viking 1975 Mission to Mars.* Vol. 1, "Prelaunch Planning, Implementation, and Testing", D. J. Mudgway and M. R. Traxler, Jan. 1, 1977; Vol. 2, "Launch Through Landing of Viking 1", D. J. Mudgway and M. R. Traxler, March 15, 1977; Vol. 3, "Planetary Operations", D. J. Mudgway, Sept. 1, 1977.

3. Stuhr, F. V., *Viking 75 Project Orbiter System, Lander System and Viking Mission Control and Computing Center System to Tracking and Data System Interface Requirements Document,* Vol. II, "Viking Orbiter System to Deep Space Network", Viking Project Document ID-3703111, Dec. 19, 1973.

Table 1. Single-page Viking Orbiter signal levels for peak signal day

Expected signal levels for mid January 1978

Uplink (20 kW, command modulation only)	VO-1	VO-2
26-m, Orbiter HGA	−114	−108.5
26-m, Orbiter LGA	−134	−128.5
64-m, Orbiter HGA	−103.5	−98
64-m, Orbiter LGA	−123.5	−118

Deltas (may be combined)

(a) If station transmits at 10 kW, make levels 3 dB weaker than shown.
(b) If station also has ranging modulation, make levels 9 dB weaker.
(c) If station has no uplink modulation, make levels 2.5 dB stronger.

Single subcarrier downlinks (ranging channel on for Orbiter, 1.5 h from rise/set)

	Downlink AGC	33-1/3 SNR	8-1/3 SNR
26-m, Orbiter HGA	−139.5	> +23[b]	> +23[b]
26-m, Orbiter LGA	−160.5	+5	+10
64-m, Orbiter HGA	−131.5	> +23[b]	> +23[b]
64-m, Orbiter LGA	−152	+15	+20

Deltas (may be combined)

(a) If Orbiter ranging channel is off, make AGC's and SNR's 0.9 dB stronger.
(b) If within one hour of station viewperiod end, make SNR's 1 dB weaker, do not change AGC's.
(c) If on station Maser 2, make SNR's 1 dB weaker, do not change AGC's.

Dual subcarrier downlinks (ranging channel on for Orbiter, 1.5 h from rise/set)

	Downlink AGC	8-1/3 SNR	33-1/3	2 kbps SNR	8 kbps
26-m, Orbiter HGA	−145	+13	+8	+10.5	n/a[c]
26-m, Orbiter LGA	−165.5	< 0	< 0	< 0	n/a[c]
64-m, Orbiter HGA	−136.5	+23	+18	+20.5	+14.5
64-m, Orbiter LGA	−157	+1.5	< 0	< 0	< 0

Deltas: use same deltas as for single subcarrier downlinks.

X-Band downlink (ranging channel on for Orbiter)

	Downlink AGC
64-m, Orbiter HGA	−141.5

Delta: if Orbiter ranging channel is off, make X-band AGC 1.4 dB stronger.

[a]Peak signal day was on Jan. 19, 1978. On this day Earth to Mars range was 97.7×10^6 km. One-way light time was 5 min, 26 s.

[b]Saturated.

[c]Data rate available at 26-m stations.

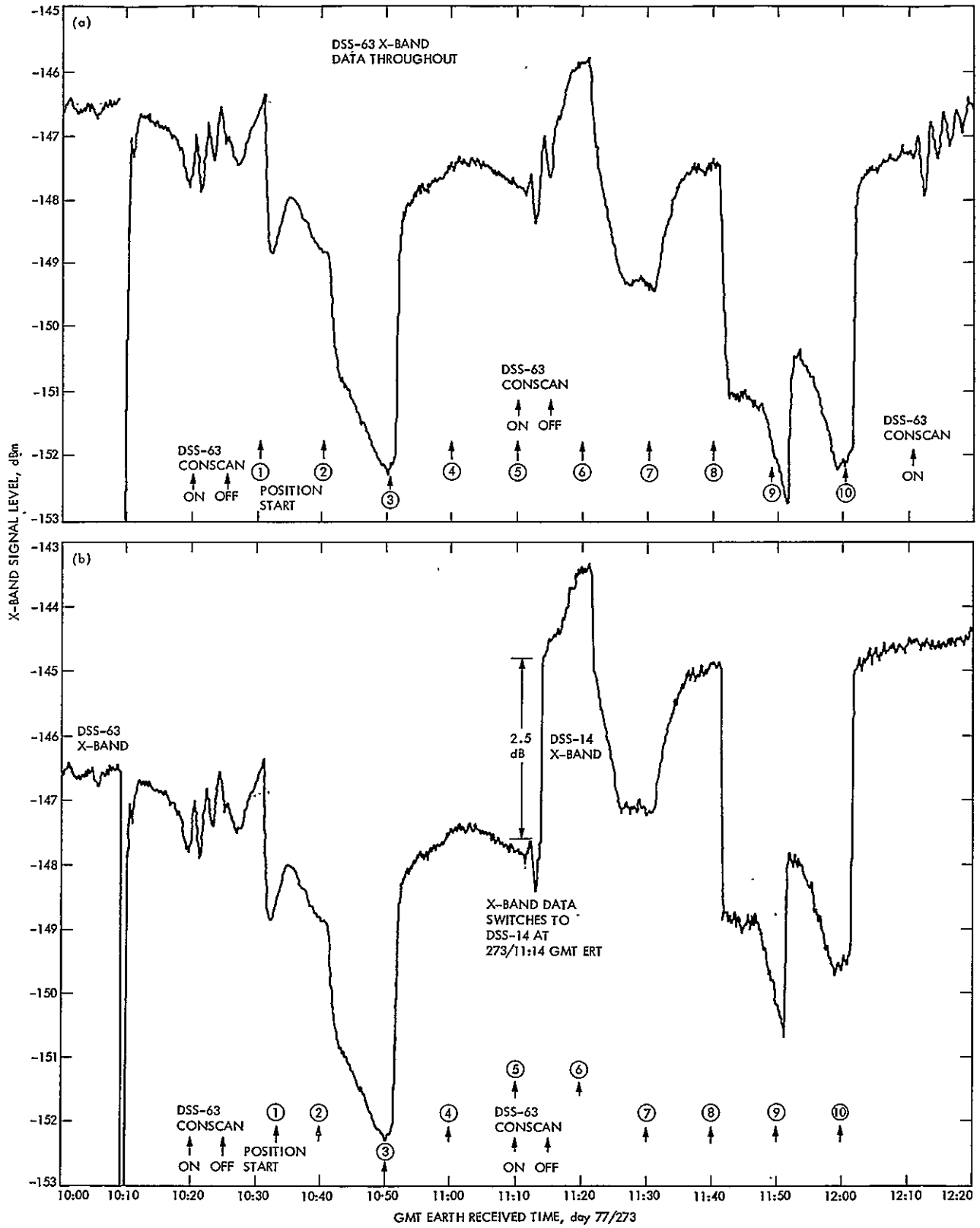

Fig. 1. Viking Orbiter high-gain antenna calibration, Sept. 30, 1977, DSS 63 X-band signal level (a) MTC-A (b) MTC-B

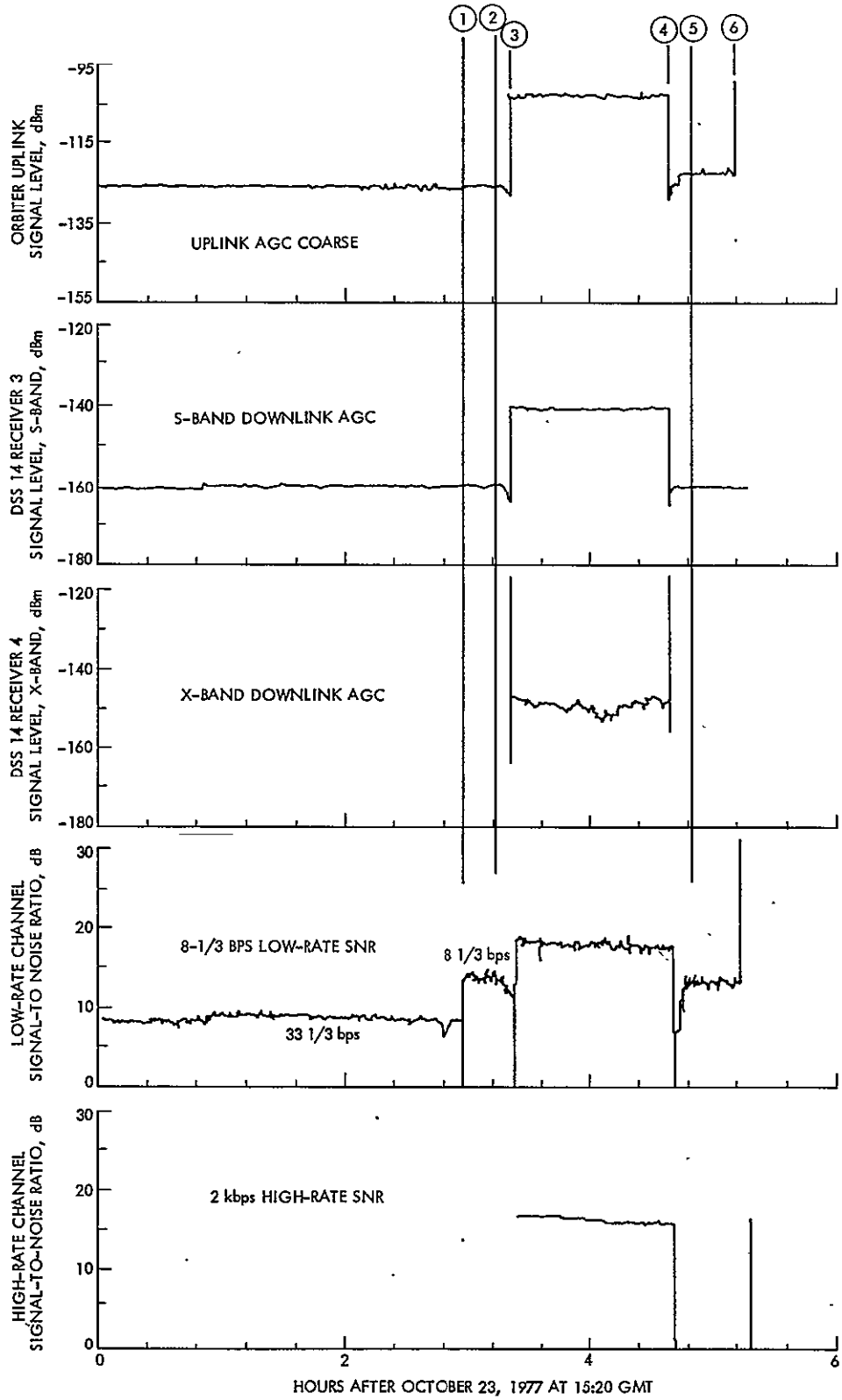

Fig. 2. Link quantities during VO-2 Mars orbit trim No. 13

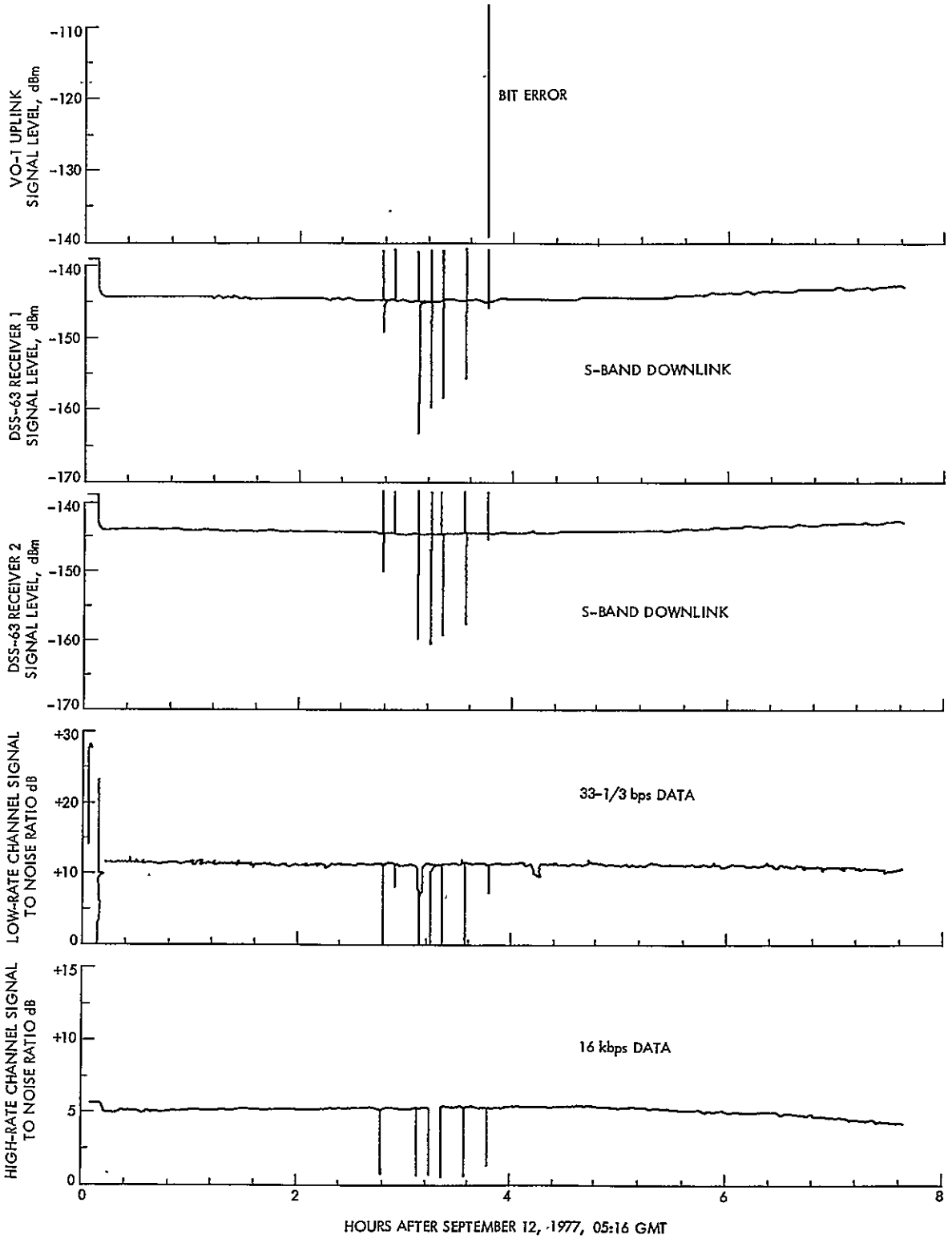

Fig. 3. Signature of uplink ranging modulation interference to one-way orbiter

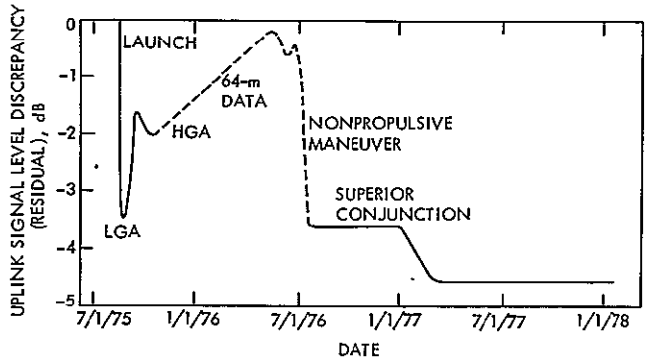

Fig. 4. Overview of variation in VO-1 uplink signal level residual since launch

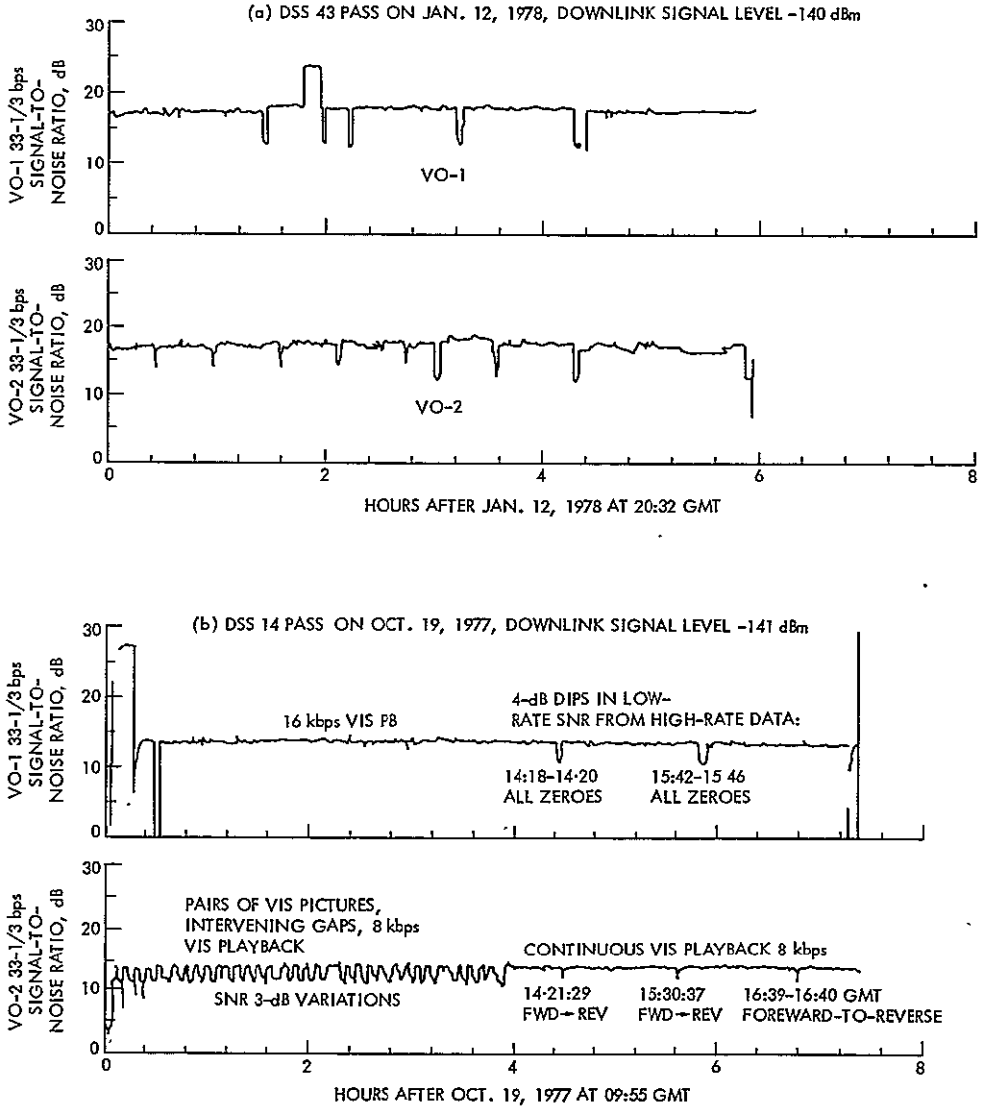

Fig. 5. Low-rate signal-to-noise degradation caused by high-rate data content

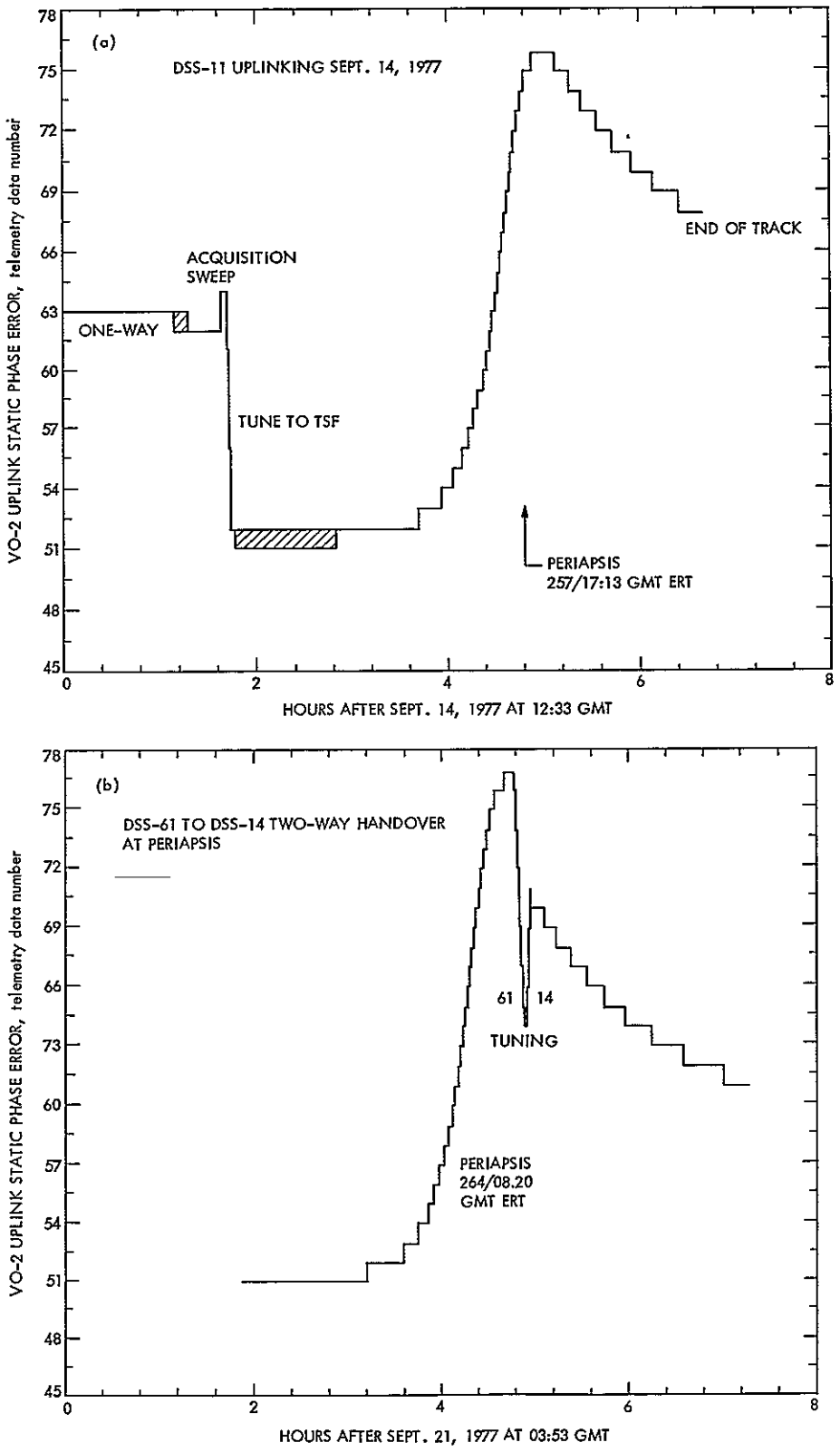

Fig. 6. Uplink acquisitions of VO-2, showing static phase error changes

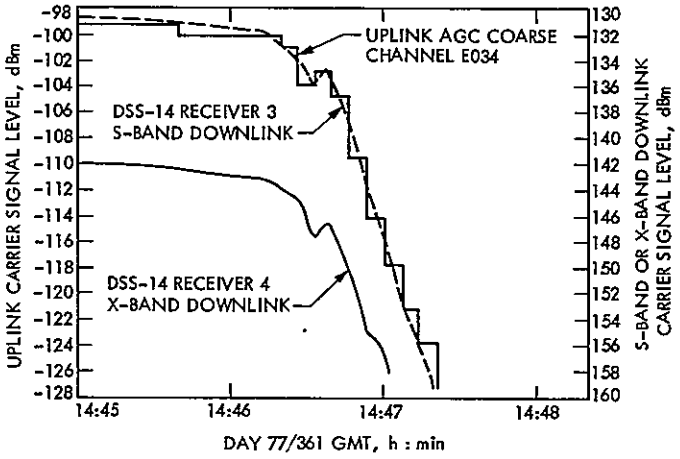

Fig. 7. VO-2 uplink and downlink signal levels prior to entering Earth occultation, DSS-14, Dec. 27, 1977

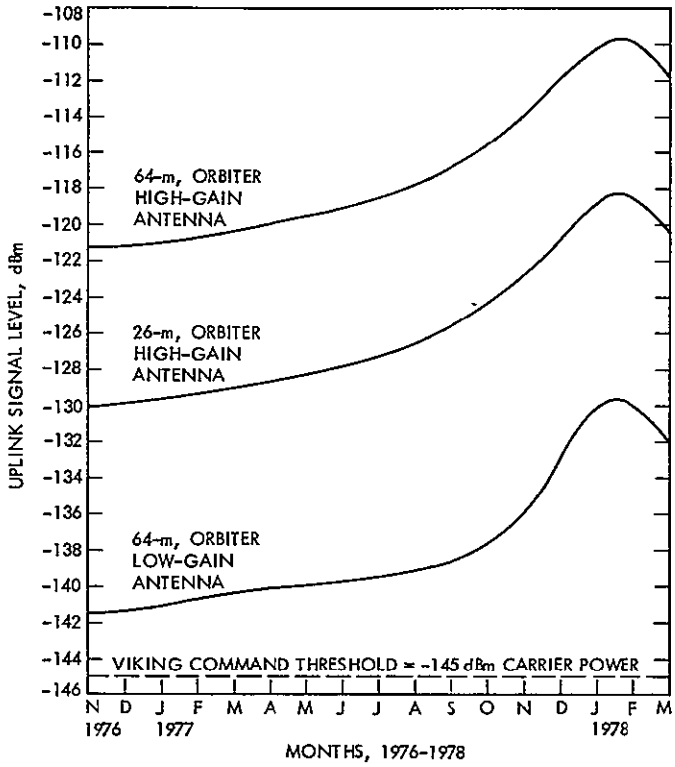

Fig. 8. Predicted uplink signal level, Viking Extended Mission 20-kW transmitter power, command and 9-dB ranging modulation on

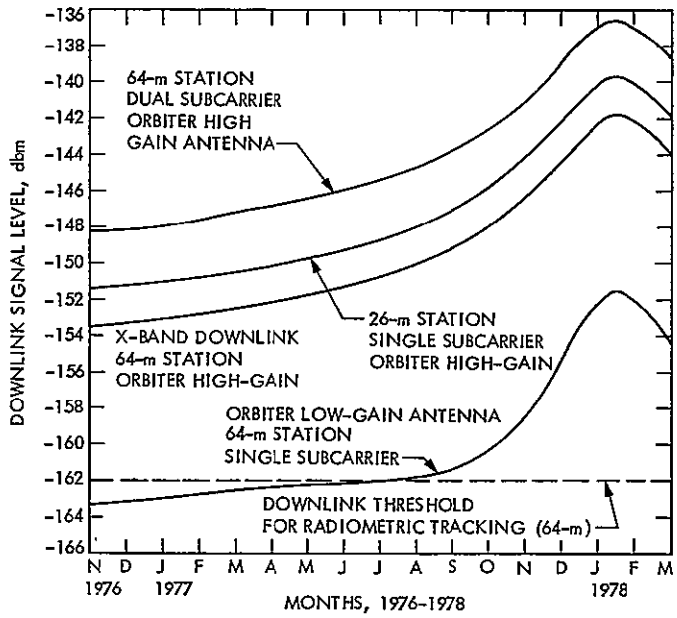

Fig. 9. Predicted downlink signal level, Viking Extended Mission Orbiter high-power mode, ranging channel on

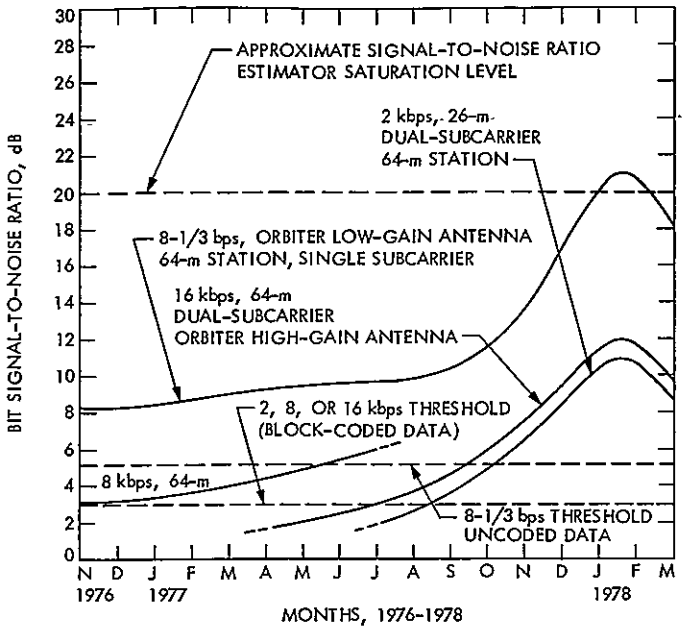

Fig. 10. Predicted signal-to-noise ratio, Viking Extended Mission Orbiter high-power mode, ranging channel on

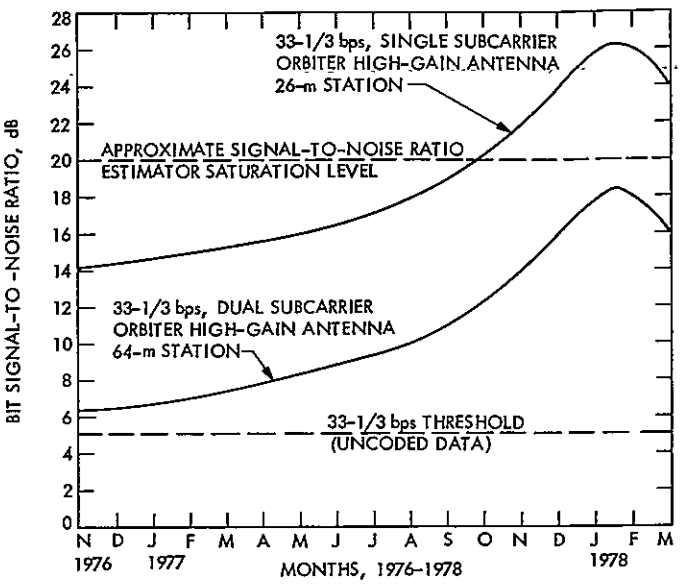

Fig. 11. Predicted signal-to-noise ratio, Viking Extended Mission Orbiter high-power mode, ranging channel on

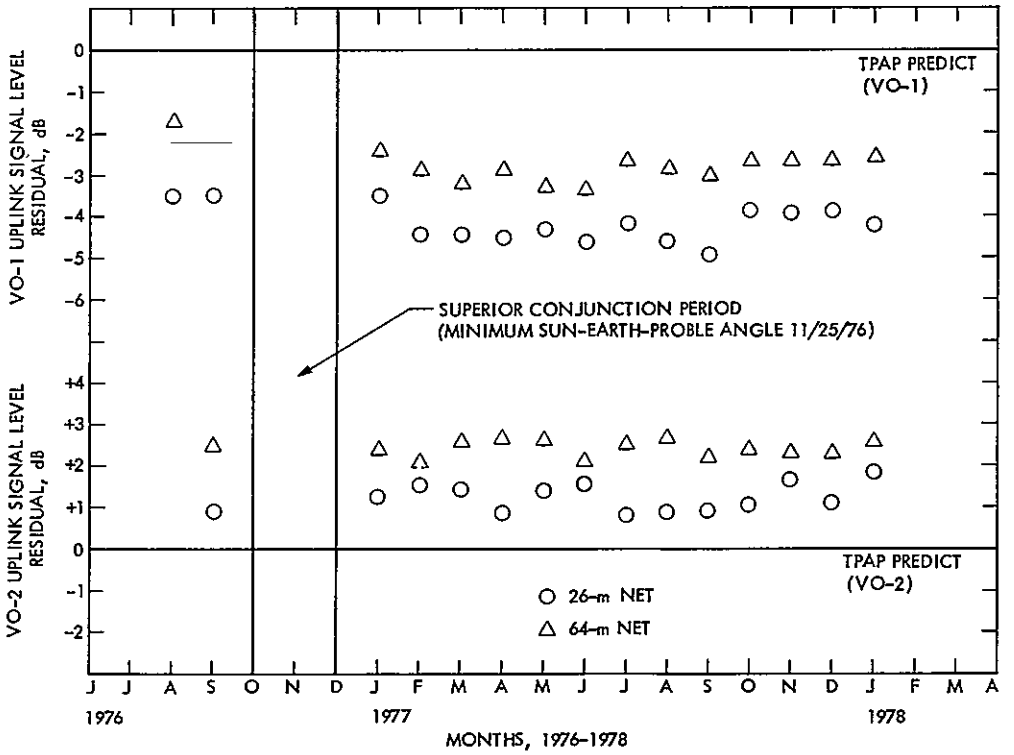

Fig. 12. VO-1 and VO-2 uplink AGC residual (observed mean difference from TPAP predict)

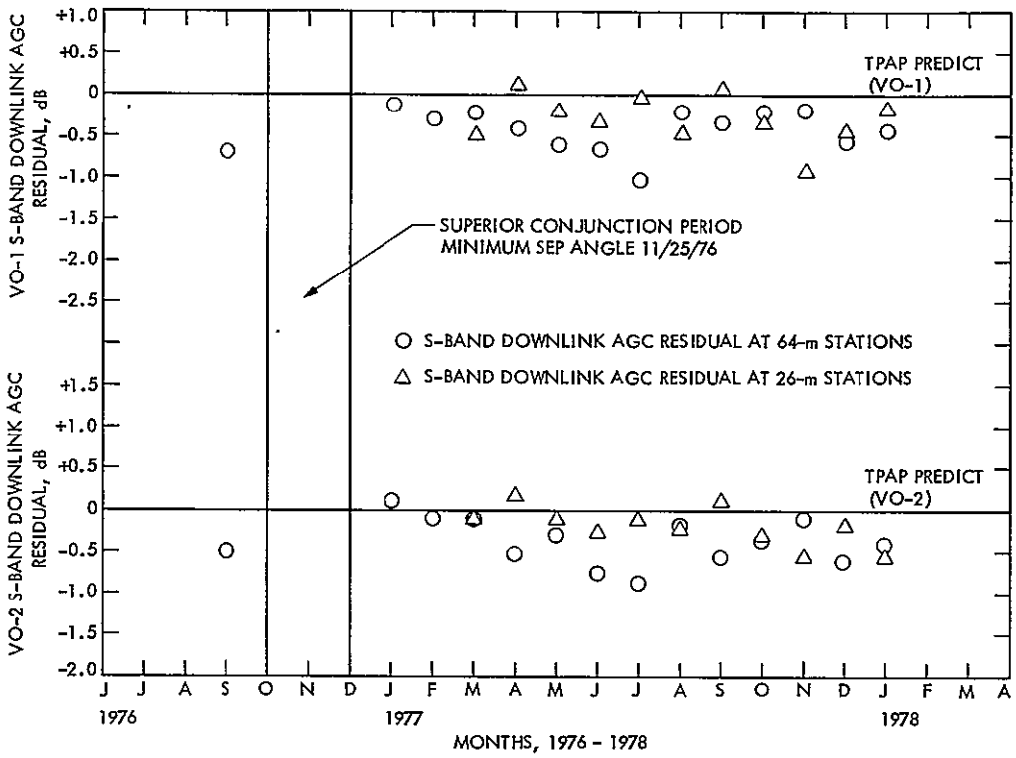

Fig. 13. S-band downlink AGC residual (dual subcarrier and single subcarrier combined, observed mean difference from TPAP predict)

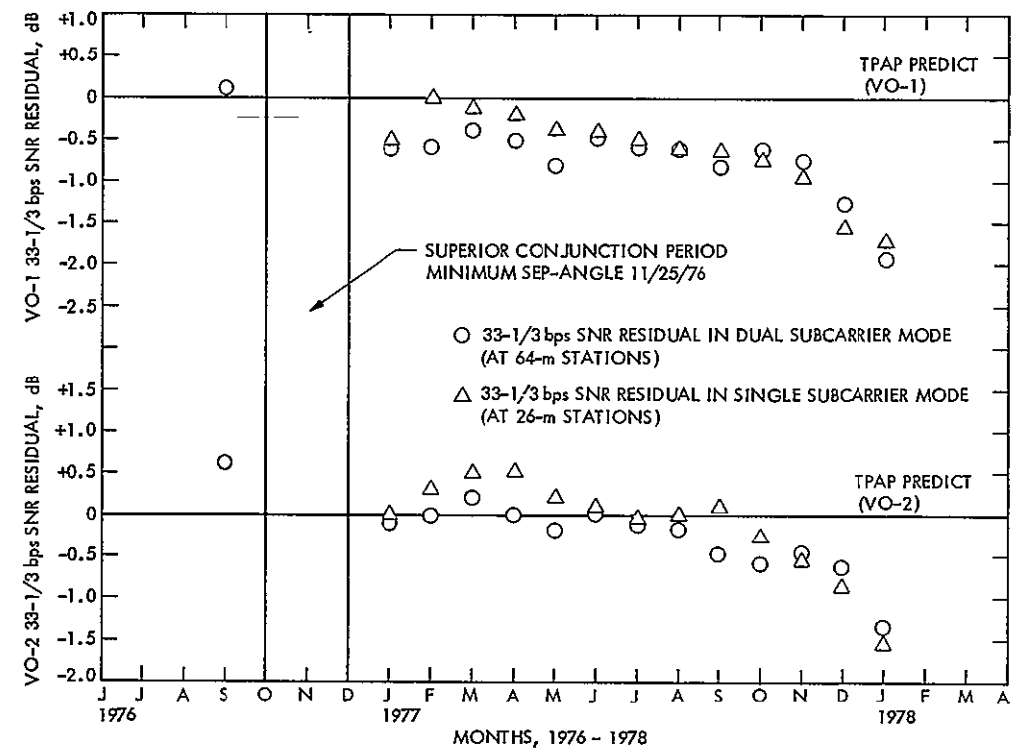

Fig. 14. VO-1 and VO-2 low-rate channel (33-⅓ bps) SNR residuals (observed mean difference from TPAP)

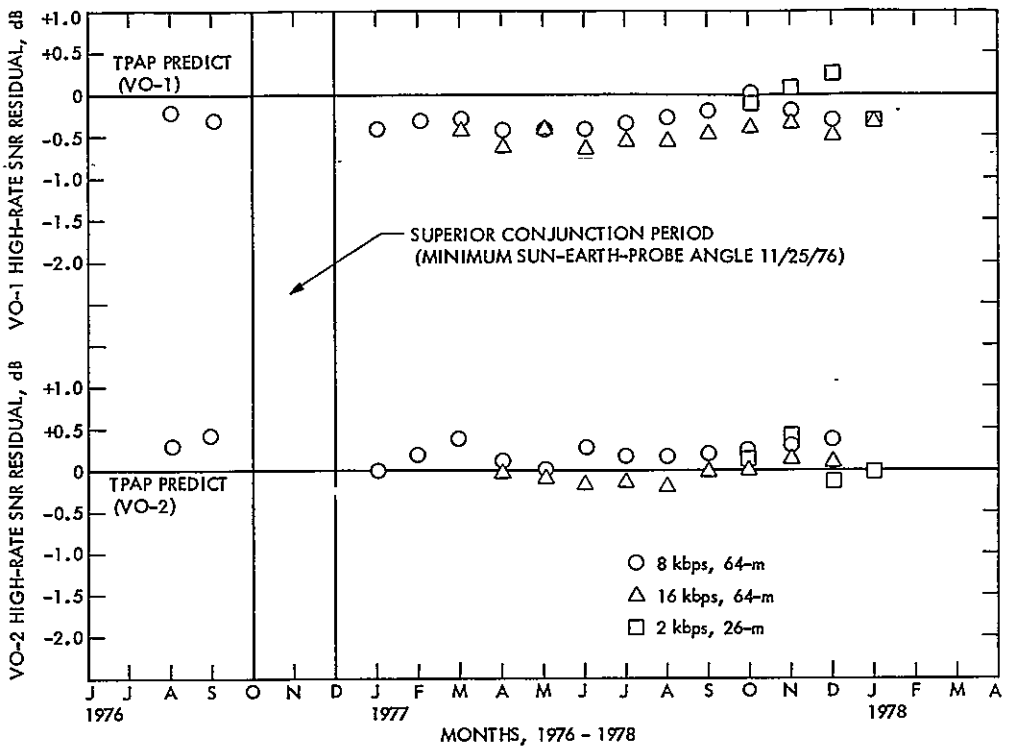

Fig. 15. VO-1 and VO-2 high-rate SNR residual (observed mean difference from TPAP predict)

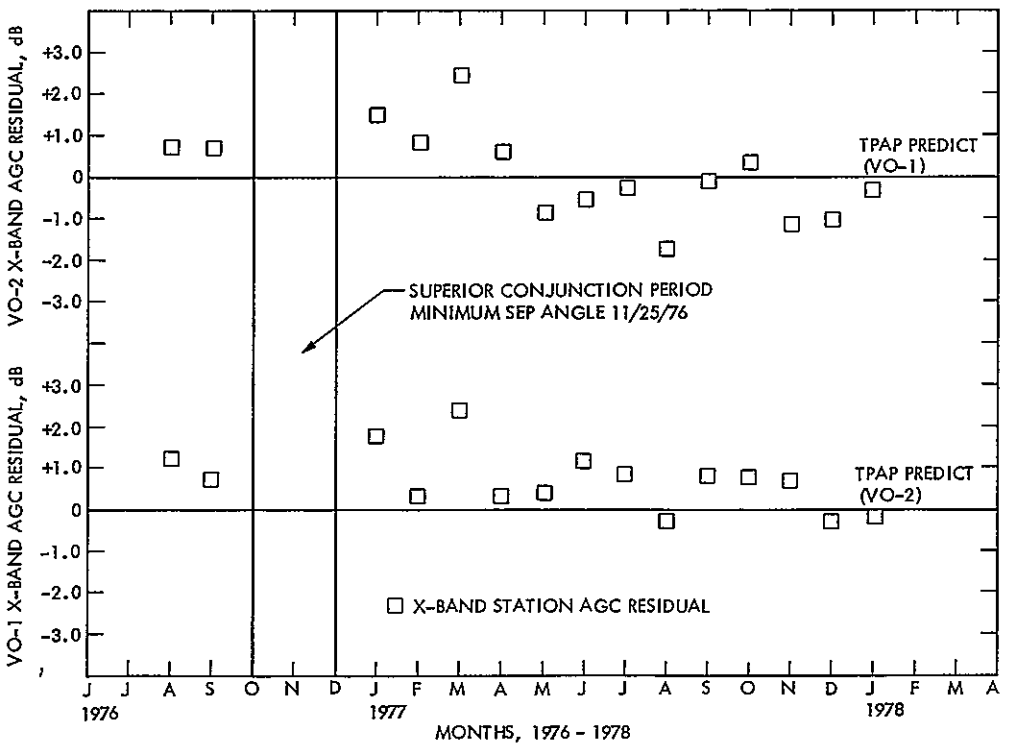

Fig. 16. VO-1 and VO-2 X-band station AGC residuals at 64-m stations (observed - TPAP predict)

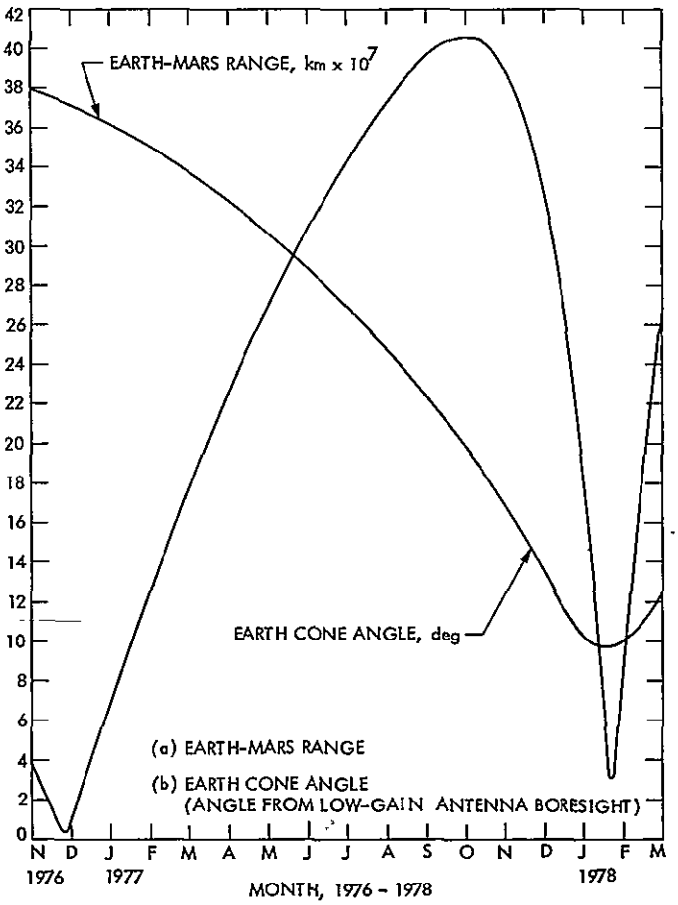

Fig. 17. Trajectory quantities for Viking Extended Mission

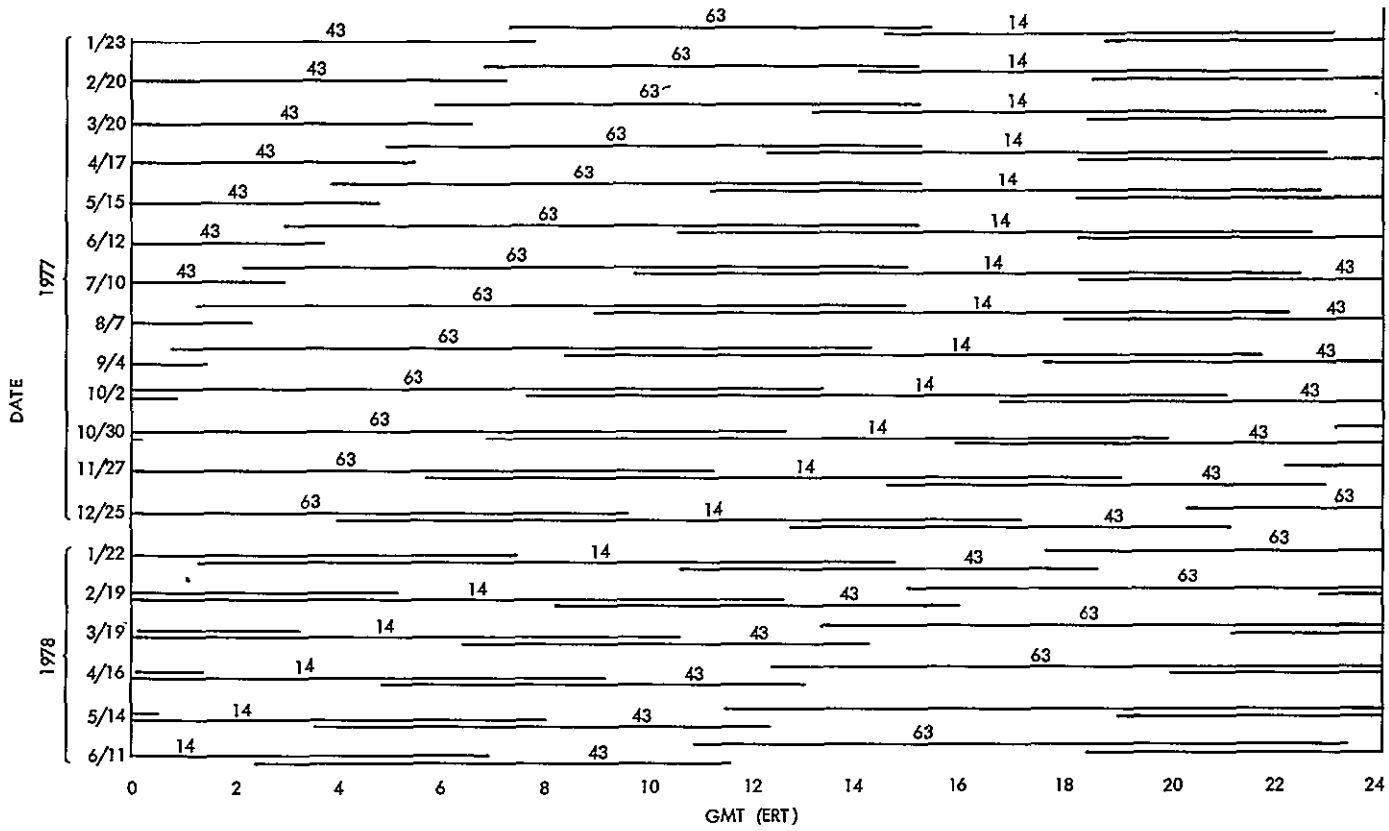

Fig. 18. Deep Space Network station view periods for Viking Extended Mission

Pioneer Venus 1978 Mission Support

R. B. Miller
TDA Mission Support Office

The Tracking and Data System and Deep Space Network preparation and support activities for the Pioneer Venus 1978 Mission are described for the period from June 3, 1977 to April 24, 1978.

I. Introduction

The article contained in Ref. 1 described the DSN support status up to June 3, 1977. The following article describes those activities which have taken place from June 3, 1977 to April 24, 1978.

In general, at the time of the writing of this article, essentially all implementation required for support of the Multiprobe Mission and for the launch of the Orbiter and Multiprobe Missions has been completed. The most significant implementation item outstanding for Pioneer Venus support is the new Radio Science Subsystem required by the Orbiter Mission.

II. Major Project Activities and Milestones

A Preship Review for the Orbiter spacecraft took place at Hughes Aircraft Company February 20, 1978. At that time, the health of the spacecraft and all of the Orbiter instruments were found to be in very good state and the Orbiter was shipped to Cape Canaveral, Florida, arriving the week of March 13, 1978. The Preship Review for the Multiprobe spacecraft took place the week of April 24, and extended for six days at Hughes Aircraft Company. The current plan is for the Multiprobe spacecraft to arrive at the Cape the week of June 5, 1978. The launch window open for the Orbiter Mission is currently scheduled for May 20, 1978, and the launch window nominally extends through the 10th of June, although seven additional days are possible by compromising the arrival date at Venus.

The principal problem encountered in preparing for the launch of the Orbiter Mission has been with the Ames portion of the Ground Data System (GDS). The Project has been tackling an extremely difficult task of trying to accommodate a very complex mission on a very modestly sized computer system at Ames Research Center. The difficulties in accomplishing this task resulted in the GDS testing starting a full two months later than originally desired. Mission Operations Systems (MOS) testing started on the 20th of March 1978. The Project was forced into the position of stretching out a phased delivery of the capabilities required for support of the mission so that GDS testing has had to overlap completely with the MOS testing in preparation for launch. A reduced set of capabilities in telemetry and command were decided upon by the Project as being adequate to support launch. There will, therefore, be an extensive and continued development of the Ames portion of the Ground Data System after the Orbiter launch, extending up to the start of Orbiter operations in

December 1978. The Multiprobe launch opportunity extends from Aug. 7 to as late as Sept. 3, 1978. The Orbiter Mission will insert into orbit around Venus on Dec. 4, and the Multiprobe entry take place on Dec. 9, 1978.

III. DSN Prelaunch Preparations

The DSN achieved the state of readiness to support Project Ground Data System testing in late December 1977, which was about one month later than originally planned. This had no impact on the Project because of complementary delays on the Project side of the Ground Data System. The month delay in the DSN preparations was caused principally by problems encountered in trying to get data throughput using the new high-speed data system which involved 22-bit error polynomials. The new high-speed data system involves automatic switching at JPL and accommodates error detection-correction. Only one significant problem was uncovered outside of communications in the early testing executed by the Network Operations Project Engineer for Pioneer Venus, and that was the initial inability of the Telemetry Processor Assembly to handle Pioneer Venus uncoded data. Once the problem with uncoded data was discovered, it was quickly rectified.

The new equipment and software involved in the high-speed data transmission from the tracking station through JPL and on to the Project has been a continual source of difficulty in preparing for the Pioneer Venus Mission. The problems encountered in communications in the last two months of 1977 and in January 1978 were principally due to trying to use kludges and various patches in order to start Pioneer Venus testing prior to the committed date for the Communications System to be fully operational. The problems from February on have been principally due to less-than-desired reliability in the new hardware and software in the Ground Communications System. Many of the initial problems, of course, were due to operator unfamiliarity with the new hardware and software. The contractor changeover on March 27, 1978 compounded the problem by introducing a fairly significant number of new personnel who required training. As test activities progressed into April, it became clear that many remaining problems were not operator-related and that, indeed, the new hardware and software in the communications system had unsatisfactory reliability. In order to put a concentrated effort on solving the reliability problems, a special task team was formed April 20, 1978 to execute a concentrated effort to find ways of improving the communications reliability.

The second area in which the DSN had difficulty in meeting the schedule and delivering satisfactory reliability performance was in the implementation of a new direct interface between the DSN and the Navigation area for the delivery of radiometric data. This new system involves the generation of Intermediate Data Records of radio metric data by the DSN for direct delivery to Navigation without using the Mission Control and Computing Center's 360-75 computers. This capability was planned to be fully operational by Feb. 1, 1978, but software development problems delayed the system going on-line until March 13, 1978. The system, therefore, went on-line nearly coincident with the contractor changeover, which compounded problems that followed in the inability of the DSN to reliably deliver the radio metric IDR to Navigation. By the second week of April, a growing concern about whether the DSN would achieve satisfactorily reliable flow of the radio metric data to Navigation prior to launch resulted in formation of a briefly lived task team whose efforts led to the Pioneer Navigation Chief declaring on April 17, 1978 that the new radio metric data interface was no longer a major concern in preparing for launch.

IV. Multiprobe Entry Preparations Status

The original plan was to complete the multiprobe entry configuration at DSS 14 by Feb. 15, 1978 and at DSS 43 by March 1, 1978, with only two exceptions. The two exceptions were (1) the redundant recorder for the differential long baseline interferometry (DLBI) experiment which was delayed to an FY-78 procurement to save FY-77 funds; and (2) the automation of certain functions in the spectrum signal indicator, which could not be completed until the summer of 1978. The purpose of establishing the multiprobe configuration well in advance of the mission need date was twofold: first, to allow for adequate procedure development using the actual configuration, and sufficient time for the complex test and training activity that would be required; and second, to enable the tracking and processing of ALSEP data well in advance in order to shake down the end-to-end DLBI wind measurement experiment system.

In mid-December, it became clear that lack of completed ECOs and cables had placed the early completion of the Multiprobe configuration very much in jeopardy. The problem was compounded by the fact that the installation of this major configuration at DSS 14 was going to have to take place concurrent with the DSN Mark III Data Systems (MDS) downtime at DSS 11. The concurrent DSS 11 downtime was a potential problem because of the consolidated Maintenance and Integration crew at Goldstone. The individual cognizant development engineers responsible for the equipment involved in the Multiprobe entry had identified all of their cable requirements well in advance. The breakdown occurred at various levels in the myriad steps that take place between the time that a CDE specifies what cables are necessary and the

delivery of an accepted ECO kit for the installation. The basic problem is that no one person or organization accepts responsibility for coordinating all of those myriad of steps which must take place in various organizations. The DSN has encountered problems of this kind in the past with major implementations, but found that the very major MDS implementation was nearly free of such difficulties because it had been projectized. In order to attack the immediate Pioneer Venus problem, an Implementation Project Manager was named within Office 430 to spearhead the coordination of the remaining implementation and installation for Pioneer Venus. Due to the concerted effort supported by a large number of people (the Cognizant Development Engineers, the Division 33 System Engineers, the Cognizant Operation Engineers, the cable group and Station Maintenance and Integration personnel) the Multiprobe configuration was completed at both stations only one week later than the original (two-year-old) plan.

The remaining implementation for the Multiprobe consists of the redundant transport for the DLBI (Digital Recording Assembly (DRA)) recorders at DSS 14 and 43 which will be shipped to the stations in late April, and microcontrollers to automate some of the Spectral Signal Indicators functions in August 1978.

The completed Multiprobe configuration was utilized in March for four 8-hour procedure development tests. The two Australian shift supervisors who will be handling the actual entry event in Australia came to JPL in April and March 1978 for training. They assisted the Network Operations Project Engineer and his staff in formulating a first cut at the detailed procedures for handling the Multiprobe entry. These procedures were executed during the procedure development tests (with the participation of the Australians) and a further refinement of the procedures developed. Several observations were noted from the March 1978 procedure development tests: First, the Multiprobe entry is every bit as operationally complex as has been anticipated. Second, the acquisition of the Probe data with closed-loop receivers in real-time will be extremely difficult because of the combination of large uncertainties, low-signal levels, and (after blackout) small subcarrier frequencies. Rapid acquisition of the signals in real-time appears to have a lower probability than anticipated. Third, the open-loop receiver precarrier detection recording must be considered the prime means of recovering the data for the actual mission. Fourth, the Spectral Signal Indicators are crucial to successful real-time acquisition of data. Fifth, successful handling of the Multiprobe entry would have been impossible if the DSN had not built simulators for the Probe entry. Sixth, the extensive test and training activity planned for the Multiprobe entry will be essential to ensure success.

V. DLBI Wind Measurement Experiment Status

A major review of the differential long baseline interferometry (DLBI) experiment was held at the Massachusetts Institute of Technology on Aug. 11, 1977. The Review Board consisted of interferometry expertise from around the country who were not directly involved with the experiment. During that review, the detailed design and plans for the Tracking and Data System support for the experiment were reviewed and accepted.

The second week of November 1977, the DSN and STDN provided equipment for the DLBI Experiment underwent a trial installation at the STDN station located at Goldstone, Calif. The primary purpose of this trial installation was to execute a complete integration test for the DSN and STDN equipment prior to the shipment of the equipment to the overseas sites at Guam and Santiago, Chile. Some minor interface problems were discovered and corrected as a part of this activity and actual Apollo Lunar Surface Experiment Package (ALSEP) signals were recorded. These ALSEP recordings were processed using developmental software at JPL to bandwidth-reduce the data to produce the deliverable recording which was further processed by MIT where the recorded ALSEP signals were detected.

The equipment necessary for supporting the DLBI wind experiment was completed at all four stations at the same time that the rest of the Multiprobe entry implementation was completed. Starting Feb. 28, ALSEP tracks have been scheduled every two weeks in order to shake-down the system. During the first ALSEP track of Feb. 28, the two stations scheduled were DSS 14 at Goldstone and the STDN Santiago station. Station 14 experienced considerable difficulties during that attempted track due to a combination of equipment and procedural problems that were compounded by the fact that two of the three ALSEPs that were to be used for that track were malfunctioning. The end-to-end DLBI experiment system involves the taking of data at at least two tracking stations, the shipment of the data from overseas to JPL, processing of that data at JPL, and then shipment of the data to MIT where the final processing takes place. It will not be known whether the equipment is working sufficiently well to enable the experiment to take place until the final processing takes place at MIT and fringes are detected. This process involves, in the best case, on the order of a month from the time data is taken until MIT can declare whether the interferometer is working and required this implementation to be completed early enough to give several chances for a complete end-to-end check by taking multiple station data using ALSEP's as the signal source. After the first track, there were problems in subsequent tracks associated with tape shipment time. One station's tapes were

lost, one station's tapes arrived degaussed, and so far one incident of hardware failure. The majority of problems encountered have not been technical. The first two-station data successfully processed at JPL was shipped to MIT on April 15, (more than a month after equipment installation was completed), but MIT had not completed evaluation at the time of this writing. Several procedural problems were resolved in the process, but protracted tape shipment problems have not yet been solved. It may be necessary for the actual mission tapes to be hand-carried from the overseas sites. Although not enough time has elapsed for MIT to evaluate the data, the processing of the data at JPL does involve a low-resolution, fast Fourier transform of the reduced-bandwidth-data-and-all indications are (as best as can be determined, looking at one station at a time) that the station equipment and the bandwidth reduction equipment at JPL are operating properly.

Reference

1. Miller, R. B., "Pioneer Venus 1978 Mission Support", in *The Deep Space Network Progress Report 42-40*, pp. 14-20, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1977.

Pioneer Venus 1978 Deep Space Network Telecommunications Compatibility Test Program Status

A. I. Bryan and R. P. Kemp
TDA Engineering Section

The Pioneer Venus 1978 Flight Project DSN Telecommunications Compatibility Test Program consists of three phases: Subsystem Design, System Design and System Verification Tests, which are to be performed at JPL and at the U.S. Air Force Eastern Test Range and Kennedy Space Center Complexes. Subsystem Design Tests were performed during April 1977. A subset of System Design Tests were performed during November 1977. This article describes the tests that have been completed through 1977.

I. Introduction

This report summarizes the DSN Pioneer Venus 1978 (PV'78) Flight Project Telecommunications Compatibility Test Program, covering the period from April 1977 through November 1977. This test program is following the standard three-phase plan of establishing telecommunications compatibility as specified in the DSN Standard Practice Document, *Deep Space Network – Flight Project Interface Compatibility Test Design Handbook,* and in the Pioneer Venus Project – DSN Spacecraft Compatibility Test Plan.

The plan specifies that telecommunications design compatibility will be established at the subsystem and system levels and conclude with a final verification at Cape Canaveral, Florida, prior to launch. In addition, the DSN and flight project equipment, software configurations, requirements and test objectives in all phases are specified.

Procedures for conducting the tests as well as test design criteria and test parameters for the ground station hardware and software were prepared by Network personnel. Spacecraft telecommunications design performance criteria and test parameters to establish flight project nominal and threshold telecommunications conditions were provided by the PV'78 Telecommunications group. The test criteria were included as part of the test procedures to provide real-time assessment of performance. All test procedures were jointly approved by the DSN and flight project representatives.

II. Types of Tests

As of this report, the test program has included Subsystem Design Tests (Phase I) with the PV'78 Prototype and System Design Tests (Phase II) with the PV'78 Orbiter, BUS and the three Small Probe Spacecrafts. System Design Tests yet to be

completed are those involving the PV'78 Large Probe and the Multisignal (five simultaneous downlink signals) tests.

III. Objectives

The objectives of the test program is to demonstrate, in the major areas of radio frequency acquisition and tracking, command and telemetry, compatibility between the spacecraft telecommunications subsystems and the Network, to establish system design compatibility between the spacecraft and the Network, and to verify continued interface integrity and maintenance of compatibility during prelaunch activities.

IV. DSN Subsystem Compatibility Tests

A. PV'78 Prototype, 5–13 April 1977

1. Test objectives. The objectives of these tests were to establish telecommunications design compatibility between the DSN and the Pioneer Venus Project in the areas of Radio Frequency acquisition, tracking and telemetry. Specifically, the objectives were as follows:

(1) Verify the capability to receive S-band and X-band carrier signals from the PV'78 Prototype Spacecraft Orbiter RF Subsystem without degradation of DSN receiver thresholds.

(2) Verify the capability to receive S-band carrier signals from the DSN transmitter without degradation of the PV'78 Orbiter receiver threshold.

(3) Determine the maximum DSN transmitter offsets and the minimum DSN transmitter sweep rates for reliable PV'78 receiver acquisition.

(4) Determine the maximum DSN transmitter sweep rates for proper PV'78 receiver tracking.

(5) Verify the DSN Telemetry System capability to receive and process coded and uncoded telemetry without degradation.

All tests performed were in accordance with JPL internal Document 810-8; Rev. B, *Deep Space Network/Flight Project Interface Compatibility Test Design Handbook.*

2. Test conditions. The Pioneer Venus Prototype Transponder and Telemetry Subsystems were located in the Compatibility Test Area RF screen room at JPL. The Radio Frequency Subsystem was configured as follows:

(1) S-band: receiver, Channel 17 (2114.335648 MHz) transmitter, equipped with solid state amplifier coherent and non-coherent modes of operation

(2) X-band: transmitter, equipped with solid state amplifier

The DSN, as represented by CTA 21, was configured to simulate a Pioneer Venus Flight Project committed Deep Space Station. The ground hardware included the Block IV Receiver-Exciter Subsystem and the Mark III Data Subsystem for telemetry.

The S-band and X-band RF links between the DSN and the PV'78 prototype transponder were coaxial cables (with suitable attenuators) which had previously been calibrated for attenuation losses and impedance characteristics. The DSN software provided at CTA 21 was the preliminary (unreleased) Pioneer Venus operational software.

3. Test results. Table 1, DSN Pioneer Venus Prototype Spacecraft Telecommunications Compatibility Test Summary provides a listing of test configurations, test criteria, parameters and results. Refer to Fig. 1 for DSN mode configurations. Significant test results and comments are discussed below:

a. Radio frequency acquisition and tracking. All objectives of this phase of testing were successfully met. However, while performing the rest frequency measurement, it was observed that the rest frequency of the VCO was quite unstable under no signal conditions. Additionally, it was discovered that the prototype receiver VCO was being "pushed" while ramping the uplink frequency (S-band) at 5 Hz per second. Investigation of these two conditions revealed that both problems were known to exist under similar test conditions at the factory. The test conditions and parameters were modified as follows:

(1) The prototype receiver was located on an open table in the RF screen room and was subjected to a changing ambient temperature environment. An insulated enclosure was constructed around the prototype receiver which allowed the VCO to meet the stability requirements of the test.

(2) The uplink ramp rate was changed from 5 Hz per second to 10 Hz per second and successful acquisition of the spacecraft receiver was accomplished.

b. Telemetry. Initial attempts to perform sequential decoding of Orbiter spacecraft convolutionally encoded data resulted in aborted acquisitions by the DSN sequential decoder. This condition was encountered regardless of data rate and/or signal level. It was observed, however, that the DSN Telemetry Subsystem performed satisfactorily on station generated simulated spacecraft encoded data and performed satisfactorily on the prototype spacecraft encoded data up to the sequential decoder itself.

During the performance of trouble-shooting procedures, the spacecraft data generation program was re-initialized, and that

seemed to fix the problem. Every morning for the remainder of the test, the spacecraft data generation program had to be re-initialized to correct the problem before testing could begin.

As the telemetry tests got underway, it was noted that during periods when the signal strength was high enough that there should not have been any errors in the telemetry, the decoder was detecting and correcting errors. This was indicated by the fact that the number of computations per bit was more than 1.00 and the deletion rate was 0.00.

By taking a dump of the raw data symbols from the telemetry processor assembly data symbol buffer, it was discovered that the spacecraft encoder reset pulse was not being applied properly at the beginning of the 24-bit tail sequence.

The problem was independently verified by the spacecraft hardware personnel by looking at the encoder signals on an oscilloscope.

After the encoder was reworked to the correct configuration, satisfactory operation of the sequential decoder was achieved.

The encoder that was initially configured incorrectly was an orbiter-type encoder. To insure that the probe-type encoder was correctly configured, an analog recording of the encoded data-stream from one of the flight units was prepared at HAC and brought to JPL for processing through the DSN telemetry equipment. This effort proved successful and demonstrated that the probe-type encoders were configured correctly.

Following resolution of the encoder problem, the telemetry threshold tests were completed. The coded telemetry tests were performed over a two-day period with a noticeable difference in performance for the two days. On the first day, the decoder was deleting frames (as expected) as the input signal level was lowered. However, during the second day of testing, the deletion rate did not exceed zero until the sequential decoder completely dropped lock from lack of input signal. Because of insufficient performance data at threshold conditions with the new DSN sequential decoder, it was not known which, if either, condition was normal.

Subsequent to the subject tests, a software error that prevented proper read-out of the frame deletion statistics was discovered and corrected. After this software error had been identified, a new series of tests were performed at CTA 21 to determine proper operation (frame deletion rate as a function of input signal level) of the new sequential decoder.

A plan was generated by the DSN to extensively test telemetry performance for sequential decoding in order to identify and document proper performance characteristics. The plan utilized was to repeat only the coded data portions of the telemetry threshold tests performed during the DSN PV'78 Prototype testing. The spacecraft telemetry generating function was simulated utilizing the Simulation Conversion Assembly (SCA) and the DSN test transmitter. The resulting signal was processed through the Block III Receiver, Subcarrier Demodulation Assembly (SDA) and the Telemetry Processor Assembly (TPA).

In order to facilitate test time, a frame deletion rate of 10^{-2} was used as the criterion for a specified signal-to-noise (ST_b/N_o) input level. Additionally, revised ST_b/N_o values were used. Based on theoretical system losses and measured performance characteristics of the old Data Decoder Assembly (DDA), these revised ST_b/N_o values reflected a more accurate estimate of sequential decoder performance and project requirements. Results of the tests were satisfactory as all frame deletion rate (DLR) measurements were well within the tolerances of the controlled test conditions, i.e., setting of ST_b/N_o by the y-factor method. In addition to DLR data, the symbol error rate (SER) for each test was also provided.

The telemetry test data shown in Table 1 are the data that were obtained during the sequential decoder performance testing without the prototype spacecraft.

V. DSN System Compatibility Tests

A. PV'78 Orbiter, 1–4 November 1977

1. Test objectives. The objectives of these tests were to establish telecommunications design compatibility between the DSN and the Pioneer-Venus Orbiter in the areas of radio frequency acquisition, tracking, telemetry, and command at strong signal levels. Specifically, the objectives were as follows:

(1) Determine the maximum DSN transmitter offsets and the minimum DSN transmitter sweep rates for reliable PV'78 receiver acquisition.

(2) Determine the maximum DSN transmitter sweep rates for proper PV'78 receiver tracking.

(3) Verify the DSN Telemetry System capability to receive and process coded and uncoded telemetry.

(4) Verify the DSN Command System capability to successfully transmit timed commands, untimed commands and contiguous commands to the PV'78 Orbiter.

(5) Verify the DSN spectral signal indicator (SSI) to display and correctly identify major PV'78 RF signal components.

(6) Verify the DSN capability to perform predetection recording and playback on actual PV'78 RF signals and telemetry data.

All tests performed (with the exception of (5) and (6)) were in accordance with JPL internal Document 810-8, Rev. B, *Deep Space Network/Flight Project Interface Compatibility Test Design Handbook.* Tests (5) and (6) were performed to demonstrate the operational capabilities of the SSI and pre-detection recording and playback.

2. Test conditions. The Pioneer-Venus Orbiter Spacecraft was located at the Hughes Aircraft Company, El Segundo, California. The Radio Frequency Subsystem was configured as follows:

(1) S-band: Receiver No. 1, Channel 11 (2112.290394 MHz); Receiver No. 2, Channel 12 (2112.629760 MHz); and Transmitter No. 1 and No. 2, equipped with 20-W solid state amplifiers

(2) X-band: Not tested

The DSN, as represented by CTA 21, was configured to simulate a Pioneer-Venus Flight Project committed Deep Space Station. The ground hardware included the Block III and IV Receiver-Exciter Subsystems and the Mark III Data Subsystem for telemetry and command.

The S-band RF link between the DSN and the PV'78 Orbiter Spacecraft was an open RF link between the Hughes Aircraft Company, El Segundo, California and the Jet Propulsion Laboratory, Pasadena, California. A description of the link is provided in the Appendix. The link had been previously tested for amplitude and phase stability and was found to be acceptable for strong signal testing.

The DSN software provided at CTA 21 was the Pioneer-Venus operational software for telemetry and command.

3. Test results. Table 2, DSN Pioneer-Venus Orbiter Spacecraft Telecommunications Compatibility Test Summary, provides a listing of test configurations, test criteria, parameters, and results. Refer to Fig. 1 for DSN mode configurations. Significant test results and comments are discussed below:

a. Radio frequency acquisition and tracking All objectives of this phase of testing were successfully met. The acquisition test was performed by off-setting the CTA 21 exciter plus and minus 5 kHz about the spacecraft receiver theoretical best-lock frequency, and ramping toward best lock until spacecraft receiver acquisition was achieved. The two acquisition frequencies were averaged and became the actual best lock frequency.

The tracking test was performed by ramping the spacecraft receiver to plus and minus 100 kHz about center (best lock) frequency at a rate of 100 Hz per second.

b. Telemetry. The telemetry tests were begun by first performing predetection recording and playback of 16.0 bits/s coded and 256.0 bits/s coded data rates. This particular set of tests was performed in order to demonstrate that the predetection recording and playback technique was viable utilizing actual spacecraft data in preparation for the multisignal tests which are scheduled to be performed at a later date. Results of these tests indicate that the system was satisfactory for recording and playback of the data.

As part of these compatibility tests, the spectral signal indicator (SSI) being developed by Section 333 was made available for displaying and analyzing the PV'78 Orbiter RF spectrum. Results of the tests that were performed with the SSI indicated quite clearly that the SSI has extremely good sensitivity and resolution. It was easily demonstrated that the RF carrier, subcarrier and data modulation components could be identified.

The remainder of the telemetry processing tests were performed with no significant problems noted. It was discovered that a 5 db error in observed SSA SNR values versus expected values was the result of a shifted AGC curve in the Block IV receiver. The AGC curves were recalibrated and testing was continued without impact.

A significant product of these tests was the telemetry subsystem lock-up times. These data provide a good measure of the software operational capabilities under relatively strong signal level conditions. In addition to the lock-up times, the sequential decoder deletion rate, average computation per bit and symbol error rate data were also provided.

c. Command. In order to test the DSN PV'78 Orbiter command interface for compatibility, a variation from the test plan, PC'464, was utilized. It was agreed that since the PMOCC was not yet ready to support the command test, as specified in PC-464, the commands would be transmitted from CTA 21 utilizing the manual mode capability of the software.

All commands scheduled to be utilized during the compatibility test were successfully transmitted by CTA 21 and successfully processed by the PV'78 Orbiter Spacecraft.

During initial attempts to send the single timed commands, it was discovered that an interference problem by the HAC ground support equipment was preventing proper reception of the DSN generated commands. The HAC ground support equipment configuration included the capability of locally

generating spacecraft commands via a hardline to the spacecraft command processor. After several unsuccessful attempts to send the single timed commands, HAC reported that it was a known problem that the hardline configuration would inhibit commands received through the RF system when command processor No. 1 was receiving commands via receiver No. 1 and when command processor No. 2 was receiving commands via receiver No. 2. However, in the cross-strapped configuration (receiver No. 1 – command processor No. 2/Receiver No. 2 – command processor No. 1), this condition did not exist.

After resolving this problem, it was discovered that several commands provided to the DSN contained erroneous bit-structures. When the proper command bit structures were made available, the remainder of the command sequences were sent successfully and processed successfully by the spacecraft.

Strip chart recordings of the DSN command subsystem displayed an obvious difference in several bit lengths of the Idle 2 mode just prior to a timed preamble. It was subsequently learned that this was a normal occurrence (and one that does not adversely affect spacecraft command processor operation) because of the manner in which the Command Modulator Assembly (CMA) must achieve 0.1-s command synchronization accuracy.

B. PV'78 Bus, 8–10 November 1977

1. Test objectives. The objectives of these tests were to establish telecommunications design compatibility between the DSN and the Pioneer Venus BUS in the areas of radio frequency acquisition, tracking, telemetry, and command at strong signal levels. Specifically, the objectives were as follows:

(1) Determine the maximum DSN transmitter offsets and the minimum DSN transmitter sweep rates for reliable PV'78 BUS receiver acquisition.

(2) Determine the maximum DSN transmitter sweep rates for proper PV'78 Bus receiver tracking.

(3) Verify the DSN Telemetry System capability to receive and process coded and uncoded telemetry.

(4) Verify the DSN Command System capability to successfully transmit timed commands, untimed commands and contiguous commands to the PV'78 Bus.

All tests performed were in accordance with JPL Document 810-8, Rev. B, *Deep Space Network/Flight Project Interface Compatibility Test Design Handbook.*

2. Test conditions. The Pioneer Venus bus spacecraft was located at the Hughes Aircraft Company, El Segundo, California. The Radio Frequency Subsystem was configured at S-band as follows: Receiver No. 1, Channel 6 (2110.583352 MHz); Receiver No. 2, Channel 8 (2111.268744 MHz); and Transmitter No. 1 and No. 2, equipped with 20-W solid-state amplifiers.

The DSN, as represented by CTA 21, was configured to simulate a Pioneer Venus Flight Project committed Deep Space Station. The ground hardware included the Block III and IV Receiver-Exciter Subsystems and the Mark III Data Subsystem for telemetry and command.

The S-band RF link between the DSN and the PV'78 Orbiter spacecraft was an open RF link between the Hughes Aircraft Company, El Segundo, California and the Jet Propulsion Laboratory. A description of the link is provided in the Appendix. The link had been previously tested for amplitude and phase stability and was found to be acceptable for strong signal testing.

The DSN software provided at CTA 21 was the Pioneer-Venus prereleased operational software for telemetry and command.

3. Test results. Table 3, DSN Pioneer Venus Bus Spacecraft Telecommunications Compatibility Test Summary provides a listing of test configurations, test criteria, parameters, and results. Refer to Fig. 1 for DSN mode configurations. Significant test results and comments are discussed below:

a. Radio frequency acquisition and tracking. All objectives of this phase of testing were successfully met. The acquisition test was performed by off-setting the CTA 21 exciter plus and minus 5 kHz about the spacecraft receiver theoretical best lock frequency, and ramping toward best lock until spacecraft receiver acquisition was achieved. The two acquisition frequencies were averaged and became the actual best lock frequency.

The tracking test was performed by ramping the spacecraft receiver to plus and minus 100 kHz about center (best lock) frequency at a rate of 100 Hz per second.

b. Telemetry. The telemetry processing tests were performed with no problems noted. Downlink signal levels were adjusted to values which yielded frame deletion rates (DLR) that were equal to or slightly less than zero. Digital original data records (DODR) were made of all bit rates for later playback and evaluation by the Pioneer Mission Operations Control Center (PMOCC). These records will aid in finalizing the PMOCC software.

c. Command. In order to test the DSN PV'78 Bus command interface for compatibility, a variation from the test

plan, PC-464, was utilized. It was agreed that since the PMOCC was not ready to support the command test, as specified in PC-464, the commands would be transmitted from CTA 21 utilizing the manual mode capability of the software.

All commands that were scheduled to be sent to the Bus spacecraft during these tests were successfully sent and processed. However, as was the case with the Orbiter spacecraft, several operational problems had to be overcome in order to complete the test.

Initial attempts to send single timed commands to the spacecraft were unsuccessful because the command bit structures were such that the probes had to be mated to the Bus for the commands to be successfully processed. Later it was discovered that one of the contiguous timed commands (CMD 8) was also in error and had to be changed. With these errors corrected, all commands were successful.

In addition to the normal command bit structures; a second group of single timed commands were successfully transmitted and processed that were composed of a leading hexidecimal "zero" at the beginning of each command. This special test assured that the spacecraft command detector maintained lock in the absence of four consecutive one's and zero's.

Strip chart recordings were made of all DSN generated commands to the spacecraft.

C. PV'78 Bus Small Probe (Dual Subcarrier), 11 November 1977

1. **Test objectives.** The objective of this test was to establish telecommunications design compatibility between the DSN and the Pioneer-Venus BUS-Small Probe (dual subcarrier) in the area of telemetry at strong signal levels. Specifically, the objective was to verify the DSN Telemetry System capability to receive and process coded telemetry from a single downlink with dual subcarriers (PCM/PSK/PM). This test was performed in accordance with JPL Document 810-8, Rev. B, *Deep Space Network/Flight Project Interface Compatibility Test Design Handbook.*

2. **Test conditions.** The Pioneer Venus Bus Small Probe Spacecraft was located at the Hughes Aircraft Company, El Segundo, California. The Radio Frequency Subsystem was configured at S-band as follows: receiver No. 2, Channel 12 (2110.583328 MHz); transmitter No. 2, equipped with 20-W solid-state Amplifier; and Auxiliary Oscillator No. 2 (2292.034784 MHz).

The DSN, as represented by CTA 21, was configured to simulate a Pioneer Venus Flight Project committed Deep Space Station. The ground hardware included the Block III and IV Receiver-Exciter Subsystems and the Mark III Data Subsystem for telemetry and command.

The S-band RF link between the DSN and the PV'78 Orbiter spacecraft was an open RF link between the Hughes Aircraft Company and the Jet Propulsion Laboratory. A description of the link is provided in the Appendix. The link had been previously tested for amplitude and phase stability and was found to be acceptable for strong signal testing.

The DSN software provided at CTA 21 was the Pioneer Venus prereleased operational software for telemetry and command.

3. **Test results.** Table 4, DSN Pioneer-Venus Bus-Small Probe Spacecraft Telecommunications Compatibility Test Data and Table 5, Telemetry Lock-Up Times provides a listing of test configurations, test criteria, parameters, and results. Refer to Fig. 1 for DSN mode configurations. Significant test results and comments are discussed below:

a. Telemetry. The dual subcarrier telemetry test was completed with no problems noted. All testing (as shown by the results in Tables 4 and 5) was performed with independent TPA strings processing. As can be seen from the lockup times in Table 5, the independent strings were not simultaneously initialized for each run. However, the signal levels associated with each run were set for that particular run. The SNR values shown in Table 4 are the result of the signal level shown.

The lock-up times and sequential decoder statistics are additional and significant data points for the strong signal level telemetry processing function.

An additional telemetry threshold test (results not included) was performed to demonstrate the dual telemetry channel single sequential decoder operation of the software. Preliminary indications were that operation was satisfactory. A more thorough and rigorous testing of this capability is planned during the DSN PV'78 multisignal compatibility test scheduled for a later date.

D. DSN PV'78 Small Probes (1, 2 and 3), 14–17 November 1977

1. **Test objectives.** The objectives of these tests were to establish telecommunications design compatibility between the DSN and the Pioneer Venus Orbiter in the areas of radio frequency acquisition and telemetry at strong signal levels. Specifically, the objectives were as follows:

(1) Determine telemetry processing degradation characteristics due to wind-shear perturbations during the planetary atmospheric entry sequence.

(2) Determine telemetry processing degradation characteristics due to pyrotechnic firing perturbations during the planetary entry sequence.

(3) Verify small probe ultra-stable oscillator frequency stability during initial RF turn on sequence.

(4) Verify the DSN Telemetry System capability to receive and process coded telemetry.

(5) Verify the DSN capability to perform predetection recording and playback on actual PV'78 Small Probe RF signals and telemetry data.

Test (4) was performed in accordance with JPL Document 810-8, Rev. B, *Deep Space Network/Flight Project Interface Compatibility Test Design Handbook.* Tests (1), (2), (3) and (5) were special tests to determine those characteristics described. The test configuration for the small probe tests is shown in Fig. 2.

2. Test conditions. The Pioneer Venus Small Probe Spacecraft were located at the Hughes Aircraft Company, El Segundo, California. The Radio Frequency Subsystems were configured as S-band as follows: Small Probe 1, 10-W transmitter (2292.282020 MHz); Small Probe 2, 10-W transmitter (2292.437720 MHz); and Small Probe 3, 10-W transmitter (2291.552280 MHz).

The DSN, as represented by CTA 21, was configured to simulate a Pioneer Venus Flight Project committed Deep Space Station. The ground hardware included the Block III and IV Receiver-Exciter Subsystems and the Mark III Data Subsystem for telemetry and command.

The S-band RF link between the DSN and the PV'78 Orbiter Spacecraft was an open RF link between the Hughes Aircraft Company, El Segundo, California and the Jet Propulsion Laboratory, Pasadena, California. A description of the link as provided in Appendix 1. The link had been previously tested for amplitude and phase stability and was found to be acceptable for strong signal testing.

The DSN software provided at CTA 21 was the Pioneer Venus prereleased operational software for telemetry and command.

3. Test results. Tables 6 through 10, DSN PV'78 small probe telemetry tests, provide a listing of test results for verifying the DSN Telemetry System capability to receive and process coded telemetry. Tables 12 through 14 provide data relative to the spacecraft frequency stability measurements. Table 14 provides data showing results of the predetection recording playback.

a. Windshear and pyrotechnic firing. These tests were simulated by applying pulsed signals to the PIN modulator attenuator in the downlink (Figs. 2 and 3). These pulses suppressed the downlink signal level to simulate conditions expected to be encountered during firing of the spacecraft pyrotechnics and the windgust perturbations during descent through the Venusian atmosphere. However, these pulses were completely masked by variations that were encountered over the HAC-JPL open RF air link and therefore no conclusions regarding their contribution to telemetry processing degradation could be discerned.

In further efforts to verify possible degradation because of windshear and pyro firing effects, a closed loop test, internal to CTA 21, was performed off line following regularly scheduled testing with the small probe spacecrafts.

Utilizing the DSN test transmitter, modulated with PV'78 type data from the simulation conversion assembly, these tests were again performed. For the windshear test, the following conclusion was made:

(1) Only when the DSN receiver was operating at threshold conditions was there any effect noted.

(2) At threshold, one telemetry data frame was deleted, for each windgust.

(3) Assuming one or two windgusts and threshold conditions at 30-km altitude and at the surface, the deletion rate did not change significantly.

For the pyro firing test, the following was observed:

(1) No telemetry system loss of lock was observed.

(2) No deleted frames.

b. Small probe ultra-stable oscillator frequency stability. Performance of these tests was satisfactory, the data compares closely with unit data measured by HAC. The systems performance test software was a valuable tool in the successful conduction of this test.

c. Telemetry system processing test (planetary entry sequence). This test was conducted by simulating events scheduled to be encountered during the small probe-entry-to-impact phase of the mission. Each small probe spacecraft was run through this sequence two times, with CTA 21 processing the telemetry data and producing a Digital Original Data Record (DODR). The DODR's will be converted to a different format and provided to the PMOCC for later processing and training purposes.

Each run was recorded (Tables 6 through 10) and this data provides DSN system lock times as well as the Sequential Decoder statistical data. A shift in the CTA 21 Receiver AGC

curve occurred during the Small Probe No. 3, Run No. 2 test, but post calibrations validated the test data. Data readout (received power, P_c) shown in Table 11 have been adjusted to reflect this change.

d. Predetection recording and playback. The predetection recording scheme, to be utilized during the actual mission, was exercised during these tests to demonstrate that the system was compatible with mission requirements. Recordings were made during those tests described in paragraph *c* above. Following those tests, selected playbacks produced the results that are shown in Table 15. As can be seen, the DSN predetection recording technique appears to function within expectations.

Table 1. DSN-Pioneer Venus prototype spacecraft telecommunications compatibility test

Test date	Test title	DSN mode	Spacecraft TM mode	Test conditions	Criteria	Performance	Test, min
4/6/77, Test 1	Receiver rest frequency	002$\frac{3}{6}$00		UL signal level: -90.0 dBm UL frequency at best lock: 2114.327016 MHz	Acquire spacecraft receiver	Acquired	26
		002$\frac{3}{6}$00		UL signal level: -140.0 dBm Offset = +2 kHz at S-band Ramp rate = 10 Hz/s		Acquired at 2114.327280 MHz	
				UL signal level: -140.0 dBm Offset = -2kHz at S-band Ramp rate = 10 Hz/s		Acquired at 2114.326752 MHz	
4/5/77, Test 2	Receiver tracking	002$\frac{3}{6}$00		UL signal level: -140.0 dBm UL center frequency: 2114.316720 MHz DL center frequency: 2296.090560 MHz Ramp rate: 55 Hz/s at S-band	Ramp spacecraft receiver to ±67.5 kHz offset and record SPE/AGC at increments of 2.5 kHz	Satisfactory	104

Table 1 (contd)

Test date	Test title	DSN mode	Spacecraft TM mode	Test conditions	Criteria	Performance	Test, min
4/5/77, Test 3	Spacecraft receiver threshold	$002\frac{3}{6}00$		S-band UL frequency: 2114.316864 MHz S-band DL frequency: 2296.090740 MHz	<-152.0 dBm	-153.0 dBm	22
4/5/77, Tests 4 and 7 combined	RF & telemetry spectrum analysis	000300	Low mod index 16 bps	S-band DL frequency: 2296.093040 MHz Spacecraft VCO mode	Observe presence of unpredicted spectral components	None observed	197
		002300		S-band DL frequency: 2296.087160 MHz S-band UL frequency: 2114.313552 MHz		None observed	
		000300		S-band DL frequency: 2296.073700 MHz Auxiliary oscillator mode		None observed	
		000300	High mod index 64 bps	S-band DL frequency: 2296.073700 MHz auxilliary oscillator mode		None observed	
		002300		S-band DL frequency: 2296.092240 MHz		None observed	

Table 1 (contd)

Test data	Test title	DSN mode	Spacecraft TM mode	Test conditions	Criteria	Performance	Time, min
4/5/77, Tests 4 and 7 combined	RF & telemetry spectrum analysis (cont'd)			S-band UL frequency: 2114.318256 MHz			
4/5/77, Test 5A	RF downlink threshold (non-coherent)	000300	2048 bps coded	S-band DL frequency: 2296.106080 MHz	-159.0 ± 1.0 dBm	-158.0 dBm	2.0
4/10/77, Test 5B	Telemetry downlink threshold (non-coherent)	000312		16-bps coded Lo mod index	ST_b/N_o = 8.39 dB DLR = 10^{-2}	Frame deletion rate = 0.0% SER = 1.2%	
				64-bps coded Hi mod index	ST_b/N_o = 6.98 dB DLR = 10^{-2}	Frame deletion rate = 0.965% SER = 2.8%	
4/5/77, Test 6A	RF downlink threshold (coherent) S- and X-band	$002\frac{3}{6}00$		S-band UL frequency: 2114.321328 MHz			50
				S-band DL frequency: 2296.095560 MHz	-159.0 ± 1.0 dBm	-158.0 dBm	
				X-band DL frequency: 8419.017220 MHz	-151.0 ± 1.0 dBm	-150.5 dBm	
4/12/77, Test 6B	Telemetry downlink threshold (coherent)	002312		8.0 bps coded Low mod index	ST_b/N_o = 8.5 dB DLR = 10^{-2}	Frame deletion rate = 0.0% SER = 1.0%	

Table 1 (contd)

Test date	Test title	DSN mode	Spacecraft TM mode	Test conditions	Criteria	Performance	Time, min
4/12/77, Test 6B	Telemetry downlink threshold (coherent) (cont'd)	002312		128.0-bps coded Hi mod index Y-factor = 8.92 dB Pad = 19.98 dB	ST_b/N_o = 6.6 dB DLR = 10^{-3}	Frame deletion rate = 1.2% SER = 3.2%	
				170.66-bps coded Hi mod index Y-factor = 10.66 dB Pad = 19.98 dB	ST_b/N_o = 6.3 dB DLR = 10^{-3}	Frame deletion rate = 0.26% SER = 3.2%	
				170.66-bps uncoded Hi mod index Y-factor = 12.2 dB Pad = 19.98 dB ST_b/N_o = 8.45 dB	SSA SNR = 7.0 dB	SSA SNR = 7.0 dB	
				256.0-bps coded Hi mod index Y-factor = 11.27 dB Pad = 19.98 dB	ST_b/N_o = 6.1 dB DLR = 10^{-3}	Frame deletion rate = 0.56% SER = 2.3%	
				341.33 uncoded Hi mod index Y-factor = 14.29 dB Pad = 19.98 dB ST_b/N_o = 8.15 dB	SSA SNR = 7.0 dB	SSA SNR = 6.85 dB	
				341.33 coded Hi mod index Y-factor = 12.25 dB Pad = 19.98 dB	ST_b/N_o = 6.0 dB DLR = 10^{-3}	Frame deletion rate = 0.965% SER = 4.5%	

Table 1 (contd)

Test date	Test title	DSN mode	Spacecraft TM mode	Test conditions	Criteria	Performance	Time, min
4/12/77, Test 6B	Telemetry downlink threshold (coherent) (cont'd)	002312		512.0-bps coded Hi mod index Y-factor = 15.18 dB Pad = 19.98 dB	ST_b/N_o = 5.8 dB DLR = 10^{-3}	Frame deletion rate = 0.504% SER = 4.3%	
				682.66 uncoded Hi mod index Y-factor + 7.92 dB Pad = 10.34 dB ST_b/N_o = 7.8 dB	SSA SNR = 7.0 dB	SSA SNR = 7.08 dB	
				682.66 coded Hi mod index Y-factor - 7.83 dB Pad = 10.34 dB	ST_b/N_o = 5.7 dB DLR = 10^{-3}	Frame deletion rate = 0.45% SER = 4.3%	
				1024.0 coded Hi mod index Y-factor = 8.8 dB Pad = 10.34 dB	ST_b/N_o = 5.5 dB DLR = 10^{-3}	Frame deletion rate = 0.492% SER = 5.2%	
				2048.0 uncoded Hi mod index Y-factor = 12.2 dB Pad = 10.34 dB ST_b/N_o = 7.8 dB	SSA SNR = 7.0 dB	SSA SNR = 6.82 dB	
				2048.0 coded Hi mod index Y-factor = 13.4 dB Pad = 10.34 dB	ST_b/N_o = 4.8 dB DLR = 10^{-3}	Frame deletion rate = 0.847% SER = 5.6%	

Table 1 (contd)

Test date	Test title	DSN mode	Spacecraft TM mode	Test conditions	Criteria	Performance	Time, min
4/6/77, Test 8	Residual carrier phase jitter	000$\frac{3}{4}$00		S-band DL frequency: 2296.102820 MHz (one-way)	6.6 deg RMS	2.69 deg RMS	100
				DL signal level: -80.0 dBm			
				UL signal level: -90.0 dBm			
				X-band DL frequency: 8419.043740 MHz (one-way)	24.2 deg RMS	7.39 deg RMS	
				DL signal level: -100.0 dBm			
				S-band UL frequency: 2114.321088 MHz	5.8 deg RMS	1.84 deg RMS	
				S-band DL frequency: 2296.095320 MHz			
				DL signal level: -80.00 dBm			
				UL signal level: -90.0 dBm			
				X-band DL frequency: 8419.016370 MHz	21.3 deg RMS	6.21 deg RMS	
				DL signal level: -100.0 dBm			

Table 1 (contd)

Test date	Test title	DSN mode	Spacecraft TM mode	Test conditions	Criteria	Performance	Time, min
4/6/77, Test 9	Subcarrier frequency and phase jitter	$000\frac{31}{42}0$		S-band DL frequency: 2296.117780 MHz DL signal level: -80.0 dBm	Time period variation shall be less than 0.5% over 10 consecutive square wave periods	1.46 deg RMS Subcarrier frequency: 16384.0 Hz	35

Table 2. Deep Space Network/Pioneer Venus Orbiter telecommunications compatibility test data

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/1/77, Test 1	Receiver rest frequency	002300	Transmitter 1 Receiver 1	UL signal level: -131.6 dBm UL frequency at best lock: 2112.290352 MHz	Acquire spacecraft receiver		
				Offset = +5 kHz at S-band Ramp rate = 15 Hz/s at S-band		Acquired at 2112.290352 MHz VCO temp: 28.9°C	
				Offset = -5 kHz at S-band Ramp rate = 15 Hz/s at S-band		Acquired at 2112.290352 MHz VCO temp: 29.86°C	
		002300	Transmitter 2 Receiver 2	UL signal level: -131.6 dBm UL frequency at best lock: 2112.628056 MHz			
				Offset = +5 kHz at S-band Ramp rate = 15 Hz/s at S-band		Acquired at 2112.627840 MHz VCO temp: 28.81°C	
				Offset = -5 kHz at S-band Ramp rate = 15 Hz/s at S-band		Acquired at 2112.628272 MHz VCO temp: 28.81°C	

Table 2 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/1/77, Test 2	Receiver tracking	002300	Transmitter 1 Receiver 1	UL signal level: -131.6 dBm UL center frequency: 2112.288624 MHz VCO temp: 31.78°C	Spacecraft receiver maintain phase lock to UL signal while ramping ±100 kHz	Satisfactory	2 h 57 min
				DL center frequency: 2293.888180 MHz	About center frequency		
				Ramp rate: 100 Hz/s at S-band			
		002300	Transmitter 2 Receiver 2	UL signal level: -131.4 dBm UL center frequency: 2112.629808 MHz VCO temp: 29.8°C		Satisfactory	
				DL center frequency: 2294.258615 MHz			
				Ramp rate: 100 Hz/s at S-band			
11/3/77, Test 3	Telemetry processing	001411	Transmitter 1 Receiver 1 Telemetry processor 1	UL frequency at S-band: 2112.290304 MHz UL signal level: -125.5 dBm DL frequency at S-band: 2293.889984 MHz	Process telemetry		

Table 2 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/3/77, Test 3	Telemetry processing (contd)			Subcarrier frequency: 16.384 kHz			
				Bit rate: 8.0 bps coded data Mod index: 37.2 deg DL signal level: -142.4 dBm		SSA lock: 328 s SEQ lock: 621 s SSA SNR: 4.75 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.6442%	
				Bit rate: 16.0 bps coded data Mod index: 37.2 deg DL signal level: -138.0 dBm		SSA lock: 161 s SEQ lock: 363 s SSA SNR: 7.27 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.0925%	
				Bit rate: 32.0 bps coded data Mod index: 37.2 deg DL signal level: -136.2 dBm		SSA lock: 321 s SEQ lock: 354 s SSA SNR: 5.83 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.415%	
11/3/77, Test 3	Telemetry processing	001411	Transmitter 1 Receiver 1 Telemetry processor 1	Bit rate: 64.0 bps coded data Mod index: 67.6 deg DL signal level: -144.0 dBm	Process telemetry	SSA lock: 64 s SEQ lock: 98 s SSA SNR: 7.13 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.109%	
				Bit rate: 128.0 bps coded data Mod index: 67.6 deg DL signal level: -141.5 dBm		SSA lock: 49 s SEQ lock: 64 s SSA SNR: 4.81 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.921%	

Table 2 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/3/77, Test 3	Telemetry processing (cont'd)			Bit rate: 170.66 bps coded data Mod index: 67.6 deg DL signal level: -139.5 dBm		SSA lock: 90 s SEQ lock: 45 s SSA SNR: 5.64 dB SEQ DLR: 0.00% SEQ ACB: 1.003 SEQ SER: 0.3023%	
				Bit rate: 256.0 coded data Mod index: 67.6 deg DL signal level: -137.0 dBm		SSA lock: 84 s SEQ lock: 59 s SSA SNR: 8.21 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.0305%	
				Bit rate: 256.0 bps uncoded Mod index: 67.6 deg DL signal level: 133.0 dBm		SSA lock: 49 s SSA SNR: 13.38 dB	
				Bit rate: 341.33 bps coded data Mod index: 67.6 deg DL signal level: -136.3 dBm		SSA lock: 48 s SEQ lock: 36 s SSA SNR: 7.07 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.0652%	
				Bit rate: 512 bps coded data Mod index: 67.6 deg DL signal level: -134.1 dBm		SSA lock: 199 s SEQ lock: 149 s SSA SNR: 7.53 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.0202%	

Table 2 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/3/77, Test 3	Telemetry processing (cont'd)			Bit rate: 682.66 bps coded data Mod index: 67.6 deg DL signal level: -132.4 dBm		SSA lock: 43 s SEQ lock: 5 s SSA SNR: 8.05 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.0163%	
				Bit rate: 682.66 bps uncoded Mod index: 67.6 deg DL signal level: -128.4 dBm		SSA lock: 52 s SSA SNR: 12.72 dB	
				Bit rate: 1024.0 bps coded data Mod index: 67.6 deg DL signal level: -131.0 dBm		SSA lock: 17 s SEQ lock: 9 s SSA SNR: 8.13 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.0154%	
				Bit rate: 2048.0 bps coded data Mod index: 67.6 deg DL signal level: -124.7 dBm		SSA lock: 64 s SEQ lock: 7 s SSA SNR: 10.81 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.0%	
				Bit rate: 2048.0 bps uncoded Mod index: 67.6 deg DL signal level: -125.0 dBm		SSA lock: 25 s SSA SNR: 12.53	

Table 2 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/4/77, Test 3	Telemetry processing	001411	Transmitter 2 Receiver 2 Telemetry processor 2	UL frequency S-band: 2112.629760 MHz UL signal level: -125.2 dBm DL frequency S-band: 2294.258528 MHz Subcarrier frequency: 16.384 kHz Bit rate: 8.0 bps coded data Mod index: 37.2 deg TLM format: PSB DL signal level: -142.4 dBm	Process telemetry	SSA lock: 321 s SEQ lock: 602 s SSA SNR: 5.22 dB SEQ DLR: 0.00% SEQ ACB: 1.0008 SEQ SER: 0.45%	
				Bit rate: 16.0 bps coded data Mod index: 37.2 deg TLM format: PSD DL signal level: -140.0 dBm		SSA lock: 160 s SEQ lock: 342 s SSA SNR: 3.66 dB SEQ DLR: 0.00% SEQ ACB: 1.025 SEQ SER: 1.859%	
				Bit rate: 32.0 bps coded data Mod index: 37.2 deg TLM format: PSE DL signal level: -136.4 dBm		SSA lock: 129 s SEQ lock: 176 s SSA SNR: 6.98 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.07%	

Table 2 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/4/77, Test 3	Telemetry processing (cont'd)			Bit rate: 64.0 bps coded data Mod index: 67.6 deg TLM format: ASB DL signal level: -144.3 dBm		SSA lock: 83 s SEQ lock: 165 s SSA SNR: 4.78 dB SEQ DLR: 0.00% SEQ ACB: 1.04 SEQ SER: 0.70%	
				Bit rate: 128.0 bps coded data Mod index: 67.6 deg TLM format: ASA DL signal level, -141.2		SSA lock: 49 s SEQ lock: 58 s SSA SNR: 4.6 dB SEQ DLR: 0.00% SEQ ACB: 1.007 SEQ SER: 0.72%	
				Bit rate: 170.66 bps coded data Mod index: 67.6 deg TLM format: PSC DL signal level: -137.7 dBm		SSA lock: 192 s SEQ lock: 90 s SSA SNR: 6.85 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.118%	
				Bit rate: 256.0 bps coded data Mod index: 67.6 deg TLM format: PSA DL signal level: -137.6 dBm		SSA lock: 64 s SEQ lock: 40 s SSA SNR: 6.61 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.13%	
				Bit rate: 341.33 bps coded data Mod index: 67.6 deg TLM format: ENG DL signal level: -136.7 dBm		SSA lock: 48 s SEQ lock: 35 s SSA SNR: 7.04 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.06%	

Table 2 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/4/77, Test 3	Telemetry processing (cont'd)			Bit rate: 512.0 bps coded data Mod index: 67.6 deg TLM format: ENG DL signal level: -134.8 dBm		SSA lock: 32 s SEQ lock: 74 s SSA SNR: 6.63 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.154%	
				Bit rate: 682.66 bps coded data Mod index: 67.6 deg TLM format: ENG DL signal level: -131.8 dBm		SSA lock: 27 s SEQ lock: 13 s SSA SNR: 7.7 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.02%	
				Bit rate: 682.66 bps uncoded Mod index: 67.6 deg TLM format: ENG DL signal level: -130.3 dBm		SSA lock: 49 s SSA SNR: 11.32 dB	
				Bit rate: 1024.0 bps coded data Mod index: 67.6 deg TLM format: ENG DL signal level: -131.0 dBm		SSA lock: 33 s SEQ lock: 9 s SSA SNR: 6.44 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.13%	
				Bit rate: 2048.0 bps coded data Mod index: 67.6 deg TLM format: ENG DL signal level: -127.6 dBm		SSA lock: 129 s SEQ lock: 7 s SSA SNR: 8.48 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.014%	

Table 2 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/4/77, Test 3	Telemetry processing (cont'd)			Bit rate: 2048.0 bps uncoded Mod index: 67.6 deg TLM format: ENG DL signal level: -124.4 dBm		SSA lock: 33 s SSA SNR: 11.26 dB	
11/2/77, Test 4	Command	021100	Transmitter 1 Receiver 1 Command processor 1	UL S-band frequency: 2112.290304 MHz UL signal level: -127.1 dBm	Spacecraft process and verify command execution		
				Single timed commands			
				Preamble with time: AAAAAAAAAAAA8			
				Single timed commands			
				CMD 1: EB04D4000E5D8		Verified	
				Preamble with time: AAAAAAAAAAAA8			
				CMD 5: EB04D4000E988		Verified	
				Preamble with time: AAAAAAAAAAAA8			
				CMD 2: EB04D68011D28		Verified	

Table 2 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/2/77, Test 4	Command (cont'd)			Preamble with time: AAAAAAAAAAAA8			
				CMD 3: EB04D6800BEC8		Verified	
				Preamble with time: AAAAAAAAAAAA8			
				CMD 4: EB04D6801D828		Verified	
				Single untimed commands			
				CMD 1: EB04D000E5D8		Verified	
				CMD 5: EB04D4000E988		Verified	
				CMD 2: EB04D68011D28		Verified	
				CMD 3: EB04D6800BEC8		Verified	
				CMD 4: EB04D6801D828		Verified	
				Spacecraft command timing check			
				CMD 11: EB04D700023E8		Verified	
				CMD 12: EB04D7000E6E8		Verified	
				CMD 13: EB04D7000DBF8		Verified	
				CMD 16: EB04D70001EF8		Verified	

Table 2 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/2/77, Test 4	Command (cont'd)			CMD 14: EB04CF2582E68		Verified	
				CMD 15: EB04CF4B02F78		Verified	
		021100	Transmitter 2 Receiver 2 Command processor 2	UL S-band frequency: 2112.628056 MHz			
				UL signal level: -125.2 dBm			
				Contiguous timed commands			
				Preamble with time: AAAAAAAAAAAA8			
				CMD 6: EB04D4000E178		Verified	
				CMD 7: EB05DA8011918		Verified	
				CMD 8: EB05DA800BAF8		Verified	
				CMD 9: EB05DA801DC18		Verified	
				CMD 10: EB05D4000ED28		Verified	
11/1/77	Special predetection recording and playback	001211	Transmitter 1 Receiver 1 Telemetry processor 1	Bit rate: 16.0 bps coded data Mod index: 32.7 deg DL signal level: -138.0 dBm	Maximum of dB between real-time SNR and playback SNR	Real-time SSA SNR: 7.27 dB Playback SSA SNR: 7.13 dB	

Table 2 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/1/77	Special predetection recording and playback (cont'd)			Bit rate: 64.0 bps coded data Mod index: 67.6 deg DL signal level: -144.0 dBm		Real-time SSA SNR: 7.13 dB Playback SSA SNR: 7.24 dB	
				Bit rate: 256.0 bps coded data Mod index: 67.6 deg DL signal level: -137.0 dBm		Real-time SSA SNR: 8.21 dB Playback SSA SNR: 7.95 dB	

Table 3. Deep Space Network PV'78 bus spacecraft telecommunications compatibility test data

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/8/77, Test 1	Receiver rest frequency	002300	Transmitter 1 Receiver 1	UL signal level: -138.0 dBm UL frequency at best lock: 2110.583352 MHz	Acquire spacecraft receiver		2 h 3 min
				Offset = +5 kHz at S-band Ramp rate = 15 Hz/s at S-band		Acquired at 2110.583568 MHz VCO temp: 28.9°C	
				Offset = -5 kHz at S-band Ramp rate = 15 Hz/s at S-band		Acquired at 2110.583136 MHz VCO temp: 28.9°C	
		002300	Transmitter 2 Receiver 2	UL signal level: -137.8 dBm UL frequency at best lock: 2111.268744 MHz			
				Offset = +5 kHz at S-band Ramp rate = 15 Hz/s at S-band		Acquired at 2111.268912 MHz VCO temp: 27.63°C	
				Offset = -5 kHz at S-band Ramp rate = 15 Hz/s at S-band		Acquired at 2111.268576 MHz VCO temp: 27.63°C	

Table 3 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/8/77, Test 2	Receiver tracking	002300	Transmitter 1 Receiver 1	UL signal level: -139.3 dBm UL center frequency: 2110.583040 MHz	Spacecraft receiver maintain phase lock to UL signal while ramping ±100 kHz	Satisfactory	2 h 43 min
				DL center frequency: 2292.035960 MHz Ramp rate: 100 Hz/s at S-band	About center frequency		
		002300	Transmitter 2 Receiver 2	UL signal level: -137.2 dBm UL center frequency: 2111.267712 MHz		Satisfactory	
				DL center frequency: 2292.779520 MHz			
				Ramp rate: 100 Hz/s at S-band			
11/10/77, Test 3	Telemetry processing	001411	Transmitter 1 Receiver 1 Telemetry processor 2	UL frequency S-band: 2110.583040 MHz UL signal level: -131.5 dBm DL frequency S-band: 2292.02400 MHz	Process telemetry		

Table 3 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/10/77, Test 3	Telemetry processing (contd)			Subcarrier frequency: 16.384 kHz			
				Bit rate: 8.0 bps coded data Mod index: 37.2 deg DL signal level: -147.5 dBm TLM format: ENG		SSA lock: 320 s SEQ lock: 569 s SSA SNR: 4.41 dB SEQ DLR: 0.00% SEQ ACB: 1.0092 SEQ SER: 1.21%	
				Bit rate: 16.0 bps coded data Mod index: 37.2 deg DL signal level: -144.8 dBm TLM format: ENG		SSA lock: 576 s SEQ lock: 696 s SSA SNR: 2.05 dB SEQ DLR: 0.00% SEQ ACB: 1.113 SEQ SER: 3.01%	
				Bit rate: 128.0 bps coded data Mod index: 67.6 deg DL signal level: -146.7 dBm TLM format: ENG		SSA lock: 145 s SEQ lock: 161 s SSA SNR: 2.74 dB SEQ DLR: 0.00% SEQ ACB: 1.0785 SEQ SER: 2.398%	
				Bit rate: 170.66 bps coded data Mod index: 67.6 deg DL signal level: -145.1 dBm TLM format: ENG		SSA lock: 193 s SEQ lock: 87 s SSA SNR: 3.18 dB SEQ DLR: 0.00% SEQ ACB: 1.062 SEQ SER: 1.95%	

Table 3 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/10/77, Test 3	Telemetry processing (contd)			Bit Rate: 256.0 coded data Mod Index: 67.6 deg DL signal level: -143.9 dBm TLM format: ENG		SSA lock: 129 s SEQ lock: 39 s SSA SNR: 1.36 dB SEQ DLR: 0.00% SEQ ACB: 1.2 SEQ SER: 4.07%	
				Bit rate: 256.0 bps uncoded Mod index: 67.6 deg DL signal level: -142.8 dBm TLM format: ENG		SSA lock 48 s SSA SNR: 6.14 dB	
				Bit rate: 341.33 bps coded data Mod index: 67.6 deg DL signal level: -140.5 dBm TLM format: ENG		SSA lock: 48 s SEQ lock: 34 s SSA SNR: 4.39 dB SEQ DLR: 0.00% SEQ ACB: 1.004 SEQ SER: 0.849%	
				Bit rate: 341.33 bps uncoded Mod index: 67.6 deg DL signal level: -139.9 dBm TLM format: ENG		SSA lock: 96 s SSA SNR: 11.02 dB	

Table 3 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/10/77, Test 3	Telemetry processing (contd)			Bit rate: 512 bps coded data Mod Index: 67.6 deg DL signal level: -139.8 dBm TLM format: ENG		SSA lock: 65 s SEQ lock 12 s SSA SNR: 2.98 dB SEQ DLR: 0.00% SEQ ACB: 1.033 SEQ SER: 2.04%	
				Bit rate: 682.66 bps coded data Mod index: 67.6 deg DL signal level: -138.4 dBm TLM format: ENG		SSA lock: 52 s SEQ lock: 13 s SSA SNR: 3.59 dB SEQ DLR: 0.520% SEQ ACB: 2.603 SEQ SER: 1.54%	
				Bit rate: 1024.0 bps coded data Mod index: 67.6 deg DL signal level: -137.5 dBm TLM format: ENG		SSA lock: 34 s SEQ lock: 27 s SSA SNR: 2.61 dB SEQ DLR: 0.52% SEQ ACB: 2.652 SEQ SER: 2.53%	
				Bit rate: 1024.0 bps uncoded Mod index: 67.6 deg DL signal level: -138.0 dBm TLM format: ENG		SSA lock: 32 s SSA SNR: 7.96 dB	

Table 3 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/10/77, Test 3	Telemetry processing (contd)			Bit rate: 2048.0 bps coded data Mod index: 67.6 deg DL signal level: -132.5 dBm TLM format: ENG		SSA lock: 64 s SEQ lock: 6 s SSA SNR: 6.14 dB SEQ DLR: 0.00% SEQ ACB: 1.113 SEQ SER: 3.01%	
11/9/77, Test 3	Telemetry processing	001411	Transmitter 2 Receiver 2 Telemetry processor 1	UL freq. S-band: 2111.265696 MHz UL signal level: -130.9 dBm DL freq. S-band: 2292.777344 MHz Subcarrier frequency: 16.384 kHz	Process telemetry		
				Bit rate: 8.0 bps coded data Mod index: 37.2 deg TLM format: ENG DL signal level: -147.0 dBm		SSA lock: 315 s SEQ lock: 653 s SSA SNR: 2.6 dB SEQ DLR: 0.00% SEQ ACB: 1.146 SEQ SER: 1.994%	
				Bit rate: 16.0 bps coded data Mod index: 37.2 deg TLM format: ENG DL signal level: -144.3 dBm		SSA lock: 160 s SEQ lock: 308 s SSA SNR: 5.22 dB SEQ DLR: 0.00% SEQ ACB: 1.0007 SEQ SER: 0.39%	

Table 3 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/9/77, Test 3	Telemetry processing (contd)			Bit rate: 32.0 bps coded data Mod index: 37.2 deg TLM format: ENG DL signal level: -141.3 dBm		SSA lock: 326 s SEQ lock: 377 s SSA SNR: 4.65 dB SEQ DLR: 0.00% SEQ ACB: 1.0026 SEQ SER: 0.64%	
				Bit rate: 64.0 bps coded data Mod index: 67.6 deg TLM format: ENG DL signal level: -149.0 dBm		SSA lock: 58 s SEQ lock: 34 s SSA SNR: 4.01 dB SEQ DLR: 0.00% SEQ ACB: 1.028 SEQ SER: 0.967%	
				Bit rate: 128.0 bps coded data Mod index: 67.6 deg TLM format: ENG DL signal level: -146.7 dBm		SSA lock: 49 s SEQ lock: 58 s SSA SNR: 2.02 dB SEQ DLR: 0.00% SEQ ACB: 1.039 SEQ SER: 1.846%	
				Bit rate: 170.66 bps coded data Mod index: 67.6° TLM format: ENG DL signal level: -144.6 dBm		SSA lock 96 s SEQ lock: 51 s SSA SNR: 4.84 dB SEQ DLR: 0.00% SEQ ACB: 1.01 SEQ SER: 0.56%	
11/8/77				Bit rate: 256.0 bps coded data Mod index: 67.6 deg TLM format: ENG DL signal level: -142.6 dBm		SSA lock: 65 s SEQ lock: 41 s SSA SNR: 4.91 dB SEQ DLR: 0.00% SEQ ACB: 1.0025 SEQ SER: 0.5%	

Table 3 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/9/77, Test 3	Telemetry processing (contd)			Bit rate: 256 bps uncoded Mod index: 67.6 deg TLM format: ENG DL signal level: -139.6 dBm		SSA lock: 59 s SSA SNR: 11.05 dB	
				Bit rate: 341.33 bps coded data Mod index: 67.6 deg TLM format: ENG DL signal level: -140.9 dBm		SSA lock: 144 s SEQ lock: 77 s SSA SNR: 5.73 dB SEQ DLR: 0.00% SEQ ACB: 1.00008 SEQ SER: 0.243%	
				Bit rate: 512.0 bps coded data Mod index: 67.6 deg TLM format: ENG DL signal level: -139.6 dBm		SSA lock: 65 s SEQ lock: 14 s SSA SNR: 5.99 dB SEQ DLR: 0.00% SEQ ACB: 1.0004 SEQ SER: 0.217%	
				Bit rate: 682.66 bps coded data Mod index: 67.6 deg TLM format: ENG DL signal level: -137.4 dBm		SSA lock: 25 s SEQ lock: 11 s SSA SNR: 7.2 dB SEQ DLR: 0.00% SEQ ACB: 1.000 SEQ SER: 0.046%	
				Bit rate: 682.66 bps uncoded Mod index: 67.6 deg TLM format: ENG DL signal level: -137.4 dBm		SSA Lock: 48 s SSA SNR: 9.8 dB	

Table 3 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/8/77, Test 3	Telemetry processing (contd)			Bit rate: 1024.0 bps coded data Mod index: 67.6 deg TLM format: ENG DL signal level: -135.0 dBm		SSA lock: 33 s SEQ lock: 8 s SSA SNR: 5.56 dB SEQ DLR: 0.00% SEQ ACB: 1.193 SEQ SER: 3.328%	
				Bit rate: 2048.0 bps coded data Mod index: 67.6 deg TLM format: ENG DL signal level: -135.0 dBm		SSA lock: 63 s SEQ lock: 7 s SSA SNR: 3.92 dB SEQ DLR: 0.00% SEQ ACB: 1.019 SEQ SER: 1.144	
11/9/77				Bit rate: 2048.0 bps uncoded Mod index: 67.6 deg TLM format: ENG DL signal level: -132.5 dBm		SSA lock: 17 s SSA SNR: 9.17 dB	
11/9/77, Test 4	Command	021100	Transmitter 2 Command processor 2	UL S-band frequency: 2110.582848 MHz UL signal level: -132.5 dBm Single timed commands Preamble with time: AAAAAAAAAAAA8	Spacecraft process and verify command execution		

Table 3 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/9/77, Test 4	Command (contd)			CMD 1: EB00D4000E3A8 Preamble with time: AAAAAAAAAAAA8		Verified	
				Single timed commands			
				CMD 5: EB00D8000EF68 Preamble with time: AAAAAAAAAAAA8		Verified	
				CMD 2: EB00D68011B58 Preamble with time: AAAAAAAAAAAA8		Verified	
				CMD 3: EB00D6800B8B8 Preamble with time: AAAAAAAAAAAA8		Verified	
				CMD 4: EB00D6801DE58		Verified	
				Single untimed command			
				CMD 1: EB00D4000E3A8		Verified	
				CMD 5: EB00D8000EF68		Verified	
				CMD 2: EB00D68011B58		Verified	
				CMD 3: EB00D6800B8B8		Verified	

Table 3 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/9/77, Test 4	Command (contd)			CMD 4: EB00D6801DE58		Verified	
				Contiguous timed commands			
				Preamble with time: AAAAAAAAAAAA8			
				CMD 6: EB01D4000E708		Verified	
				CMD 7: EB01DA8011F68		Verified	
				CMD 8: EB01DA800BC88		Verified	
				CMD 9: EB01DA801DA68		Verified	
				CMD 10: EB01D8000EBC8		Verified	
				Spacecraft command timing check			
11/9/77		021100	Transmitter 1 Receiver 1 Command processor 1	UL S-band frequency: 2111.265792 MHz			
				UL signal level: -131.0 dBm			
				CMD 11: EB00D70002598		Verified	
				CMD 12: EB00D7000E098		Verified	
				CMD 13: EB00D7000DD88		Verified	
				CMD 16: EB00D70001888		Verified	

Table 3 (contd)

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/9/77, Test 4	Command (contd)			CMD 14: EB00CF2582818		Verified	
				CMD 15: EB00CF1302C48		Verified	
11/10/77				Single untimed commands with leading HEX zero			
				CMD 1: 0EB00D4000E3A8		Verified	
				CMD 5: 0EB00D8000EF68		Verified	
				CMD 2: 0EB00D68011B58		Verified	
				CMD 3: 0EB00D6800B8B8		Verified	
				CMD 4: 0EB00D6801DE58		Verified	

Table 4. Deep Space Network/PV'78 bus-small probe telecommunications compatibility test data

Test date	Test title	DSN mode	Spacecraft RFS mode	Test conditions	Criteria	Performance	Time, min
11/11/77 Test 1	Dual Subcarrier	0004 11/22	Transmitter 2 Aux. oscillator 2 TLM processor 2	DL S-band frequency: 2292.034784 mHz Bus subcarrier Frequency: 16.384 kHz S.P. subcarrier Frequency: 4.096 kHz	Process telemetry	See below and Table 2	

	Bus		Small probe									
					Bus Sequential decoder			Small probe Sequential decoder				
P_c, dBm	Bit rate, coded data	Mod index, deg	Bit rate, coded data	Mod index, deg	DLR	ACB	SER	DLR	ACB	SER	Bus SSA SNR, dB	Small probe SSA SNR, dB
-141.0	8.0	37.2	16.0	44.7	0.00	1.000	0.00	0.00	1.000	0.00	9.98	12.86
-143.9	8.0	37.2	64.0	44.7	0.00	1.000	0.105	0.00	1.018	0.82	9.39	4.95
-141.0	16.0	37.2	64.0	44.7	0.00	1.000	0.036	0.00	1.000	0.00	8.16	13.25
-143.9	16.0	37.2	16.0	44.7	0.00	1.003	0.83	0.337	1.012	0.643	4.91	5.28
-143.9	16.0	37.2	64.0	58.7	0.00	1.006	0.938	0.00	1.043	0.529	5.09	6.01

Table 5. Telemetry lock-up times

	Bus		Small Probe		Bus			Small Probe		
P_c, dBm	Bit rate, coded data	Mod. index	Bit rate, coded data	Mod. index	TPA init h:min:s	SSA lock h:min:s	SEQ lock h:min:s	TPA init h:min:s	SSA lock h:min:s	SEQ lock h:min:s
-141.0	8.0	37.2	16.0	44.7	17:32:30	17:37:50	17:41:28	16:52:04	16:54:45	16:59:06
-143.9	8.0	37.2	64.0	44.7	17:32:30	17:37:50	17:41:28	18:05:24	18:06:26	18:06:55
-141.0	16.0	37.2	64.0	44.7	18:34:24	18:44:00	18:46:30	18:05:24	18:06:26	18:06:55
-143.9	16.0	37.2	16.0	44.7	18:34:24	18:44:00	18:46:30	19:09:34	19:12:15	19:13:44
-143.9	16.0	37.2	64.0	58.7	18:34:24	18:44:00	18:46:30	20:09:47	20:10:51	20:11:09

Table 6. DSN-PV'78 Small Probe No. 1 telemetry test, Run No. 1 (TPA No. 1, 2292.282020 MHz)

Event GMT	Bit rate mod. index, coded data	Received power, dBm P_T	P_c	P_D	RCVR lock (GMT)	SDA lock (GMT)	TPA init (GMT)	SSA lock (GMT)	SEQ lock (GMT)	SSA SNR, dB	SEQ SER, %	SEQ DLR %	SEQ ACB
E - 22 21:55:07	None	-136.86	-136.86	None	21 h 55 min 22 s								
E - 17 22:00:07	64 at 58.44 deg	-134.38	-140.00	-135.77		21 h 56 min 44 s	21 h 55 min 07 s	22 h 01 min 12 s	22 h 01 min 46 s	8.94	0.00	0.00	1.000
E = 0 22:17:07	64 at 58.44 deg	2-s Blackout -138.78	-144.4	-140.17	22 h 17 min 47 s	No out-of-lock indication		22 h 18 min 47 s	22 h 19 min 14 s	6.55	0.00	0.00	1.000
E + 16 22:33:07	16 at 44.7 deg	-139.63	-142.6	-142.69			22 h 33 min 27 s	22 h 36 min 09 s	22 h 38 min 59 s	10.2	0.00	0.00	1.000
E + 27 22:44:07	16 at 44.7 deg	-138.03	-141.0	-141.09						8.82	0.00	0.00	1.000
E + 38 22:55:07	16 at 44.7 deg	-140.93	-143.9	-143.99						8.13	0.00	0.00	1.000
E + 49 23:06:07	16 at 44.7 deg	-139.93	-142.9	-142.99						8.36	0.00	0.00	1.000

Table 6 (contd)

Event GMT	Bit rate mod. index, coded data	Received power, dBm P_T	P_c	P_D	RCVR lock (GMT)	SDA lock (GMT)	TPA init (GMT)	SSA lock (GMT)	SEQ lock (GMT)	SSA SNR, dB	SEQ SER, %	SEQ DLR %	SEQ ACB
E + 60	16 at												
23:17:07	44.7 deg	-139.83	-142.8	-142.89						7.43	0.048	0.00	1.000

Table 7. DSN-PV'78 Small Probe No. 1 telemetry test, Run No. 2 (TPA 1, at 2292.282020 MHz)

Event GMT	Bit rate mod. index, coded data	Received power, dBm P_T	P_C	P_D	RCVR lock	SDA lock	TPA init	SSA lock	SEQ lock	SSA SNR, dB	SEQ SER, %	SEQ DLR, %	SEQ ACB
E - 22					23 h 26 min								
23:24:05	None	-139.96	-139.96	None	08 s								
E - 17	64 at					23 h 29 min	23 h 29 min	23 h 31 min	23 h 31 min	5.05	0.55	0.00	1.0058
23:29:05	58.44 deg	-139.38	-145.0	-140.77									
E = 0	64 at	20-s Blackout			23 h 47 min	No out-of-lock indi-cation		23 h 47 min 15 s	No out-of-lock indi-cation	3.21	1.84	0.00	1.027
23:46:05	58.44 deg	-140.78	-146.4	-142.17	09 s								
E + 16	16 at						00 h 02 min		00 h 07 min	None	None	None	None
00:02:05	44.7 deg	-142.83	-145.8	-145.89			13 s	None	27 s				
E + 27	16 at						00 h 09 min	00 h 19 min	00 h 21 min				
00:13:05	44.7 deg	-144.53	-147.5	-147.59			47 s	24 s	51 s	3.21	1.84	0.00	1.027

Table 7 (contd)

Event GMT	Bit rate mod. index, coded data	Received power, dBm P_T	Received power, dBm P_c	Received power, dBm P_D	RCVR lock	SDA lock	TPA init	SSA lock	SEQ lock	SSA SNR, dB	SEQ SER, %	SEQ DLR, %	SEQ ACB
E + 38	16	-140.03	-143.0	-143.09									
	at		to										
00:24:05	44.7 deg	-148.03	-151.0	-151.09		Link variations were too excessive to be meaningful for 0.5 dB changes. Data were averaged for remainder of test.							
E + 49	16		-143.0										
	at		to	Same									
00:35:05	44.7 deg		-151.0							6.05	0.238	0.00	1.000
E + 60	16		-143.0										
	at	Same	to	Same									
00:46:05	44.7 deg		-151.0										

Table 8. DSN-PV'78 Small Probe No. 2 telemetry test, Run No. 1, TPA 1 at 2292.437720 MHz

Event GMT	Bit rate mod. index, coded data	Received power, dBm			RCVR lock	SDA lock	TPA init	SSA lock	SEQ lock	SSA SNR, dB	SEQ SER, %	SEQ DLR, %	SEQ ACB
		P_T	P_c	P_D									
E - 22					20 h 07 min								
20:04:05	None	-142.5	-142.5	None	27 s								
E - 17	64 at					20 h 09 min	20 h 09 min	20 h 11 min	20 h 11 min				
20:09:05	58.44 deg	-136.88	-142.5	-138.27		07 s	17 s	25 s	58 s	10.47	0.00	0.00	1.000
E = 0	64 at	20-s Blackout			20 h 26 min	No out-of-lock		20 h 27 min	No out-of-lock				
20:26:05	58.44 deg	-131.38	-137.0	-132.77	29 s	indication		25 s	indication	9.90	0.036	1.339	1.000
E + 16	16 at						20 h 42 min	20 h 44 min	20 h 46 min				
20:42:05	44.7 deg	-139.53	-142.5	-142.59			13 s	55 s	54 s	15.38	0.00	0.00	1.000
E + 27	16 at												
20:53:05	44.7 deg	-134.03	-137.0	-137.09						8.61	0.00	0.00	1.000
E + 38	16 at												
21:04:05	44.7 deg	-141.53	-144.5	-144.59						7.17	0.00	0.00	1.000
E + 49	16 at												
21:15:05	44.7 deg	-140.03	-143.0	-143.09						6.49	0.048	0.00	1.000

Table 8 (contd)

Event GMT	Bit rate mod. index, coded data	Received power, dBm			RCVR lock	SDA lock	TPA init	SSA lock	SEQ lock	SSA SNR, dB	SEQ SER, %	SEQ DLR, %	SEQ ACB
		P_T	P_c	P_D									
E + 60	16 at												
21:26:05	44.7 deg	-142.03	-145.0	-145.09						7.47	0.00	0.00	1.000

Table 9. DSN-PV'78 Small Probe No. 2 telemetry test, Run No. 2, TPA No. 1, at 2292.437720 MHz

Event GMT	Bit rate mod. index, coded data	Received power, dBm P_T	P_c	P_D	RCVR lock	SDA lock	TPA init	SSA lock	SEQ lock	SSA SNR, dB	SEQ SER, %	SEQ DLR, %	SEQ ACB
E - 22 20:22:50	None	None	None	None									
E - 17 20:27:50	64 at 58.44 deg	-139.68	-145.3	-141.07	20 h 28 min 22 s	20 h 28 m 59 s	20 h 29 min 04 s	20 h 30 min 09 s	20 h 30 min 35 s	4.8	0.695	0.00	1.000
E = 0 20:44:50	64 at 58.44 deg	20-s Blackout -140.18	-142.8	-142.89	20 h 48 min 04 s	No out-of-lock indication		20 h 49 min 53 s	No out-of-lock indication	4.76	1.11	0.00	1.118
E + 16 21:00:50	16 at 44.7 deg	-139.83	-142.8	-142.89			21 h 00 min 53 s	21 h 03 min 34 s	21 h 05 min 40 s	4.12	1.027	0.00	1.010
E + 27 21:11:50	16 at 44.7 deg	-144.03	-147.0	-147.09						4.17	1.108	0.00	1.010
E + 38 21:22:50	16 at 44.7 deg	-144.63	-147.6	-147.69						3.84	1.49	0.00	1.010
E + 49 21:33:50	16 at 44.7 deg	-144.93	-147.9	-147.99						2.97	2.47	0.00	1.085

Table 9 (contd)

Event GMT	Bit rate mod. index, coded data	Received power, dBm P_T	P_c	P_D	RCVR lock	SDA lock	TPA init	SSA lock	SEQ lock	SSA SNR, dB	SEQ SER, %	SEQ DLR, %	SEQ ACB
E + 60	16												
21:44:50	at 44.7 deg	-145.33	-148.3	-148.39						2.64	3.75	0.00	1.147

Table 10. DSN-PV'78 Small Probe No. 3 telemetry test, Run No. 1, TPA 1, at 2291.552280 MHz

Event GMT	Bit rate mod. index, coded data	Received power, dBm P_T	P_c	P_D	RCVR lock	SDA lock	TPA init	SSA lock	SEQ lock	SSA SNR, dB	SEQ SER, %	SEQ DLR, %	SEQ ACB
E - 22					16 h 48 min								
16:48:05	None	-142.5	-142.5	None	29 s								
E - 17	64 at					16 h 53 min	16 h 53 min	16 h 54 min	16 h 55 min				
16:53:05	58.44 deg	-136.88	-142.5	-138.27		19 s	23 s	27 s	02 s	9.03	0.00	0.00	1.000
E = 0	64 at	20-s Blackout			17 h 10 min	No out-of-lock		17 h 11 m	No out-of-lock				
17:10:10	58.44 deg	-136.38	-142.0	-137.77	32 s	indi-cation		31 s	indi-cation	7.95	0.146	1.293	1.000
E + 16	16 at						17 h 26 min	17 h 28 min	17 h 31 min				
17:26:05	44.7 deg	-140.53	-143.5	-143.59			14 s	55 s	27 s	10.53	0.00	0.00	1.000
E + 27	16 at												
17:37:05	44.7 deg	-139.03	-142.0	-142.09						8.44	0.00	0.00	1.000
E + 38	16 at												
17:48:05	44.7 deg	-139.03	-142.0	-142.09						8.15	0.00	0.00	1.000
E + 49	16 at												
17:59:05	44.7 deg	-140.53	-143.5	-143.59						8.55	0.00	0.00	1.000

Table 10 (contd)

Event GMT	Bit rate mod. index, coded data	Received power, dBm P_T	Received power, dBm P_c	Received power, dBm P_D	RCVR lock	SDA lock	TPA init	SSA lock	SEQ lock	SSA SNR, dB	SEQ SER, %	SEQ DLR, %	SEQ ACB
E + 60	16												
18:10:05	at 44.7 deg	-142.03	-145.0	-145.09						7.26	0.00	0.00	1.000

Table 11. DSN-PV'78 Small Probe No. 3 telemetry test, Run No. 2, TPA No. 1, at 2291.552360 MHz

Event GMT	Bit rate mod. index, coded data	Received power, dBm P_T	P_c	P_D	RCVR lock	SDA lock	TPA init	SSA lock	SEQ lock	SSA SNR, dB	SEQ SER, %	SEQ DLR, %	SEQ ACB
E - 22		-145.5	-145.5	None	16 h 56 min 18 s								
E - 17	64 at 58.44 deg	-140.38	-146.0	-141.77		17 h 01 min 15 s	17 h 01 min 17 s	17 h 02 min 53 s	17 h 02 min 54 s	0.95	4.10	0.00	1.338
E = 0	64 at 58.44 deg	20-s Blackout -142.38	-148.0	-143.77	17 h 18 min 28 s	No out-of-lock indication		17 h 19 min 26 s	No out-of-lock indication	0.64	5.57	2.272	
E + 16	16 at 44.7 deg	-146.03					17 h 34 min 13 s	17 h 37 min 26 s	17 h 39 min 26 s	-0.75	4.88	0.00	1.158
E + 27	16 at 44.7 deg	-146.43									5.47	0.00	2.93
E + 38	16 at 44.7 deg	-146.13								0.31	6.59	0.00	4.29
E + 49	16 at 44.7 deg	-145.63								-0.37	6.05	0.00	1.353

Table 11 (contd)

Event GMT	Bit rate mod. index, coded data	Received power, dBm P_T	P_c	P_D	RCVR lock	SDA lock	TPA init	SSA lock	SEQ lock	SSA SNR, dB	SEQ SER, %	SEQ DLR, %	SEQ ACB
E + 60	16 at 44.7 deg	-146.23								-0.93	7.275	0.00	1.734

Table 12. Small Probe 1 stable oscillator measurements

Run 1. Initial frequency 2292.282020 MHz		
Time	Point, Hz	Residual, Hz
E −10	0.0	0.50
E +2	−1.27	−0.77
E +12	−2.86	−2.36
E +20	−3.97	−3.47
E +29	−5.17	−4.67
E +40	−7.26	−6.67
E +50	−8.82	−8.32

Frequency drift for 1 h from E −10 to E +50 = −8.82 Hz, stability = 3.85×10^{-9} for 1 h.

Run 2. Initial frequency 2292.282020 MHz		
E −10	0.0	0.83
E −0	−2.04	−1.21
E +31	−4.71	−3.88
E +40	−5.46	−4 63
E +50	−6.80	−5.97
E +60	−7.58	−6.75

Frequency drift for 1 h from E −0 to E +60 = 5.54 Hz, stability = 2.42×10^{-9}

Table 13. Small Probe 2, stable oscillator measurements

Run 1. Initial frequency 2292.437720 MHz		
Time	Point, Hz	Residual, Hz
E +4	0.0	1.27
E +10	0.17	1.44
E +20	0 25	1.52
E +30	0 38	1.65
E +40	0.27	1.54
E +50	0.34	1.61
E +60	0.33	1.60
E +64	0.26	1.53

Frequency drift for 1 h from E +4 to E +64 = 0.26 Hz, stability = 1.13×10^{-10} for 1 h.

Run 2. Initial frequency 2292.437780 MHz		
E −10	0 0	0.81
E −0	0.31	1.12
E +11	0.61	1 42
E +20	0.77	1.58
E +30	1.01	1.82
E +39	1.04	1.85
E +50	1.07	1.88
E +60	1.13	1.94

Frequency drift for 1 h from E −0 to E +60 = 0.82 Hz, frequency stability = $3\ 57 \times 10^{-10}$ for 1 h.

Table 14. Small Probe 3, stable oscillator measurements

Run 1. Initial frequency 2291 552280 MHz		
Time	Point, Hz	Residual, Hz
E −5	0.0	−0.38
E −1		
E +1	0.35	−0.73
E +10	−1.07	−1.45
E +20	−2.33	−2.71
E +30	−3.15	−3.53
E +40	−3.02	−3.40
E +50	−4.70	−5.08
E +59		
Frequency drift for 1 h from E −1 to E +59 = 5.24 Hz, stability = 2.29×10^{-9} for 1 h.		
Run 2. Initial frequency 2291.552360 MHz		
E +2	0.0	1.27
E +10	0 55	1.83
E +21	−0.08	1.19
E +30	−0.32	0.95
E +40	−1.23	0.04
E +50	−1.32	−0.05
E +60	−2.10	−0.83
Frequency drift for 58 min. from E +2 to E +60 = −2.10 Hz, stability = $9\ 16 \times 10^{-10}$ for 58 min.		

Table 15. Predetection recording playback

Event	P_c	Rate, bits/s coded data	Real time SSA SNR	Playback SSA SNR
Small probe 1, Run 2 Nov. 15, 1977				
E –17	–145.0 dBm	64	5.05	No data
E –0	–146.4 dBm	64	4.53	4.88
E +16	–145.8 dBm	16	–	5.60
E +27	Link variations of	16	No data to	2.39
E +38	up to 8 dB	16	compare	1.17
E +49		16	–	End of tape
E +60	–143 to –151 dBm	16	–	–
On playback the SNR varied from –1 5 to 5 6 dB in places. This agrees with the real-time variations. All data recorded was recoverable with zero deletions.				
Small probe 3, Run 2 Nov. 16, 1977, 3 dB error in RCV AGC settings due to AGC curve shift				
E –17	–143.0 dBm	64	Computer	1.70
E –0	–145.0 dBm	64	problems no	1.02
E +16	–146.0 dBm	16	real time	0.98
E +27	–146.4 dBm	16	data	1.75
Shift in AGC curves made settings 3 dB lower than expected. At true levels of –145 dBm, SNR values of 4.8 dB were obtained. Actual levels obtained were –148 dBm which yielded SNRs of approximately 1.8 dB, which were obtained at playback. All data is recoverable from tape.				
Small probe 2, Run 1, Nov. 15, 1977				
E –22	–142.5 dBm	64	–	–
E –17	–142 5 dBm	64	10.47	10.96
E –0	–137.0 dBm	64	9.90	9.47
E +16	–142.5 dBm	16	15.38	15.74
E +27	–137.0 dBm	16	8.61	8.86
E +38	–144.5 dBm	16	7.17	6.52
E +49	–143.0 dBm	16	6.49	End of tape
E +60	–143.0 dBm	16	7 47	–
Small probe 3, Run 1, Nov. 15, 1977				
E –17	–142.0 dBm	64	9.03	Data did not lock up
E –0	–142.0 dBm	64	7.95	8.27
E +16	–143.5 dBm	16	10.53	10.21
E +27	–142.0 dBm	16	8.44	8.17
E +38	–142.0 dBm	16	8.15	8.99
E +49	–143.5 dBm	16	8.55	End of tape
E +60	–145.0 dBm	16	7 26	–

Table 16. Definition of terms for Tables 1 Through 15

ACB	Average computations per bit time
ARC	Ames Research Center
BER	Bit error rate
Bit rate	Clock frequency of the telemetry bit Information
CPA	Command Processor Assembly
CMF	Communications Monitor and Formatting Assembly
CTA 21	Deep Space Network Ground Station Compatibility Test Area at JPL
dBm	Decibel referenced to one milliwatt
DDA	Data decoder assembly
DL	RF downlink signal
DLR	Frame deletion rate
DSN mode	The Deep Space Network Ground Station Operational configuration
FDS	Spacecraft Flight Data Subsystem
HAC	Hughes Aircraft Corp.
MCD	Maximum likelihood convolutional decoder
MDA	Metric Data Assembly
MDS	Spacecraft Modulation/Demodulation Subsystem
MDS	The DSN-MARK III Data Subsystems Implementation Project
No	Noise spectral density
P_c	power in RF carrier
P_T	power total
PRA	Planetary Ranging Assembly
PFR	Problem/Failure Report
RDA	Ranging Demodulator Assembly
RF	Radio frequency
RFS	Spacecraft Radio Frequency Subsystem
RU	Range unit
SAF	Spacecraft Assembly Facility (JPL Building 179)
S/C RFS mode	Spacecraft Radio Frequency Subsystem Operational configuration
SCA	Simulation Conversion Assembly
SDA	Subcarrier Demodulator Assembly
SE	Spacecraft Ground Support Equipment
SER	Symbol error rate
SEQ	Sequential decoder
SNR	Signal-to-noise ratio
SPS	Symbols per second
SSA	Symbol Synchronizer Assembly
SSF	Space Simulator Facility (JPL Building 150)
ST_b/N_0	Signal-to-noise spectral density ratio
Symbol rate	Clock frequency of the telemetry symbol information
TBD	To be determined
TBS	To be supplied
TDL	Telemetry Development Laboratory
TLM	Telemetry
TPA	Telemetry Processor Assembly
TWT	Travelling wave tube amplifier
UL	RF uplink signal
Uplink doppler	Ramp rate of uplink RF carrier frequency
Uplink offset	Uplink RF carrier frequency displacement Relative to the spacecraft receiver test frequency
USO	Ultrastable oscillator
VCO	Voltage-controlled oscillator

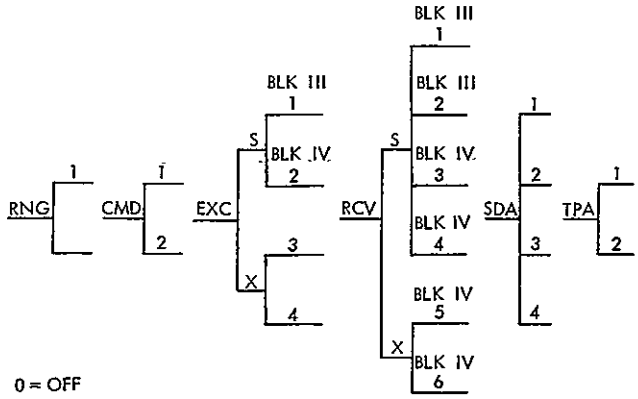

Fig. 1. DSN modes

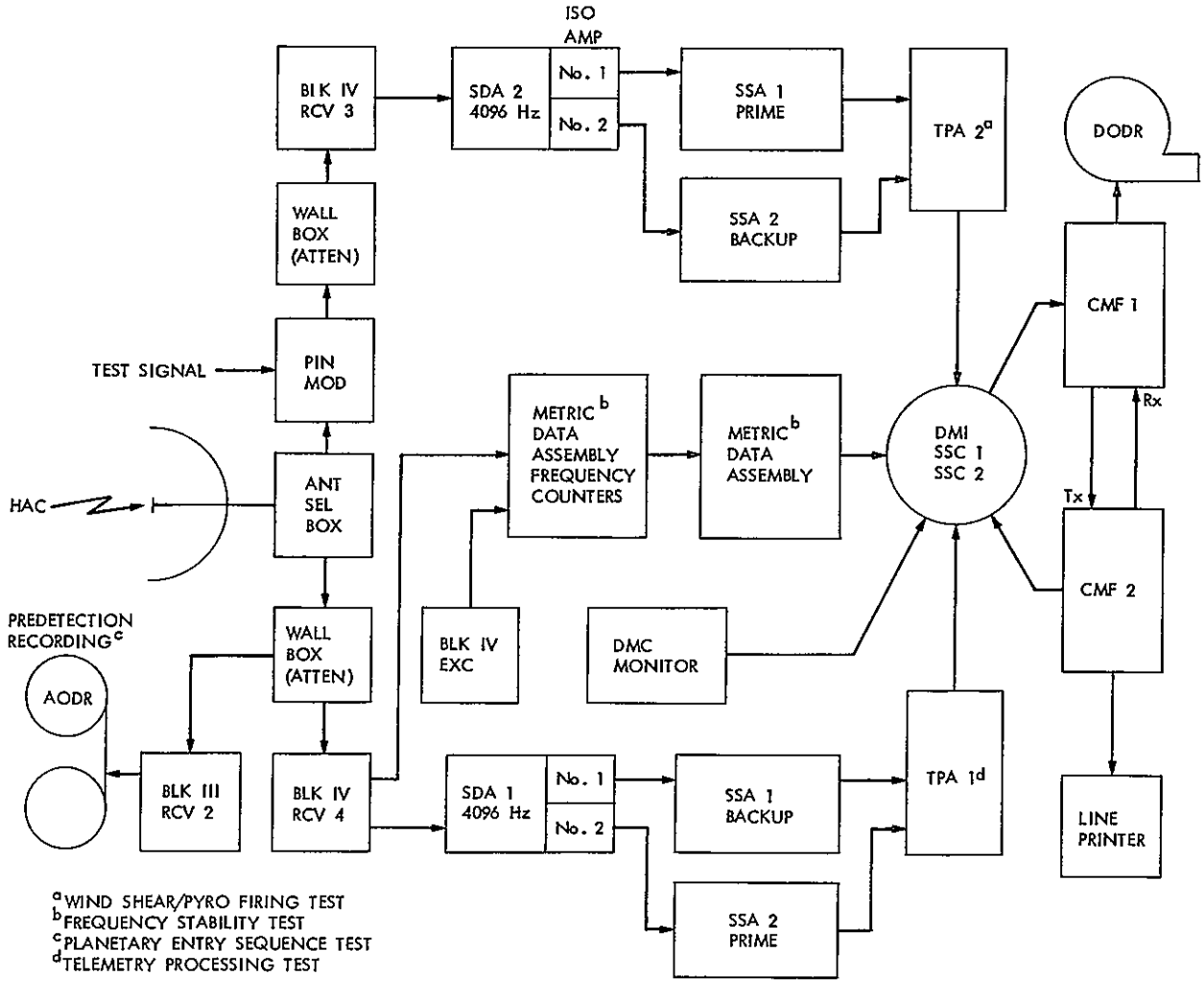

Fig. 2. DSN-PV'78 Small Probe test configuration

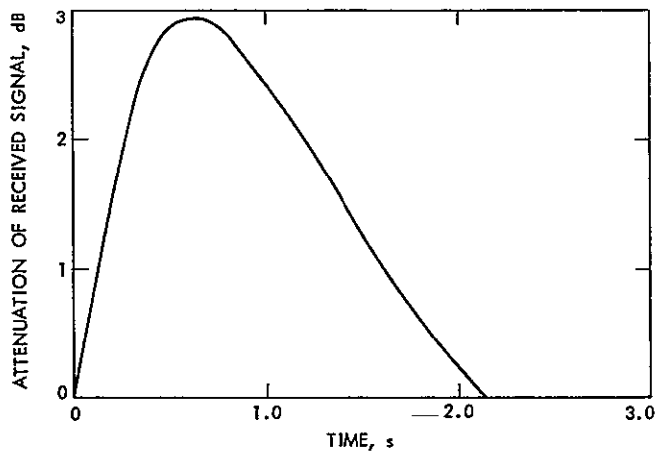

Fig. 3. Small Probe response to 5-ms step wind gust

Appendix A
S-band Microwave Link Terrain Profile

Table A-1. CTA 21/HAC link analysis

Antenna gains at midband frequency (2.296 GHz)		
	Loss, dB	Gain, dB
Cable (CTA 21 to JPL antenna)		
76 m (250 ft) of 1.9 cm (7/8 in.) hardline at −2.25 dB/30 m (100 ft)	−5.6	
Link 1 (JPL to tower): 3,590 m		
Free space loss (32.4 + 20 log f + 20 log d) where f = 2.296 GHz and d = 3590 m	−110.7	
Antenna gain		
1.8 m (6 ft) antenna (with radome)		29.1
2.4 m (8 ft) antenna		32.4
Misalignment	−2.0	
Link 2 (HAC to tower) 33,220 m		
Free space loss (32.4 + 20 log f + 20 log d) where f = 2.296 GHz and d = 33,220 m	−130.0	
Antenna gain (same as Link 1)		61.5
Misalignment	−2.0	
Cable loss (spacecraft to HAC antenna)		
122 m (400 ft) of 1.9 cm (7/8 in.) hardline at −2.25 dB/30 m (100 ft)	−9.0	
Totals	−257.3	124.0
Overall link loss	−136.3 dB	

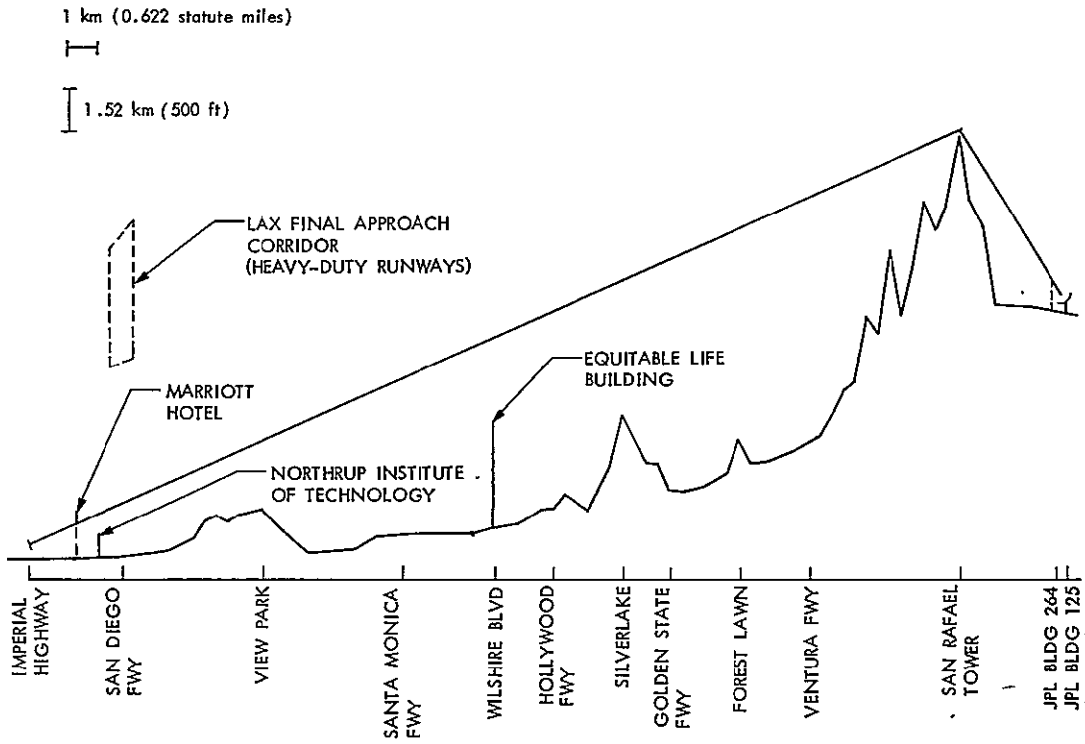

Fig. A-1. S-band microwave link terrain profile

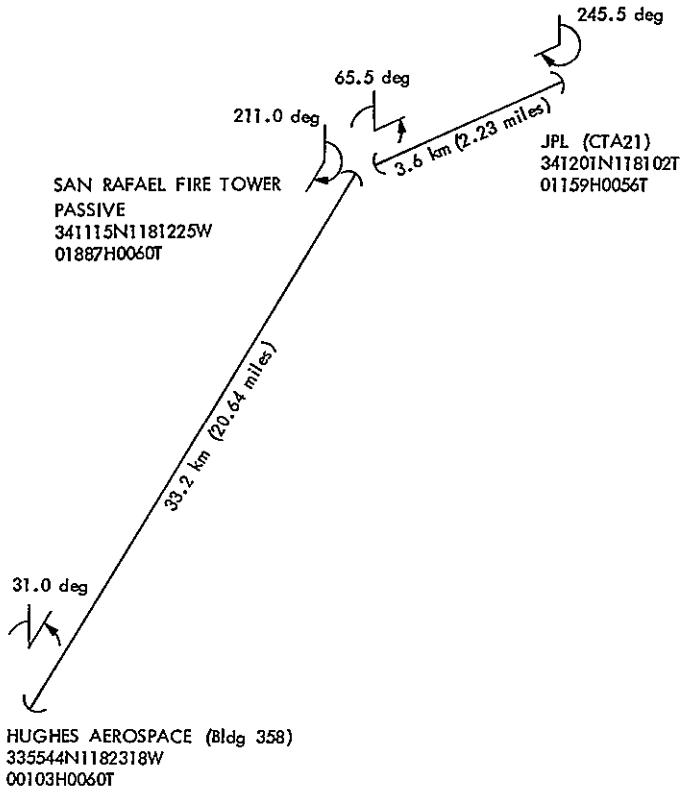

Fig. A-2. JPL/HAC microwave link

Helios Mission Support

P. S. Goodwin
TDA Support Section

W. N. Jensen and G. M. Rockwell
Deep Space Network Operations Section

This article reports on activities of the DSN Network Operations Organization in support of the Helios Project from February 15, 1978 through April 15, 1978.

I. Introduction

This article is the twenty-first in a continuing series of reports that discuss Deep Space Network support of Helios Mission Operations. Included in this article is information on Mark III Data Subsystems (MDS) testing at the Deep Space Station 11 (Goldstone, Calif.) and other mission-related activities.

II. Mission Operations and Status

As reported earlier (Ref. 1), the Helios 2 spacecraft had developed a slow gas leak from one of the spin thrusters. Since the leak rate in the spin-up direction was not correctable, the Project decided to purge the spacecraft of all remaining gas before more uncontrollable effects occurred. On March 1, 1978, Helios 2 was commanded to execute a maneuver to position itself to the stable point with an accuracy of better than 0.05 deg, (the original requirement was for better than 0.5 deg). This maneuver was very successful. The spacecraft spin rate was then corrected to a rate of 60.711 rpm. This means the spin rate will now vary between 60.711 rpm at aphelion and 60.234 rpm at perihelion due to temperature effects. Finally, the spacecraft was commanded to execute a special gas dumping procedure in order to purge all remaining gas. Subsequent analysis of data indicates that the gas dumping sequence went nominally in that the anomalous spin-up component is no longer present.

At this writing both spacecraft are fully configured and ready for the coming perihelions on April 29 and 30.

Overall coverage of both Helios spacecrafts for this period is listed in Table 1.

III. Special Activities

A. DSN-Mark III Data Subsystem (MDS) Update:

DSS 11 has completed its MDS upgrade and is now undergoing operational verification testing. To date DSS 11 has successfully performed three Helios demonstration passes.

B. Support of On-Board and Ground Experiments

Each 64-m station is preparing for the Faraday experiment by periodically exercising all of the polarimetry and Meteorological Monitor Assembly (MMA) procedures. Data from these practice passes are sent to the experimenter for validation.

References

1. Goodwin, P. S., and Rockwell, G. M., "Helios Mission Support," in the *Deep Space Network Progress Report 42-43,* Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1977.

Table 1. Helios tracking coverage

Month, 1978	Spacecraft	Station type	Number of tracks	Tracking time, h/min
February	Helios 1	26 m	25	82·19
		64 m	17	79:04
	Helios 2	26 m	37	220:28
		64 m	1	3:24
March	Helios 1	26 m	33	150.50
		64 m	14	60:08
	Helios 2	26 m	45	266:49
		64 m	1	2:45

A Demonstration of Differenced Dual-Station One-Way Doppler Conducted with Pioneer 11

C. C. Chao
Navigations Systems Section

V. J. Ondrasik and H. L. Siegel
Tracking Systems and Applications Section

In early 1976 simultaneous one-way doppler demonstrations were performed on Pioneer 11. The data noise of the differenced one-way doppler as computed from the pseudo residuals is around 0.002 Hz, which is nearly an order of magnitude smaller than the one-way doppler noise and about the same as differenced two-way/three-way doppler noise.

This report describes the demonstration results and discusses the applicability of differenced one-way radio metric data to spacecraft navigation.

I. Introduction

Differenced doppler derived from simultaneous 2-way and 3-way doppler observed at two DSN Deep Space Stations (DSS) is an effective data type for spacecraft navigation. Its advantages over single station doppler is its insensitivity to unmodeled spacecrafts accelerations and its cancelation of the solar plasma. It does suffer, however, from a restricted view period and the loss of geocentric range rate information. Rourke and Ondrasik (Ref. 1) described the information content of differenced doppler. It was shown that the information content in determining geocentric right ascension and declination is equivalent to single station tracking if the equatorial projection of the baseline vector between the two stations is comparable to the distance from the spin axis of a single station. The limitation of common view period (about 4 to 5 h) degrades the information content by approximately a factor of four when compared to full-pass single-station doppler of equivalent data quality. However, if significant unmodeled spacecraft accelerations or space plasma effects exist, the differenced doppler may yield an order of magnitude improvement in navigation accuracy over conventional two-way doppler (Ref. 2).

Normally, differenced doppler is obtained while spacecraft is in 2-way lock with one station, with an additional station in receive-only mode. This is referred to as 2-way/3-way differenced doppler. However, when the spacecraft receiver is out of lock, the doppler signal derived from an on-board oscillator,

called one-way doppler, is available at all DSSs in view. Differenced one-way doppler is equivalent to 2-way/3-way doppler and is the subject of this report.

One-way doppler alone has both advantages and disadvantages. On the positive side, one-way doppler passes through the transmission media only once, thus it is affected less than two-way doppler. In addition, for small Sun-Earth-probe angles, the one-way S-X dual-frequency doppler will not have the uplink scintillation problem; thus it is effective in removing the charged particle effects. However, the major difficulty associated with one-way doppler data is the noise from a relatively unstable reference frequency generated from an oscillator on-board the spacecraft. The induced data noise would seriously degrade the data quality for navigation purposes. The ultra-stable oscillator (USO) on the Voyager spacecraft should improve the one-way data quality by at least one order of magnitude compared with earlier on-board oscillators. However, it is not expected that the USO will provide one-way doppler of the same quality as the current 2-way doppler. An easy way to remove the oscillator noise is to difference the one-way doppler simultaneously received at two stations.

The purpose of this study is to demonstrate the concept of removing the oscillator noise by differencing simultaneously received one-way doppler data. If the one-way doppler is simultaneously received at two stations, the noise due to the on-board oscillator is common and, therefore, should cancel after differencing. For two widely separated stations, the differenced one-way doppler should have the same information content as the differenced two-way/three-way doppler. The demonstration was performed in two stages. We first demonstrated the quality of this new data type for performing short baseline and long baseline simultaneous tracking. Two passes of simultaneous short baseline one-way doppler were received at DSSs 42 and 43 from Pioneer 11, and two passes of long baseline one-way doppler were received at DSS 12 and DSS 61/62 from the same spacecraft. Secondly we compared the data quality of differenced one-way doppler with that of differenced 2-way and 3-way doppler. Three passes of simultaneous one-way doppler were obtained with 2-way and 3-way tracking at the beginning and end of each pass. After differencing one can easily compare the data quality of the two types of differenced doppler.

II. Results and Discussion

In February 1976, two passes (60 min each) of one-way doppler were simultaneously received at DSSs 42 and 43. Later in March 1976, two passes of 1-hour one-way doppler were received at the California and Spanish stations (DSS 12/DSS 62 and DSS 12/DSS 61). The received data are at S-band with a 60-s count time.

A. One-Way Doppler

Before we difference the data, let us first examine the quality of the one-way doppler. Pseudo-residuals are shown in Figs. 1 and 2. The average data noise of one-way doppler is around 0.015 Hz, which is about three times noisier than the two-way data taken shortly after the two passes. The increase in data noise is apparently due to the noise from the on-board oscillator. The long bulge in the residuals is believed to be the warm-up drift of the oscillator. The change in residual size of 0.5 Hz over an hour interval suggests that the stability of the oscillator on-board Pioneer 11 is no better than $\Delta f/f = 2 \times 10^{-10}$. Voyager is equipped with an ultra-stable oscillator (USO) and the stability is an order of magnitude better[1]. Whether the one-way data USO is useful for high precision navigation will depend on the ability to calibrate at linear drifts to within a few parts in 10^{12}. In that case, simple one-way doppler would warrant a closer look.

B. Differenced One-Way Doppler

Once the short baseline simultaneous data were differenced, nearly all the noise due to the oscillator cancels. The bias in the differenced one-way doppler residuals is down to the insignificant level of mHz or less with data noise less than 2 mHz (1σ) (Fig. 3). This shows about the same data quality as the two passes of two-way/three-way doppler taken at the same two stations two years earlier (Ref. 3). The two stations share a common frequency standard, thus no bias due to frequency offset is expected in the residuals.

The results of the two long baseline passes are not much different from that of the short baseline data in data noise which is around 2 mHz. There are biases as large as 12 mHz with slopes of 2 and 4 mHz per hour (Table 1). These biases are believed to be due to the relative frequency offset between the two stations. The slopes of 2 to 4 mHz/hr are too large to be due to the drift between the two frequency standards (Ref. 3). These slopes are probably due to the ionospheric effect together with the orbit error of Pioneer 11. The local time was 1 to 2 P.M. at DSS 12 and 9 to 10 P.M. at DSS 61/62. The estimated effect due to ionosphere is in the range of 2 to 4 mHz/h. We have reasonable confidence that neither the biases nor the slopes are due to the on-board oscillator. Some two-way/three-way doppler data at the beginning and end of each one-way/one-way pass would determine if the bias and slope are due to the on-board oscillator or other causes.

[1]According to the MJS 77 Functional Requirements Book, MJS77-4-2002, the USO will drift about 5×10^{-11} per 24 h. However, the linear drift can be calibrated to within 4×10^{-12} for an integration time of 1 s over a sample period of 600 s.

In May of 1976 three 2-h passes of simultaneous doppler were received at DSSs 12 and 62. Each pass consisted of three consecutive segments. The first and third segments were 1/2 h of 2-way/3-way tracking. The second segment was an hour of two-way station one-way tracking. This arrangement was intended to give a direct comparison between the two data types. The spacecraft Pioneer 11 was in 2-way lock at DSS 62 and concurrently assumed 3-way tracking at DSS 12. About a half hour later tracking switched to one-way mode at both stations. After one hour of one-way tracking, it returned to the 2-way/3-way mode for another half hour.

Figures 4 to 6 show the pre-fit residuals of one-way, 2-way and 3-way doppler data, with the differenced 2-way/3-way and one-way (DSS 62) doppler. Even though the one-way drifted tens of Hz due to the warm up of the spacecraft oscillator and was quite noisy (~0.1 Hz), differenced doppler based on one-way data was just as good as differenced doppler based on 2-way/3-way data. The two data types show no relative bias and no inherently different noise (Table 2) characteristics when space plasma is at low levels. Therefore, the navigation performance based upon differenced doppler data is the same if either 2-way/3-way data or 1-way/1-way data is used. The constant bias in both the 2-way/3-way and 1-way/1-way residuals is due to the frequency offset between the two station clocks. This suggests that differenced one-way doppler may be used as a tool to determine frequency offset in the same manner as differenced 2-way and 3-way doppler.

III. Conclusion

The encouraging results of these demonstrations strongly indicate that the differenced one-way doppler is equally good in data quality as the 2-way/3-way differenced doppler. Thus the differenced one-way doppler becomes a new data type for navigation without any hardware changes. This data type is effective in removing unmodeled spacecraft accelerations and space plasma effects. Future missions such as Pioneer/Venus and Galileo will be using one-way doppler for navigating the probe before impact to the planet. About 20 min of one-way doppler will be used to pinpoint the impact location on Venus during the critical period of Pioneer/Venus probe mission.

The recent receiver failure on the Voyager 2 spacecraft may require the navigation team to use differenced one-way doppler which can provide differenced doppler of the same quality as specified by Voyager mission requirements for 2-way/3-way data.

Acknowledgement

The authors would like to acknowledge the many people who helped make this demonstration possible. In particular we wish to thank R. Ryan, J. Nash, and W. Ross for scheduling of passes.

References

1. Rourke, K. H., and Ondrasik, V. J., "Application of Differenced Tracking Data Types to the Zero Declination and Process Noise Problems," in Technical Report 32-1526, Vol. IV, *The Deep Space Network Progress Report* Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1971.

2. Chao, C. C., and Ondrasik, V. J., "The QVLBI Doppler Demonstration Conducted with Mariner 10," JPL *Deep Space Network Progress Report, 42-47,* Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1975.

3. Chao, C. C. et al., "Short Baseline QVLBI Demonstration, Part II," *DSN Progress Report 42-20,* April 15, 1974.

Table 1. Summary of statistics of differenced one-way doppler (60-s count time)

DSS difference, (date)	Bias, H_z	Slope, H_z/h	Noise, (1σ), H_z
42–43 (2/8/76)	0.0011	0.0	0.0014
42–43 (2/19/76)	0.00057	0.0	0.0020
62–12 (3/17/76)	0.0126	0.0043	0.0018
61–12 (3/17/76)	0.0011	0.0027	0.0025

Table 2. Summary of differenced doppler data noise

Pass and date	$\sigma_{2\text{ way}-3\text{ way}}$ (start of pass), Hz	$\sigma_{1\text{ way}-1\text{ way}}$, Hz	$\sigma_{2\text{ way}-3\text{ way}}$ (end of pass), Hz
First pass, May 15	0.00385	0.00307	0.00265
Second pass, May 18	0.00402	0.00310	0.00295
Third pass, May 24	0.00367	0.00249	0.00301

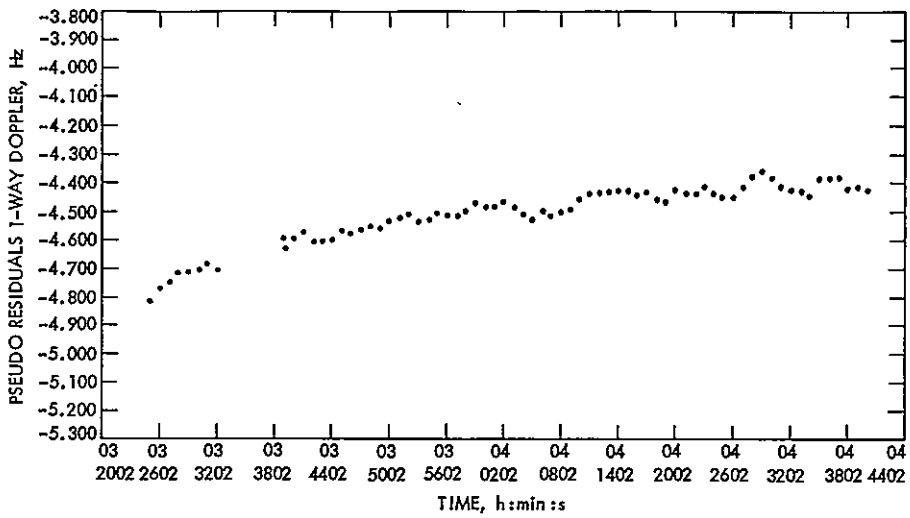

Fig. 1. Pseudo residuals of one-way doppler at DSS 42, Feb. 8, 1976

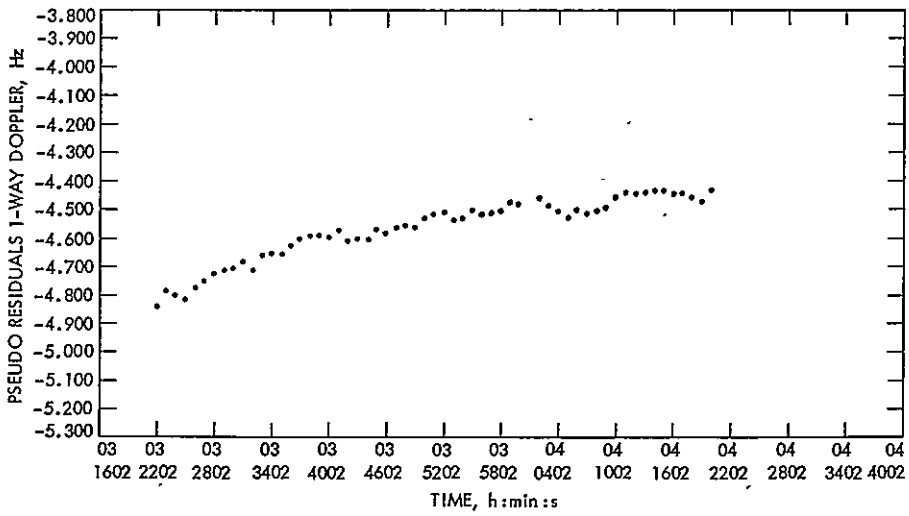

Fig. 2. Pseudo residuals of one-way doppler at DSS 43, Feb. 8, 1976

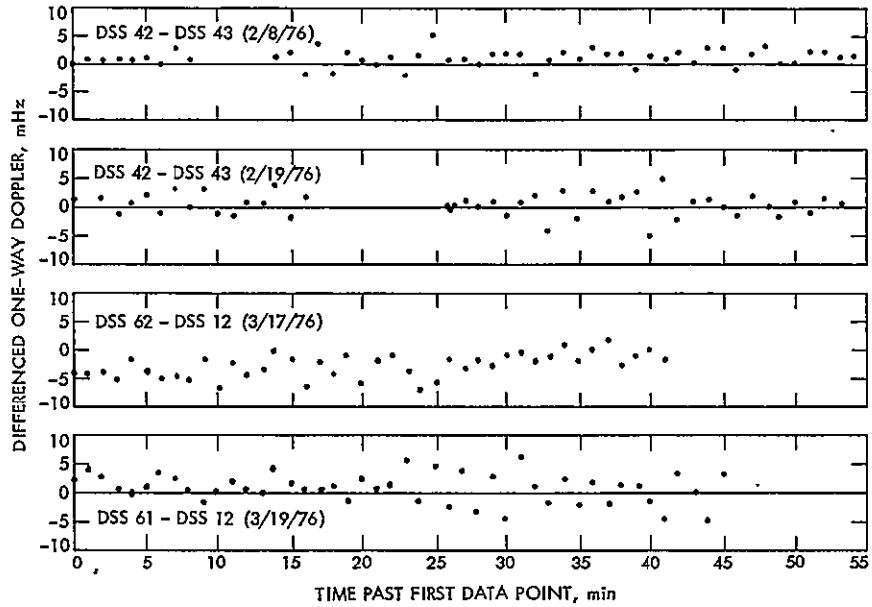

Fig. 3. Pseudo residuals of differenced two-station one-way doppler from Pioneer 11

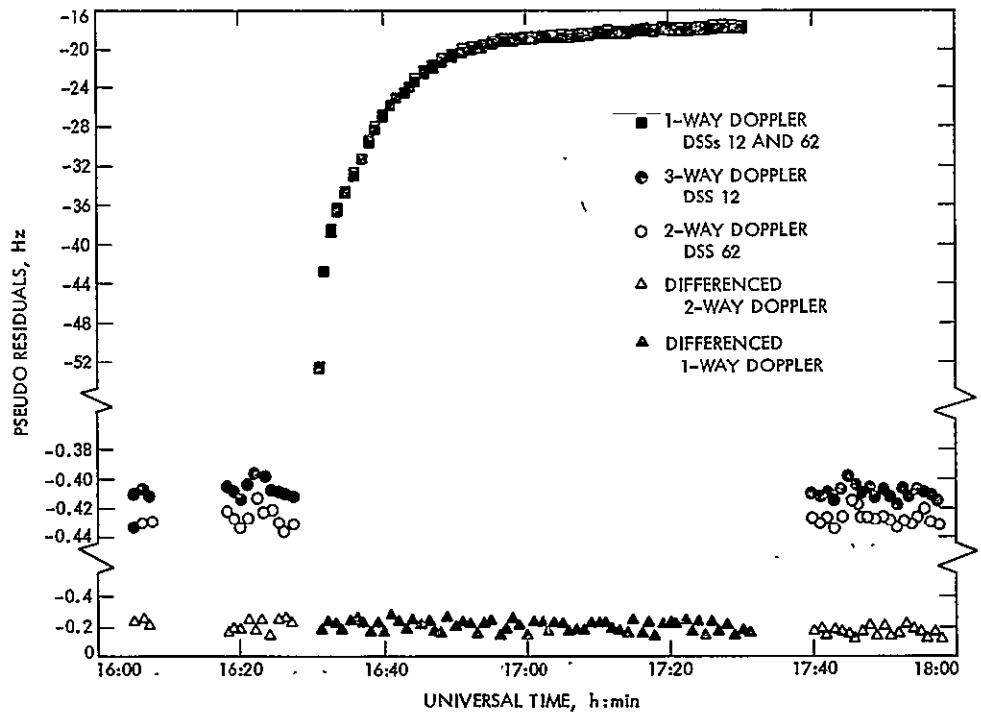

Fig. 4. Pseudo residuals of simultaneous and differenced doppler from Pioneer 11

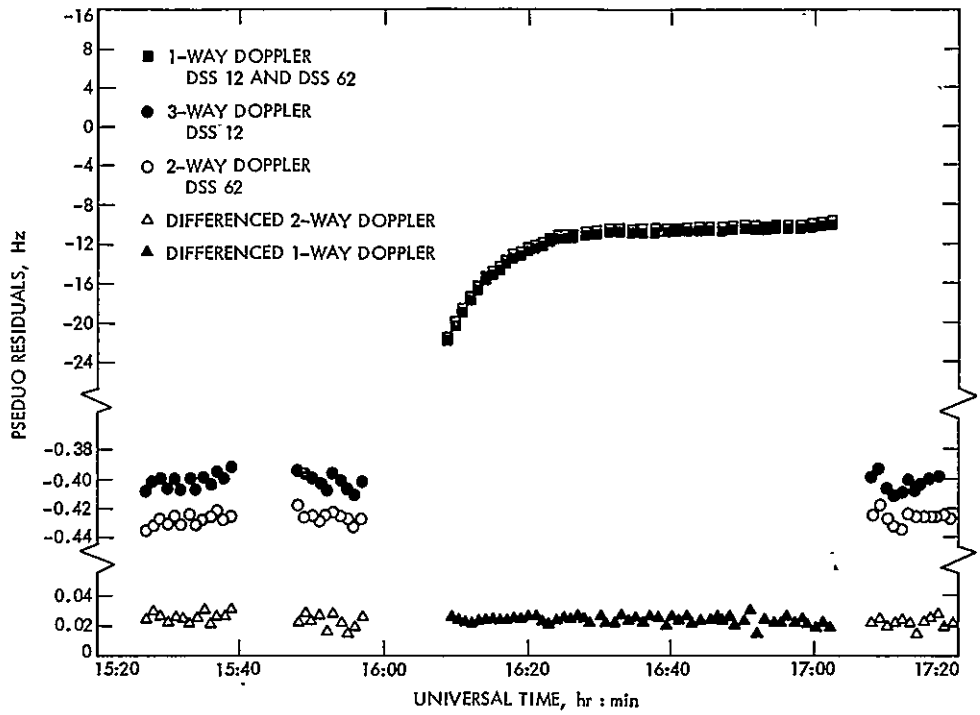

Fig. 5. Pseudo residuals of simultaneous and differenced doppler from Pioneer 11, May 18, 1976

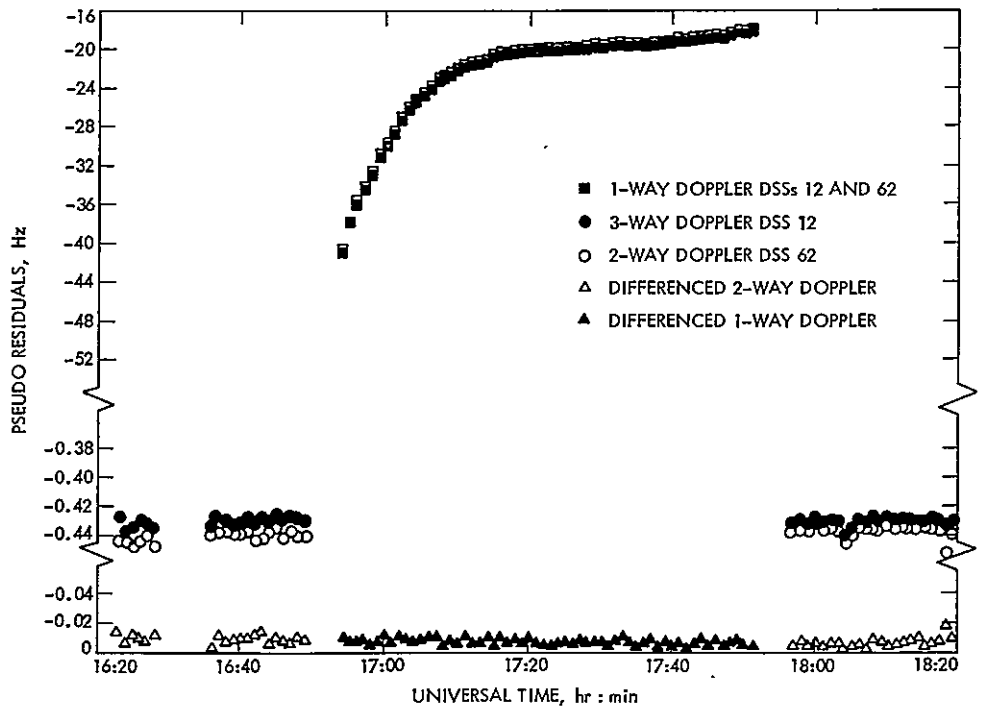

Fig. 6. Pseudo residuals of simultaneous and differenced doppler from Pioneer 11, May 15, 1976

Δ VLBI Spacecraft Tracking System Demonstration: Part I. Design and Planning

D. L. Brunn, R. A. Preston, S. C. Wu, H. L. Siegel, D. S. Brown
Tracking Systems and Applications Section

C. S. Christensen and D. E. Hilt
Navigation Systems Section

The current status of planned Δ VLBI Navigation demonstrations during Voyager Jupiter encounters is discussed. Error analysis indicates angular accuracies of 0.05 μrad are possible. Near real-time data transmission can be achieved with minimal station hardware and software modifications. A software correlator and phase tracking program, operating on the SFOF IBM 360, will be used for data processing. Data quality tests will begin on July 1978 and formal demonstrations will start on January 1979.

I. Introduction

This article is the first of a planned series which will describe the DSN development of an improved spacecraft navigation system utilizing Very Long Baseline Interferometry (VLBI) techniques. Each article is intended to provide an overview of the work accomplished for the reporting phase. Detailed analysis will be covered in separate articles. This report covers the Design and Planning phase of a planned Δ VLBI navigation demonstration using the Voyager spacecraft near Jupiter encounter. For background information the reader is referred to Ref. 1, which develops the basic theory of this navigation technique.

First it is worth while to discuss why we are developing this technique, how it relates to current DSN VLBI development efforts, and what the basic techniques are.

The main drivers may be summarized as follows:

(1) Increased efficiency of tracking station utilization for navigation

(2) Potentially more accurate estimates of right ascension α, declination δ, and their time derivatives than conventional doppler and range

(3) Lower processing costs

(4) Not limited by round-trip light times

Achievement of these desirable features appears to be well within current spacecraft and DSN technological capabilities. The need for increased operational efficiency and lower costs are well established facts. Preliminary studies have shown that

the increased measurement accuracy can provide significant mission enhancement for future projects such as Galileo.

Figure 1 shows the interrelationship of current VLBI technology applications and the development of same for spacecraft navigation purposes. Interferometric techniques have been used by radio astronomers for extra-galactic radio source (EGRS) position and structure measurements for a number of years. The DSN has been actively building a precision VLBI capability in order to better calibrate the network as an instrument for conventional radio metric navigation (Refs. 2-9).

When the interferometer is used to track essentially the downlink carrier, the data type obtained is called differential one-way doppler; when a broadband signal is supplied and tracked, the data type obtained is referred to as differential one-way ranging. When the spacecraft signal and a nearby (in angular measure) EGRS are alternately tracked, the data is referred to as Δ VLBI (the Δ in this case refers to the fact that the spacecraft and the EGRS are being differentially tracked). There are two fundamental types of Δ VLBI data: narrowband and wide-band, again depending on the bandwidth being supplied (and used) by the spacecraft RF waveform. The distinguishing features of each technique are as follows.

A. Narrowband Δ VLBI

The observables are the spacecraft S- or X-band carrier signals and a nearby (<10 deg) radio source. A single baseline is used that has a large equatorial component. The signals must be observed throughout the common view period, since the geometric changes occurring due to the Earth's rotation are needed to separate the declination and right ascension components. The measurement accuracy is declination angle dependent and angular accuracy for $\delta > 20$ deg is < 0.05 μrad. Typical measurement times are four hours.

B. Differential One-Way Range (DOR)

The observables are spacecraft generated signals that have a spanned bandwidth in the order of 50 MHz. The differences in time of arrival are interferometrically measured over two baselines and the spacecraft angular position is determined from this information. The measurement accuracy is declination angle independent. Angular accuracies of 0.05 μrad are achievable with spanned bandwidths on the order of 50 MHz and sub-nanosecond clock synchronization. Measurement time is typically ten minutes per baseline.

C. Wideband Δ VLBI

This tracking technique alternately measures the wideband spacecraft signals and signals from a nearby EGRS. The measurements are made over two baselines and the spacecraft position with respect to the EGRS are determined. The measurement accuracy is declination angle independent. Angular accuracies of < 0.05 μrad are possible since there is a large degree of error source cancellation due to the differential nature of the measurement. A significant advantage is relaxation of the sub-nanosecond clock synchronization requirement as needed by DOR. Measurement time is typically twenty minutes per baseline.

The Voyager spacecraft does not produce the wide bandwidth signals needed for precision DOR or wideband Δ VLBI measurements. Restricted bandwidth measurements, using harmonics of the unmodulated 360-kHz telemetry subcarrier, are being attempted for engineering purposes. Narrowband Δ VLBI measurements are quite possible using the spacecraft S- or X-band carriers and the Goldstone/Canberra or Goldstone/Spain baselines. The following sections will describe the planning and design of such a demonstration to be performed during the Jupiter encounter phases of Voyagers 1 and 2. Information concerning trajectory and source positions, error allocation estimates, data requirements, data acquisition and processing system and the demonstration plan will be presented.

II. Narrowband Δ VLBI Demonstration Plan

The objective of the demonstration is to demonstrate the utility and viability of Δ VLBI as a deep space navigation data type. The plan for achieving the objective is to first determine the position of Jupiter with respect to an EGRS (OJ287) using Δ VLBI and conventional doppler and range from Voyager 1. Then, using this improved Jupiter ephemeris and Δ VLBI for Voyager 2, the encounter conditions of this spacecraft will be predicted at about 30 days before the encounter takes place (Fig. 2).

Δ VLBI for Voyager 1 will be taken pre- and post-encounter in order to determine the motion of the spacecraft with respect to OJ287 (vector B) (Fig. 14). The pre- and post-encounter conventional doppler and range will produce a very accurate determination of Voyager 1 with respect to Jupiter (vector A). Combining these will yield the motion of Jupiter with respect to OJ287 (vector C). Δ VLBI for Voyager 2 taken pre-encounter will then be used to determine the spacecraft motion with respect to OJ287 (vector D). The motion of Voyager 2 with respect to Jupiter (vector E) can then be predicted.

A. Data Acquisition Strategies

The data will first be used for system testing and data quality checks. Thus, from July 1978 to December 1978, one Δ VLBI pass per month will be sufficient for testing purposes.

During January 1979 to April 1979 eight passes of Δ VLBI will be taken from Voyager 1 and OJ287. Four of the passes will be pre-encounter and four post-encounter. If possible, these passes will be symmetrically spaced with respect to encounter, e.g., one 40 days before, one 40 days after, one 20 days before, one 20 days after, etc. Further, antenna availability permitting, at least one of the passes will be within 10 days of encounter.

The gathering of Δ VLBI from Voyager 2 will begin in May 1979 and end in June of 1979. Four passes taken on a once per week basis, will be used to predict the encounter of Voyager 2.

B. Data Differencing Strategies

Narrow-band Δ VLBI from 2 stations can be defined as:

$$Z(t) = \{\rho_2(t) - \rho_2(t_0) - [\rho_1(t) - \rho_1(t_0)]\}$$

$$- \{\rho_2^*(t) - \rho_2^*(t_0) - [\rho_1^*(t) - \rho_1^*(t_0)]\} \qquad (1)$$

where the

$\rho_i(t)$ = range to station i at time t

'*' = indicates range from the EGRS

t_0 = beginning of pass

This data type is then seen to be the difference of accumulated phase of the spacecraft, and EGRS, from two stations. The expression in the first set of brackets is the difference of conventional doppler. In this mode, the two stations alternate in both receiving from the spacecraft and then the quasar.

If four stations are used, then two receive only from the spacecraft, one at each end of the baseline, and the other two receive only from the EGRS.

In order to nearly cancel the transmission media effects on the received signals, it is necessary to choose different times for differencing the spacecraft and EGRS data. In general, then, Δ VLBI will be defined as

$$Z(t, t^1) = \{[\rho_2(t) - \rho_2(t_0)] - [\rho_1(t) - \rho_1(t_0)]\}$$

$$- \{[\rho_i^*(t^1) - \rho_i^*(t_0^1)] - [\rho_j^*(t^1) - \rho_j^*(t_0^1)]\} \qquad (2)$$

where

$i = 2, j = 1$ for 2 station measurements

$i = 3, j = 4$ for 4 station measurements

t, t^1 are chosen so that the residual transmission media effects on $Z(t,t^1)$ are a minimum, as discussed in Section IV.

C. Demonstration Conduct

The demonstration will be conducted on a non-interference basis with Voyager operations using the near real-time system described in Section VII.

The processing of the data will be in parallel with Voyager operations. The navigation processing of Δ VLBI will be done on the Univac 1108 with a demonstration version of the Orbit Determination Program (ODP).

For Voyager 1, the demonstration will use the conventional radio-determined spacecraft trajectory as *a priori* for improvements to be obtained using Δ VLBI. The ephemeris of Jupiter will be the same as that used by the Navigation Team. After the encounter of Voyager 1, the Navigation Team trajectory and Δ VLBI will be processed to improve the ephemeris of Jupiter relative to OJ287. All spacecraft acceleration histories, timing and polar motion parameters, and astrodynamic constants will be supplied by the Navigation Team.

For Voyager 2 the spacecraft event histories will be supplied by the Navigation Team. The improved ephemeris of Jupiter relative to OJ287 and Δ VLBI data will be the only inputs different from those used by the Navigation Team.

III. Extra Galactic Radio Source Positions

Since the magnitudes of most Δ VLBI systematic error sources are proportional to the angular separation of the spacecraft and the EGRS, it is important to have a high spatial density of radio sources along the spacecraft flight path. Over the last few years, the DSN VLBI development effort has found a number of VLBI sources near the path of the ecliptic plane in the sky. Due to the upcoming Δ VLBI demonstration, increased emphasis has recently been placed on identifying sources along the appropriate portions of the Voyager trajectories.

For a celestial radio source to be useful for VLBI, it should possess components that are no larger in angular extent than the fringe spacing of the interferometer. The interferometer

fringe spacing is equal to the observed wavelength divided by the length of the component of the baseline that is perpendicular to the source direction. So for S-band observations with antennas separated by intercontinental distances, the interferometer sees only the radio flux coming from source components which are smaller than about 10 or 20 nanoradians in extent. Only a small fraction of celestial radio sources are that small. The search for VLBI sources has proceeded by examining with VLBI only those celestial radio sources whose intrinsic properties (e.g., spectra, variability) indicate that compact components might be present.

Figure 3 shows the present status of the search for VLBI sources in the region of the ecliptic in which the Voyager demonstration will occur. Also plotted are the nominal trajectories for the two Voyager spacecraft (JST and JSX). S-band VLBI source strengths, or correlated flux densities, are shown in parentheses next to each source (1 jansky = 10^{-26} W/m^2-Hz). X-band searches have been performed, but not yet analyzed. This figure shows that at least 25 VLBI sources with flux densities greater than 0.1 Jansky are potentially available for use in the Voyager demonstration. If the lower limit on source strength is raised to 0.4 Jansky, then the number of possible sources falls to under 10. It should be noted that the spatial distribution of these sources is rather uneven. Future searches for VLBI sources will attempt to fill in some of these gaps.

The brightest VLBI source close to the Voyager flight paths is OJ287. A time history of correlated flux density for OJ287 over the last three years using Goldstone-Australia and Goldstone-Spain baselines is shown in Fig. 4. More than a factor of two decrease in S-band source strength is apparent during that time period. A variation in VLBI source strength by a factor of two or more on a time scale of months or years is a very common property of VLBI sources. In addition, equally large variations in source strength may occur during a few hour period of observation. These short-term variations are not due to variations in the source itself, but are due to the fact that the interferometer fringe pattern changes throughout an observing period of a few hours and hence allows the interferometer to see different size elements of the source at different times. This effect also allows the source strength to vary from baseline to baseline. In the case of OJ287, the source is unusually compact, so that variations in strength caused by baseline orientation or length are not significant.

Due to both the baseline induced and intrinsic variations in source strength, the quoted strengths of the VLBI sources in Fig. 3 should be assumed to be only approximations. Any set of sources that might eventually be utilized for Δ VLBI navigation should be regularly monitored for both types of source strength variation on the same baselines that will be utilized for Δ VLBI navigation measurements. In addition, Fig. 4 shows that the X-band strength of OJ287 differs from the S-band strength. The difference is due not only to different intrinsic source strengths at X- and S-band, but also to the different size fringe spacing at the two frequencies. Hence, the source monitoring program should be carried out at the same frequency(s) which will be utilized during Δ VLBI navigation measurements.

Further VLBI measurements of the ecliptic VLBI sources will also be required in order to perform accurate determinations of their relative positions. The accurate positional information is important if accurate navigation from EGRS to EGRS or EGRS to planet is to be performed. Even with poorer knowledge of source positions, precision measurement of spacecraft-EGRS angular rate is still a valuable navigation tool.

IV. Estimate of Error Sources

This section provides estimates of the error sources affecting narrow-band Δ VLBI performance. Since Δ VLBI measures only the relative angular position between the spacecraft and a nearby EGRS, the spacecraft angular position accuracy with respect to a planet is the RSS of the measurement accuracy and the accuracy of the planetary ephemeris. In the following, each error source to which the Δ VLBI measurement is sensitive is studied separately. Numerical estimates are calculated for the geometry during Jupiter encounters of the two Voyager spacecraft. The RSS error is then compared to the 0.05 μrad error budget. Planetary ephemeris errors will be considered separately.

A. Baseline Error

The information content of a narrowband Δ VLBI can be represented by the equation:

$$\Delta\rho = -B_e \cos\delta_* [\cos(\alpha_B - \alpha_*) - \cos(\alpha_{B,i} - \alpha_*)]\ \Delta\alpha$$

$$-B_e \sin\delta_* [\sin(\alpha_B - \alpha_*) - \sin(\alpha_{B,i} - \alpha_*)]\ \Delta\delta \quad (3)$$

where

$\Delta\rho$ = accumulated differential range delay

B_e = equatorial component of the baseline

α_*, δ_* = right ascension and declination of the EGRS

$\Delta\alpha$, $\Delta\delta$ = separations in right ascension and in declination of the spacecraft from the EGRS

α_B = right ascension of the baseline center, a function of time

$\alpha_{B,i}$ = initial value of α_B

The separation $(\Delta\alpha, \Delta\delta)$ can be determined from the diurnal variation of the above equation. For this purpose Eq. (3) is rearranged into the following form:

$$\Delta\rho = A[\cos(\omega_e t + \alpha_\odot + \lambda_B - \alpha_* - \Psi) - \cos(\omega_e t_0 + \alpha_\odot + \lambda_B - \alpha_* - \Psi)] \quad (4)$$

where

$$A = B_e[\cos^2\delta_*(\Delta\alpha)^2 + \sin^2\delta_*(\Delta\delta)^2]^{1/2} \quad (5)$$

$$\Psi = \tan^{-1}\frac{\Delta\delta\tan\delta_*}{\Delta\alpha}$$

ω_e = 73 μrad/sec = the Earth's spin rate

$\alpha_\odot$ = right ascension of the Sun

λ_B = longitude of the center of baseline

It is observed that an error in B_e will result in a change in the amplitude A of the diurnal variation which in turn will result in $\Delta\alpha$ and $\Delta\delta$ estimation error; an error in λ_B will result in the same amount of error in Ψ and in turn will result in $\Delta\alpha$ and $\Delta\delta$ estimation error. The sensitivity of $\Delta\alpha$ and $\Delta\delta$ to the error in B_e and λ_B can be estimated by constructing the following matrix using Eqs. (4) and (5):

$$\begin{bmatrix} \frac{\partial B_e}{\partial(\Delta\alpha)} & \frac{\partial B_e}{\partial(\Delta\delta)} \\ \frac{\partial\Psi}{\partial(\Delta\alpha)} & \frac{\partial\Psi}{\partial(\Delta\delta)} \end{bmatrix} = \frac{1}{(\Delta\alpha)^2\cos^2\delta_* + (\Delta\delta)^2\sin^2\delta_*} \begin{bmatrix} -(\Delta\alpha)B_e\cos^2\delta_* & -(\Delta\delta)B_e\sin^2\delta_* \\ -(\Delta\delta)\sin\delta_*\cos\delta_* & (\Delta\alpha)\sin\delta_*\cos\delta_* \end{bmatrix} \quad (6)$$

The inversion of this matrix yields the following simple result

$$\begin{bmatrix} \epsilon_{\Delta\alpha} \\ \epsilon_{\Delta\delta} \end{bmatrix} = \begin{bmatrix} \frac{\partial B_e}{\partial(\Delta\alpha)} & \frac{\partial B_e}{\partial(\Delta\delta)} \\ \frac{\partial\Psi}{\partial(\Delta\alpha)} & \frac{\partial\Psi}{\partial(\Delta\delta)} \end{bmatrix}^{-1} \begin{bmatrix} \epsilon_{B_e} \\ \epsilon_\Psi \end{bmatrix} \quad (7)$$

$$= \begin{bmatrix} -\Delta\alpha/B_e & -\Delta\delta\tan\delta_* \\ -\Delta\delta/B_e & \Delta\alpha\cot\delta_* \end{bmatrix} \begin{bmatrix} \epsilon_{B_e} \\ \epsilon_{\lambda_B} \end{bmatrix}$$

The angular separation error of interest is $(\epsilon^2_{\Delta\alpha}\cos^2\delta_* + \epsilon^2_{\Delta\delta})^{1/2}$. Note that the zero-δ singularity appears only with $\epsilon_{\Delta\delta}$ due to the error in λ_B.

The error components containing $\Delta\alpha$ can be eliminated by taking the spacecraft quasar observations at times separated by $\Delta t = \Delta\alpha/\omega_e$. For instance, if $\Delta\alpha = \alpha_{S/C} - \alpha_* = 2$ deg, the baseline error will be eliminated by taking spacecraft observations 8 minutes after corresponding EGRS observations. This is true since the $\Delta\rho$ so constructed is equivalent to that for $\Delta\alpha = 0$. The time offsetting may introduce another error due to *nonlinear* clock drift. This will be discussed later.

B. Polar Motion Error

It has been shown (Ref. 2) that an error in X- or Y-component of polar motion will result in an error in VLBI measurement z/a times that due to an error in x- or y-component of the baseline where z is the z-component of the baseline and a is the Earth's radius. This relationship is still valid in the case of Δ VLBI. For a Goldstone/Canberra baseline $z \doteq a$ and the error in X and Y is of the same order of magnitude as that in x- and y-components of the baseline.

Hence Δ VLBI error due to polar motion can be estimated to be statistically equal to that due to baseline error.

C. UT1 Error

From Eq. (4) it can be seen that an error in time can be translated into an error in λ_B. Hence an error in UT1 of ϵ_t will result in an error in $\Delta\alpha$ and $\Delta\delta$ of

$$\begin{aligned} \epsilon_{\Delta\alpha} &= (\omega_e \Delta\delta \tan\delta_*)\,\epsilon_t \\ \epsilon_{\Delta\delta} &= -(\omega_e \Delta\alpha \cot\delta_*)\,\epsilon_t \end{aligned} \tag{8}$$

Current capability indicates $\epsilon_t \doteqdot 2$ mS. The projected capability during Voyager encounter is $\epsilon_t \doteqdot 1$ mS.

D. System Noise

In Δ VLBI measurement the EGRS, being a distant radio source, has a very low signal level. On the other hand the receiver has an inherent noise that is orders of magnitude higher than the EGRS signal at any frequency within the receiver passband. However, the EGRS signals received at the two ends of the baseline are highly correlated, while the system noises are not. Doubling the bandwidth increases the signal level by a factor of 2 while the noise is increased only by a factor of $\sqrt{2}$. The signal to noise ratio of the correlation signal is

$$SNR = \frac{2}{\pi}\,\frac{10^{-26}}{2K}\,S\,\frac{\pi}{4}\,d_1\,d_2\sqrt{\frac{e_1 e_2 BT}{T_{S,1}\,T_{S,2}}} \tag{9}$$

where

K = Boltzmann's constant (1.38×10^{-23})W-s/deg.

S = total observed flux density of source over baseline of interest (Jansky = 10^{-26} W/m² - Hz)

d_1, d_2 = physical antenna diameter (m) at DSS_1, DSS_2

e_1, e_2 = antenna aperture efficiency at DSS_1, DSS_2

B = VLBI receiver filter bandwidth, Hz

T = integration time(s) per fringe phase point

T_{s_1}, T_{s_2} = total system noise temperature (K) at DSS_1, DSS_2

As will be shown in Section IV, spacecraft signals can each be correlated with a local model (essentially noise-free) and the resulting SNR is much higher than that of EGRS signals. Hence it is the EGRS SNR that is of concern. The effect of SNR on $\Delta\rho$ in Eq. (3) is

$$\begin{aligned} \epsilon_{\text{phase}} &= \frac{1}{\text{SNR}} \\ \therefore\ \epsilon_{\Delta\rho} &= \left(\frac{1}{\text{SNR}}\right)\frac{c}{2\pi f} \end{aligned} \tag{10}$$

where

c = speed of light

f = effective bandpass center

The minimum SNR for which EGRS signals can be phase tracked from VLBI measurements is about 5–10. This corresponds to ~ 0.42 cm error in $\Delta\rho$ at S-band. It is now desired to have a means to estimate the sensitivity of the determination of the angular separation to the system noise. Precise estimation requires an integral over the pass (~ 4 hours) which will not provide much insight. A U-V plane approach which is first used by Melbourne and Curkendall for Navigation (Ref. 1) and which is normally good to within a factor of 2, is adopted here. Let:

$$\begin{aligned} U &= \frac{1}{\cos\delta_*}\,\frac{\partial(\Delta\rho)}{\partial(\Delta\alpha)} \\ V &= \frac{\partial(\Delta\rho)}{\partial(\Delta\delta)} \end{aligned} \tag{11}$$

When Eq. (3) is substituted into Eq. (11) we see that the (U,V) vector will trace out an ellipse as the earth spins (i.e., α_B increases):

$$\left[\frac{U\text{-}B_e\cos(\alpha_{B,i}-\alpha_*)}{B_e}\right]^2 + \left[\frac{V\text{-}B_e\sin\delta_*\sin(\alpha_{B,i}-\alpha_*)}{B_e\sin\delta_*}\right]^2 = 1 \tag{12}$$

An example of such ellipse is shown in Fig. 5. The information content of angular separation in the direction of a (U,V) vector for a single instantaneous measurement is proportional to the length of the vector. The information content in the orthogonal direction is proportional to the component in this direction of the change in the (U,V) vector. For fringe-phase Δ VLBI, the change in the (U,V) vector is always in the same direction as the (U,V) vector itself near the beginning of the pass and little information in the orthogonal direction is

available. To acquire this information a long pass is required. For a long pass of Δ VLBI measurement, the information can be shown to be inversely proportional to:

$$\sigma_T^2 \doteqdot \sigma_{\Delta\rho}^2 \left[\frac{2}{S_r^2} + \frac{2}{S_\perp^2}\right] \quad (N/3) \tag{13}$$

where

$\sigma_{\Delta\rho}^2$ = the variance of the measured $\Delta\rho$.

S_r = maximum vector length $(U^2 + V^2)^{1/2}$ over the pass

$S_\perp$ = maximum accumulated change of the vector (U,V) in the direction orthogonal to the (U,V) vector with maximum length

N = number of samples over the pass

3 = minimum number of samples required to acquire information in two orthogonal directions

For $\delta_* = 20$ deg and Goldstone/Canberra baseline, a four-hour pass results in $S_r \doteqdot 3200$ km, $S_\perp \doteqdot 800$ km. With $N = 240$ (1 sample/minute) and $\Delta\rho = 0.42$ cm, the error in the estimated angular separation is $\sigma_T \doteqdot 1$ nrad.

E. Transmission Media Error

The effect of transmission media on Δ VLBI angular separation measurement can be considered to be that of a frequency noise[1]. It differs from the usual system noise (a phase noise) in that the effect may build up over the pass, resulting in a non-zero accumulated error. For instance a 1 mm/s frequency noise, sampled every 1 minute over 4 hours, will result in an accumulated error of $(0.1 \times 60) \times \sqrt{4 \times 60} \doteqdot$ 93 cm. Hence an accumulated media error of 93 cm can be approximately represented by a frequency noise of 1 mm/sec. With this frequency-noise concept, the effect on Δ VLBI angular separation measurements can be easily estimated by an information ellipse similar to subsection (D) above.

The (U,V) vector is now defined as

$$U = \frac{1}{\cos\delta_*}\frac{\partial(\Delta\dot{\rho})}{\partial(\Delta\alpha)}$$

$$V = \frac{\partial(\Delta\dot{\rho})}{\partial(\Delta\delta)} \tag{14}$$

where the dots denote time derivatives. When Eq. (3) is substituted into Eq. (14) the following information ellipse is constructed:

$$\left(\frac{U}{\omega_e B_e}\right)^2 + \left(\frac{V}{\omega_e B_e \sin\delta_*}\right)^2 = 1 \tag{15}$$

This is shown in Fig. 6.

The information is inversely proportional to

$$\sigma_T^2 \doteqdot \sigma_{\Delta\rho}^2 \left[\frac{2}{S_r^2} + \frac{2}{S_\perp^2}\right] \Big/ (N/2) \tag{16}$$

with S_r, $S_\perp$ and N as defined in Section 4 above; the minimum number of samples is 2 in this case. As an example, for the same geometry as in Section 4 we obtain $S_r \doteqdot 420$ m/sec, $S_\perp \doteqdot 220$ m/sec. Hence the error in the estimated angular separation for 1 mm/sec frequency noise is 0.66 μrad. In other words, a 1-cm accumulated media error will result in $660/93 \doteqdot 7$ nrad error in angular separation estimation. To calculate the angular separation errors the accumulated errors in cm are multiplied by the sensitivity, 7 nrad/cm, estimated by the information ellipse. Hence, the accumulated media error must be limited to not more than a few centimeters. The accumulated media error is estimated separately for troposphere, ionosphere and solar plasma in the following.

1. Troposphere. The accumulated tropospheric effect is mainly due to the difference in elevation angles between the four ray paths. Zenith tropospheric effects can be calibrated by a surface model (Ref. 10) or a seasonal model (Ref. 11) to better than 6 cm at each DSS. Hence the residual effect at *either* DSS is

$$\varepsilon = 6 \left|\frac{1}{\sin\gamma_{S/C}} - \frac{1}{\sin\gamma_*}\right| \text{ cm} \tag{17}$$

where $\gamma_{S/C}$ and γ_* are the elevation angles of the ray paths looking at the spacecraft and EGRS, respectively. The total effect is the RSS of the *accumulated* residual effects at the two DSS.

When the angular separation is in right ascension only, $\gamma_{S/C}$ and γ_* can be made equal by time offsetting as in the case of $\Delta\alpha$ component of baseline error. When the separation is in declination only a study shows that the difference of the two terms in Eq. (17) can be made nearly constant over the pass, also by time offsetting. The RSS accumulated residual effect

[1]This concept is suggested by J. K. Miller of the Navigation Systems Section.

can be reduced by up to two orders of magnitude. The optimum time offset with separation in both $\Delta\alpha$ and $\Delta\delta$ is

$$\Delta t = 4\ \Delta\alpha + 2.8\ \Delta\delta \text{ minutes} \tag{18}$$

where $\Delta\alpha$ and $\Delta\delta$ are in degrees. Of course, with such time offset the $\Delta\alpha$ component of the baseline error will not be completely eliminated. For JST Voyager encounter ($\Delta\alpha = -11.8$ deg, $\Delta\delta = 0.7$ deg) the residual accumulated effect is estimated to be reduced from 30 cm ($\Delta t = 0$) to 0.2 cm ($\Delta t = 45.2$ min). For JSX Jupiter encounter ($\Delta\alpha = 2.2$ deg, $\Delta\delta = -2.5$ deg) it is reduced from 0.6 cm ($\Delta t = 0$) to 0.4 cm ($\Delta t = 1.8$ min). That the improvement is marginal is due to the already small effect with $\Delta t = 0$, a result of near ideal separation ($\Delta\alpha \doteqdot -0.7\ \Delta\delta$) for which no time offset is needed. The 45-minute time offset for JST encounter may introduce another error due to the imperfect correlation between the measurements separated by such time offset. This is estimated to be ~ 1 cm in accumulated effect. The total tropospheric effect for JST is thus ~ 1 cm. For JSX encounter, the time offset is small and such imperfect correlation is negligible.

2. **Ionosphere.** The accumulated ionospheric effect is mainly due to the difference in the product of the elevation-angle factor $g(\gamma)$ and the solar-zenith-angle factor $f(X)$ where

$$g(\gamma) = \left\{ \left[(R_e + h_2)^2 - R_e^2 \cos^2\gamma \right]^{1/2} - \left[(R_e + h_1)^2 - R_e^2 \cos^2\gamma \right]^{1/2} \right\} \Big/ (h_2 - h_1) \tag{19}$$

with $R_e = 6370$ km, $h_1 = 215$ km and $h_2 = 454$ km, and where

$$f(X) = \begin{cases} 0.2 + 0.8 \cos^{2/3} X\,, & |X| < 90 \text{ deg} \\ 0.2\,, & |X| > 90 \text{ deg} \end{cases} \tag{20}$$

with X being the solar-zenith angle.

With time offsetting for the minimization of the tropospheric effect the $g(\gamma)$ effect of the ionosphere is also reduced to a negligible amount. However the $f(X)$ effect will be increased since the difference in X will be increased by time offsetting. The proper time offset is a compromise between minimizing these effects. For JST Voyager encounter the Sun-Earth probe angle is (SEP $\doteqdot$ 140 deg) so that the overlapped tracking will be during local night ($|X| > 90$ deg) at both DSS and the $f(X)$ effect is minimal. Hence one can use the full time offset which minimizes the tropospheric error. The estimated accumulated error, assuming a 20-cm calibration error[2] at $\gamma = 90$ deg and $X = 0$, is ~ 0.2 cm. For JSX Jupiter encounter, the small time offset ($\Delta t = 1.8$ min) which minimizes the tropospheric effect will increase $f(X)$ effect of the ionosphere only slightly and still can be used. The estimated accumulated error (SEP $\doteqdot$ 27 deg) is ~ 1.5 cm.

In actual calibration of the ionospheric effect, the geomagnetic-latitude adjustment has to be made to minimize the north-south horizontal gradient effect.

3. **Solar plasma.** The solar plasma effect on VLBI can be estimated by (Ref. 12)

$$\epsilon = 0.039 \left[10 \sin(\text{SEP}) \right]^{-1.3} \left(\frac{\tau}{10} \right) 0.75 \tag{21}$$

where τ is the time for the solar wind to travel across the distance between the two ray paths. Assuming a solar wind velocity of 400 km/sec perpendicular to the two ray paths, the τ will vary in the range $15 < \tau < 26$ sec. For Δ VLBI, the separation between the two ray paths from a DSS looking at the spacecraft and EGRS, even at 1 deg separation, is 2 orders of magnitude greater than that between the two ray paths at different DSS but looking at the same source, spacecraft or EGRS. Hence, Eq. (19) is used to estimate the residual effect of VLBI (two ray paths looking at the same sources); the accumulated residual effect for Δ VLBI will be $\sqrt{2}$ times as large since the two effects for the two VLBI measurements can be considered to be uncorrelated. It should be pointed out that the time offsetting between the spacecraft and EGRS measurements discussed earlier will not affect the solar plasma error since we do not assume any cancellation between the spacecraft and EGRS measurements.

For JST Jupiter encounter, SEP $\doteqdot$ 140 deg. Since Eq. (21) is applicable to SEP $\leqslant$ 90 deg, the estimate will be conservative if SEP = 90 deg is used for any SEP $>$ 90 deg. With this assumption we get $\epsilon < 0.4$ cm. For Δ VLBI the residual accumulated effect will be $0.4\ \sqrt{2} \doteqdot 0.6$ cm. A similar calculation for JSX Jupiter encounter yields a residual accumulated effect of 1.6 cm.

4. **Clock error.** One of the features of Δ VLBI is the complete cancellation of clock drift between the two DSS if simultaneous spacecraft and EGRS measurements are taken and differenced. With time offset between the two types of measurement (to reduce error in baseline, polar motion, UT1 and tropospheric effect) an error due to clock drift is

[2] Using on-site Faraday rotation measurement from ATS Satellite.

introduced. Clock bias and linear phase drift has no effect on Δ VLBI; the corrupting effect is the second-order clock drift (linear frequency drift). Let the frequency stability be $\Delta f/f$ over the time T of the pass, the accumulated error in Δ VLBI with a time offset Δt can be shown to be

$$\epsilon_{\Delta\rho} = \frac{c}{2} \frac{(\Delta f/f)}{T} (\Delta t)(2T + \Delta t) \tag{22}$$

where c is the speed of light.

For JST Jupiter encounter Δt = 45.2 min, and for JSX Jupiter encounter Δt = 1.8 min. Assuming $\Delta f/f = 10^{-14}$ and T = 4 hours, the accumulated error will be 0.89 cm and 0.03 cm, respectively. The resulting angular separation error (at 7 nrad/cm) will be 6.2 nrad and 0.2 nrad, respectively. These errors are at least an order of magnitude lower than the other error sources which are reduced by introducing time offset.

This analysis does underscore, however, the importance of the hydrogen maser level of clock performance implicitly assumed by using 10^{-14} in the calculation. Degradation to 10^{-13} would in effect seriously compromise the system ability to minimize the effects of calibration errors and would force the use of only very small separation angles.

5. EGRS position error. This error is here defined as the error in pinpointing the "center" or the source. Most well-known EGRS sources are determined to within 10 nrad in angular position. For OJ287, which will be approached by the Voyager spacecraft near encounters, the angular position has been determined to within 3 nrad.

The estimated error sources for the two Voyager Jupiter encounters are summarized in Figs. 7 and 8. The projected accuracy for baselines and polar motion is 0.5 m in each component; for UT1 it is 1 mS. The RSS errors are well within the 0.5 μrad budget, especially for JST Jupiter encounter.

Instrumental errors of the system will presumably be removed by the use of phase calibrators. Estimation of the residual errors will be made when the calibrator performance and filters are studied.

V. Data Requirements

For the narrowband Δ VLBI spacecraft system demonstration, radio signals from a spacecraft and EGRS are alternately received at two DSS, infinitely clipped, sampled at a rate twice the VLBI receiver filter bandwidth, and digitally crosscorrelated to produce fringe phase. The narrowband Δ VLBI system requires fringe phase data points with a signal-to-noise ratio (SNR) of 5-10 or more for successful phase tracking. Additionally, several fringe phase data points are required every five minutes for phase connection.

Assuming the EGRS is a white noise source over the VLBI receiver filter bandwidth then the fringe phase SNR for the digital crosscorrelation of signals from DSS_1 and DSS_2 is given by Eq. (7). Let the radio source for the VLBI measurement be a spacecraft carrier signal. Assume the spacecraft is a spectral line radio source in the VLBI receiver filter bandwidth. The fringe phase SNR for the digital crosscorrelation of the signal from DSS_1 and DSS_2 is given by Eq. (9).

An alternate technique to direct crosscorrelation of spacecraft signals is local model crosscorrelation. For local model crosscorrelation, the spacecraft signal is crosscorrelated with a locally generated model of the spacecraft signal to determine their phase difference. The phase SNR for the digital local model crosscorrelation at DSS_1 is

$$SNR = 2\sqrt{\frac{P_{c_1} T}{\pi K T_{s_1}}} \tag{23}$$

The phase produced by local model crosscorrelation at DSS_1 is differenced with the phase at DSS_2 to produce fringe phase. Not only does local model crosscorrelation have a distinct SNR advantage over direct crosscorrelation, but only a digital representation of a phase and not an entire spacecraft signal needs to be transmitted from the DSS to JPL for real time processing.

The total minimal number of bits required from each station per fringe phase point is:

$$N = 2BT \tag{24}$$

Figure 9 is a plot of the expected S-band EGRS VLBI performance using OJ287 on the Goldstone – Canberra baseline for the narrowband Δ VLBI spacecraft tracking system demonstration. The performance for all possible VLBI DSS combinations on this baseline is shown. Integration time refers to the time period over which crosscorrelation is performed to produce one fringe phase data point at the sample rate of 80 kbits/s.

Table 1 is a summary of data requirements for the S-band narrowband Δ VLBI spacecraft tracking system demonstration.

VI. Data Acquisition

Figures 10, 11 and 12 depict the hardware configurations needed for the near real time Δ VLBI data acquisition system at DSS 14, DSS 13, and DSS 42/43 and 61/63. The data acquisition system allows near real-time demonstration of either two or four station narrowband Δ VLBI spacecraft tracking. Additionally, a backup MK II capability simultaneously exists.

Very little additional equipment is needed at the operational stations. Specifically, a power divider is added to the VLBI Bandwidth Synthesis (BWS) Assembly which splits the baseband signal into two paths: one to the IF converter and subsequent IVC recorder (MK II) and the other to the Occultation Data Assembly (ODA) via a 40-kHz low-pass filter. The ODA contains an A/D converter and a Modcomp computer with disk and tape drive. The low-pass filter bandwidth is set by the ODA data rate capability which is specified as 80 kbps maximum. The ODAs will be installed at each 64-m station by mid-1978 and the multiplexers have two spare input ports, one of which will be assigned for VLBI use. Data will be recorded and time tagged using either modified PV-78 ODA software or new software. The resultant tape is played back via the 56-kbps GCF WBDL to the NOCC at JPL using the Telemetry Processor Assembly (TPA). Minor software changes may be needed for the TPA Modcomp to perform this function.

Since DSS 13 does not have an ODA or WBDL, the 40-kHz analog signal out of the BWS assembly at DSS 13 will be digitized, time tagged, and transmitted by microwave to DSS 14 via GCF 10 for transmission to JPL.

VII. Data Processing

The current R&D VLBI system (MK II) is designed for platform parameter calibration and clock synchronization. The measurement technique involves observation of large numbers of radio sources and utilizes wide bandwidths (2 MHz) with corresponding high sample rate (4 Mb/s). Data is recorded on video tape and is shipped to JPL for processing on a special purpose hardware correlator located at CIT.

Use of this data acquisition and processing system for spacecraft navigation has several disadvantages, i.e., long delays due to logistics, large data volume and limited use of the correlator, which is also used for radio astronomy purposes. An operational version of this system is being implemented by the DSN and is scheduled to be operational by July 1979. Unfortunately this is too late for the Voyager Δ VLBI demonstrations; therefore, an alternative R&D design is being implemented which will allow near-real time data transmission (Section V) and processing. The salient features of this processing system are described in the following paragraphs.

Radio Source VLBI data processing can be conceived of as a data compression stage followed by a parameter estimation stage. In the data compression stage we input pairs of data samples taken simultaneously at two DSS. The model geometric delay and velocity difference at the two stations are removed by delaying one bit stream and digitally heterodyning one bit stream with the differential phase rate. Then the two streams are crosscorrelated and averaged. Finally the residual velocity difference (residual fringe frequency) is detected with a Fourier transform and the phase is tracked with a least-squares technique. When there are large time gaps in the observation sequence, it becomes necessary to resolve 2π phase ambiguities, i.e., to connect phase. The parameter estimation stage improves a priori values for source positions, station baselines, UT1, station clock offsets and polar motion. This program called "Phasor" currently runs on the CIT IBM 370 (Ref. 13).

Several modifications are required for processing Δ VLBI navigation data. These are:

(1) The system must return results in "near real time". That is, Δ VLBI spacecraft orbit solutions should be available on the order of a day. The Mark II turnaround is on the order of weeks due to tape shipment delays. DSN plans are to implement the correlator and phase tracking functions on the SFOF IBM 360 computers and to transmit data sampled at 500 kbits/s to JPL via the GCF wideband data lines. The previously described R&D near real-time configuration (Section V) is being implemented to provide data sampled at a reduced 80 kbit/s rate. A preliminary version of the IBM 360 correlator will be utilized to implement this system in July 1978. It should be noted that the capability to operate the conventional 4-megabit video recorder system concurrently will be maintained.

(2) The system must be able to model spacecraft motion for local correlation. The spacecraft models will provide a local model to correlate each station separately. The extragalactic radio source data will continue to be cross-correlated in the conventional manner.

(3) The system must be able to phase track spacecraft data. Since the spacecraft signals are very narrow-band, the correlated amplitude does not fall off rapidly in the time domain. Thus a special version of the phase tracking is being prepared which just fits phase residual in one lag channel only.

(4) The system must be able to solve for spacecraft orbits. The R&D system will accomplish this by preparing a

file of phase observables – called the Δ VLBI data file on an 1108 readable tape. Then a special program, the Δ VLBI processor, will prepare the data for processing on the standard Voyager ODP (orbit determination program).

An overall block diagram of the system is shown in Fig. 13. Data is received and recorded in the NOCC by the network log processor/real-time monitor. These tapes along with observation times, baseline parameters, Earth rotation rate, EGRS α & δ, and precalculated time delay and clock offset information are input to the IBM 360 software correlator which produces the fringe frequency correlation coefficients. The correlator operates in a cross-correlation mode when processing EGRS signals and a local model or cross-correlation mode when processing spacecraft signals. This step in the processing compresses the data by a factor of approximately 10^6.

Further processing of the data is accomplished in the PHASOR program, which performs a least-squares fit to the correlator output data producing a refined estimate of the fringe phase residuals plus the best estimate model. The ODP pre-processor connects the residual fringe phase, recovers the model to produce the estimated true fringe phase versus time and formats this data for input to the ODP. From this data the ODP generates the desired spacecraft position relative to the EGRS.

References

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

2. J. G. Williams, "Very Long Baseline Interferometry and Its Sensitivity to Geophysical and Astronomical Effects", JPL Space Program Summary, 37-62 Vol. II, March 31, 1970, pp. 49-55.

3. Slade, M. A., Preston, R. A., Harris, A. W., Skjerve, L. J., Spitzmesser, D. J., "A1SEP – Quasar Differential VLBI", *The Moon*, 17, 133-147.

4. Thomas, J. B., "An Analysis of Long Baseline Radio Interferometry," in *The Deep Space Network Progress Report*, Technical Report 32-1526, Vol. VII, p. 37, Feb. 1972.

5. Thomas, J. B., "An Analysis of Long Baseline Radio Interferometry, Part II," in *The Deep Space Network Progress Report*, Technical Report 32-1526, Vol. VIII, p. 29, May 1972.

6. Thomas, J. B., "An Analysis of Long Baseline Radio Interferometry, Part III," in *The Deep Space Network Progress Report*, Technical Report 32-1526, Vol. XVI, p. 47, Aug. 15, 1973.

7. Goldstein, R., "Clock Calibration Via Quasar," SPS 37-48 Vol. II, p. 79, Nov. 1967.

8. Hurd, W. J., "Preliminary Demonstration of Clock Synchronization by Radio Interferometry," in *The Deep Space Network Progress Report* 42-37.

9. Hurd, W. J., "DSN Station Clock Synchronization by Maximum Likelihood VLBI", *The Deep Space Network Progress Report*, TR 32-1526, Vol. 10.

10. A. L. Berman, *The Prediction of Zenith Range Refraction From Surface Measurements of Meteorological Parameters*, TR 32-1602, July 15, 1976.

11. K. L. Thuleen and V. J. Ondrasik, "The Repetition of Seasonal Variations in the Tropospheric Zenith Range Effect", DSN Progress Report, TR 32-1526, Vol. VI, December 15, 1971, pp. 83-98.

12. P. S. Callahan, "An Analysis of Viking S-X Doppler Measurements of Solar Wind Columnar Content Fluctuations", DSN Progress Report PR 42-44, April 15, 1977.

13. Purcell, G., "A Procedure for Preliminary Reduction of Bandwidth Synthesis Data," in *The Deep Space Network Progress Report*, Technical Report, 42-33.

Table 1. Data requirements for the S-band narrowband Δ VLBI spacecraft tracking system demonstration

DSS combination	EGRS crosscorrelation		Spacecraft crosscorrelation		Spacecraft local model crosscorrelation	
	T	N	T	N	T	N
64 m – 64 m	1.61×10^{1}	1.3×10^{6}	1.82	1.5×10^{5}	3.37×10^{-2}	2.7×10^{3}
64 m – 26 m	1.64×10^{2}	1.3×10^{7}	1.82×10^{1}	1.5×10^{6}	3.38×10^{-1}	2.7×10^{4}
26 m – 26 m	1.67×10^{3}	1.3×10^{8}	1.83×10^{2}	1.5×10^{7}	$3\,38 \times 10^{-1}$	2.7×10^{4}

SNR = 10, B = 40 kHz
T = integration time in seconds per fringe phase point
N = number of bits per station per fringe phase point
OJ287 quasar radio source (0.8J correlated flux density for Goldstone — Canberra baseline)
Voyager spacecraft at JSX Jupiter encounter (64 m P_c = – 150.3 dBm, 26 m P_c = –158.5 dBm)

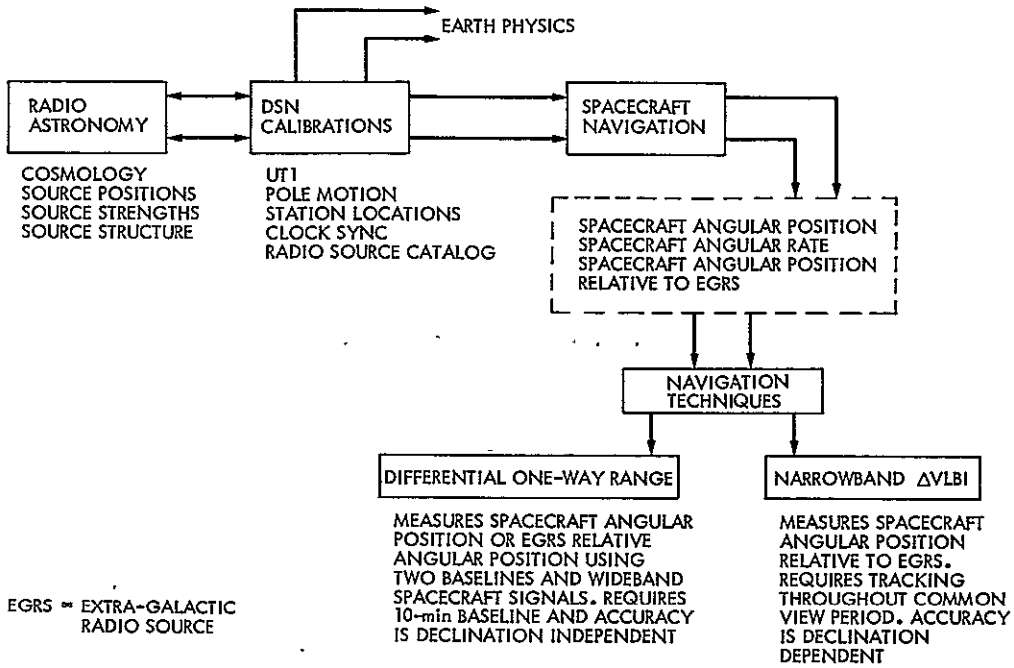

Fig. 1. Applications of very long baseline interferometry

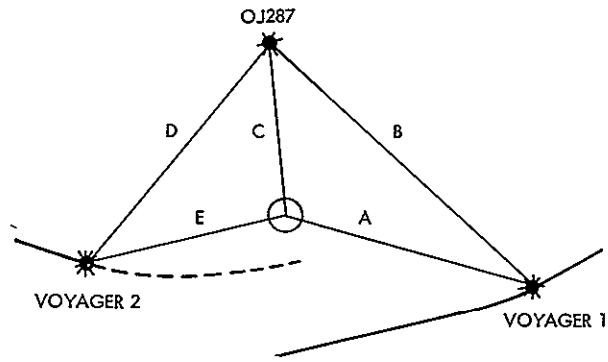

Fig. 2. Voyager 1/2 Jupiter encounter (plane of sky)

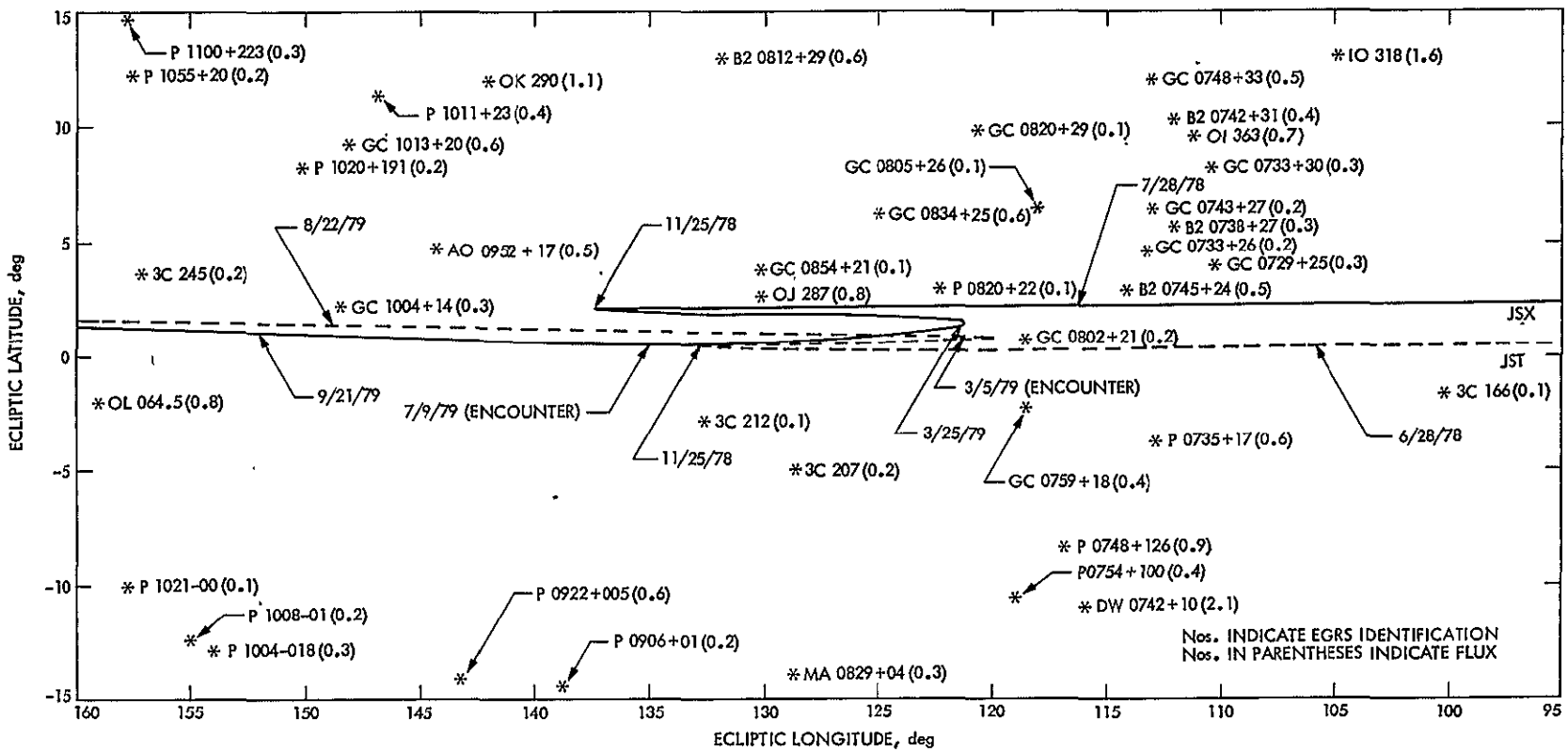

Fig. 3. Ecliptic VLBI source catalog for Voyager Δ VLBI demonstration

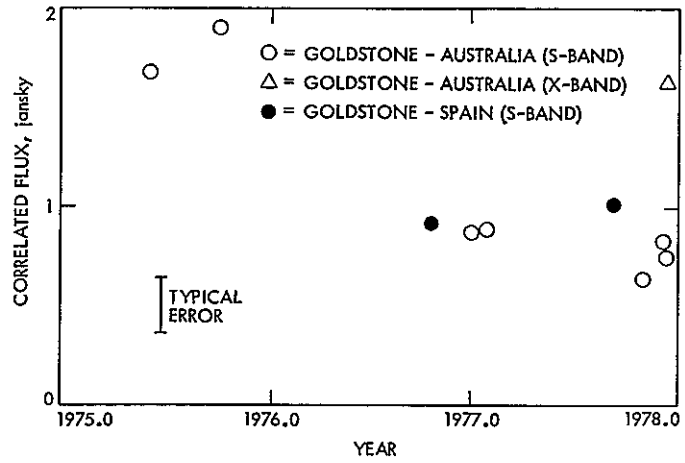

Fig. 4. OJ287 correlated flux history

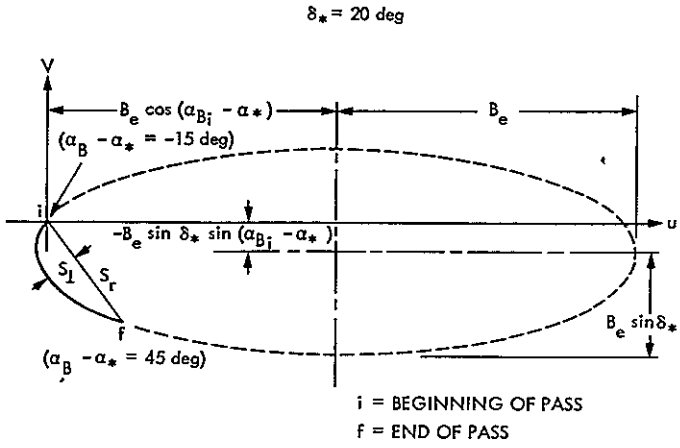

Fig. 5. Information content ellipse of fringe phase Δ VLBI

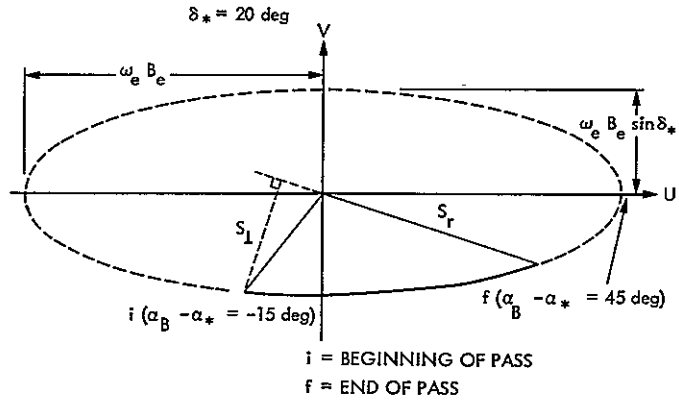

Fig. 6. Information content ellipse of fringe-phase-rate Δ VLBI

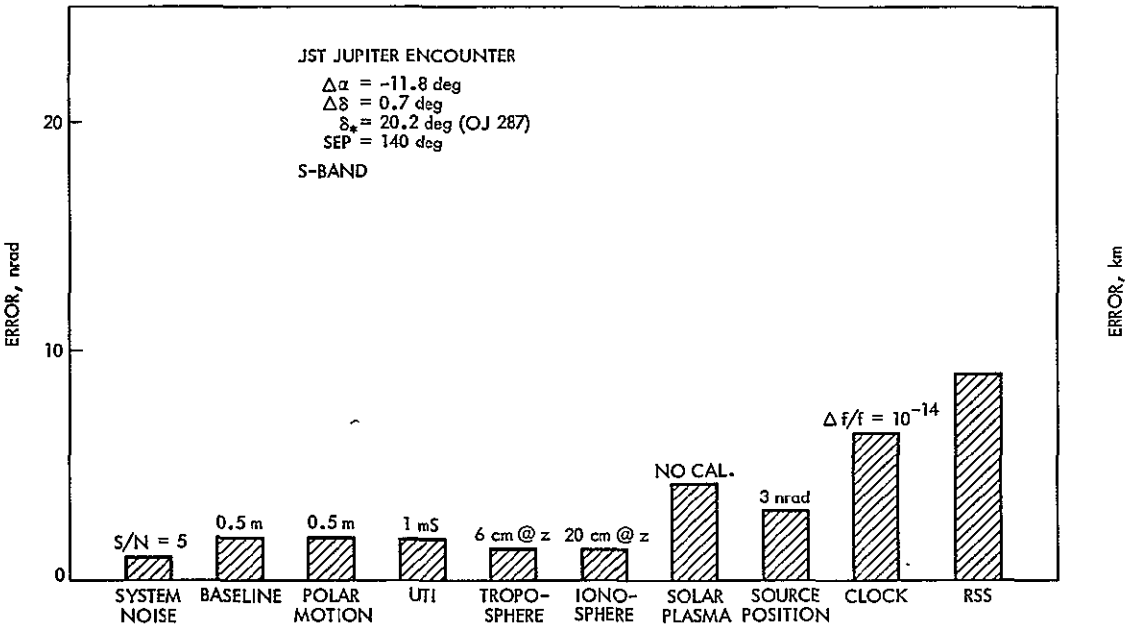

Fig. 7. Δ VLBI error sources during JST Jupiter encounter

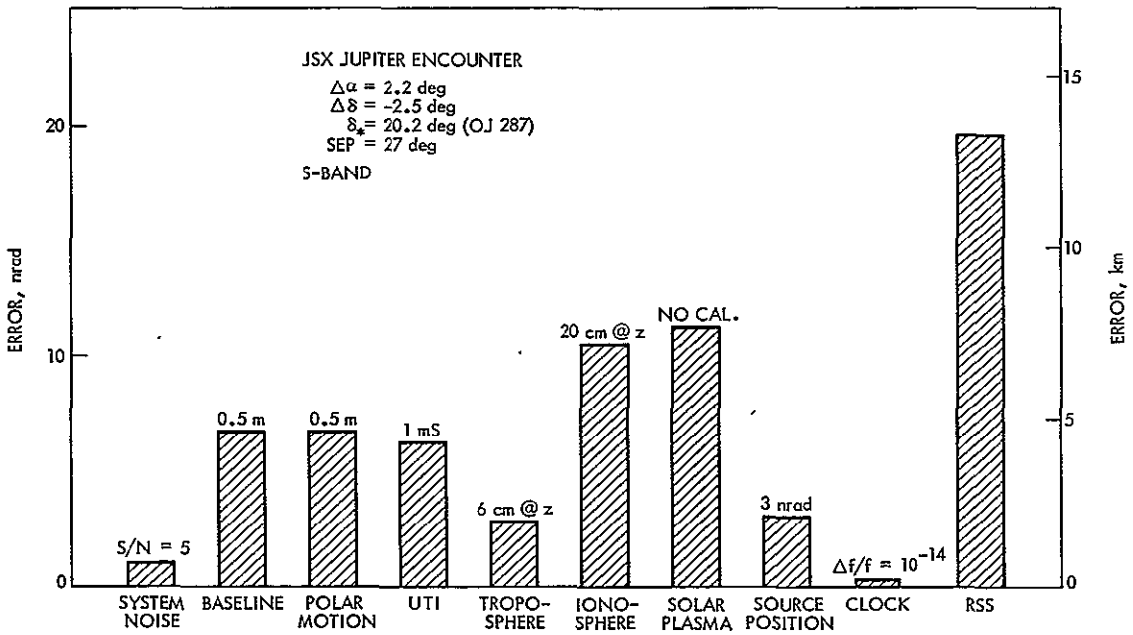

Fig. 8. Δ VLBI error sources during JSX Jupiter encounter

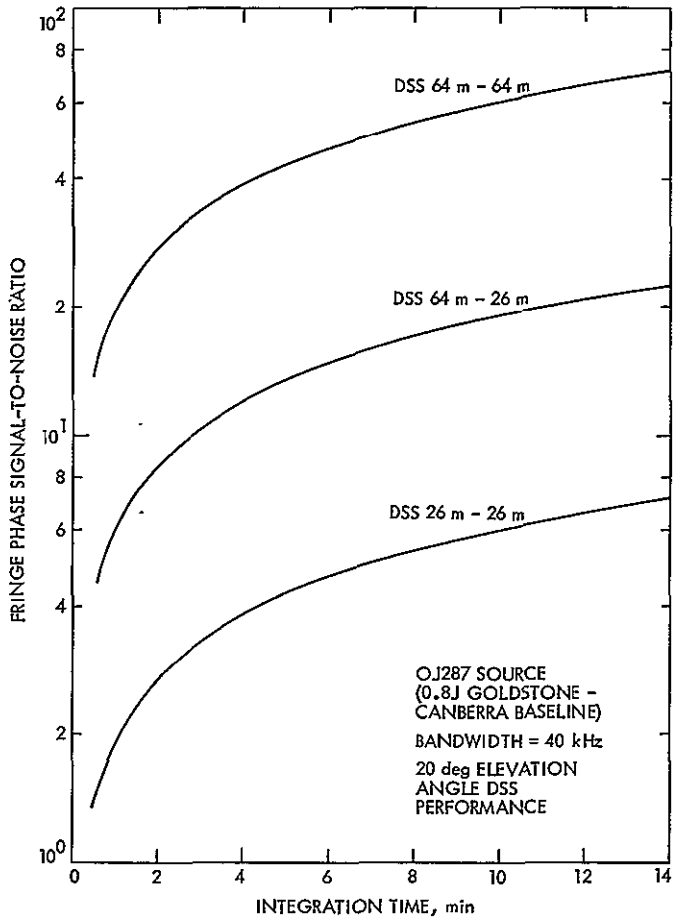

Fig. 9. S-band narrow-band EGRS Δ VLBI performance OJ287 source (0.8J Goldstone-Canberra baseline)

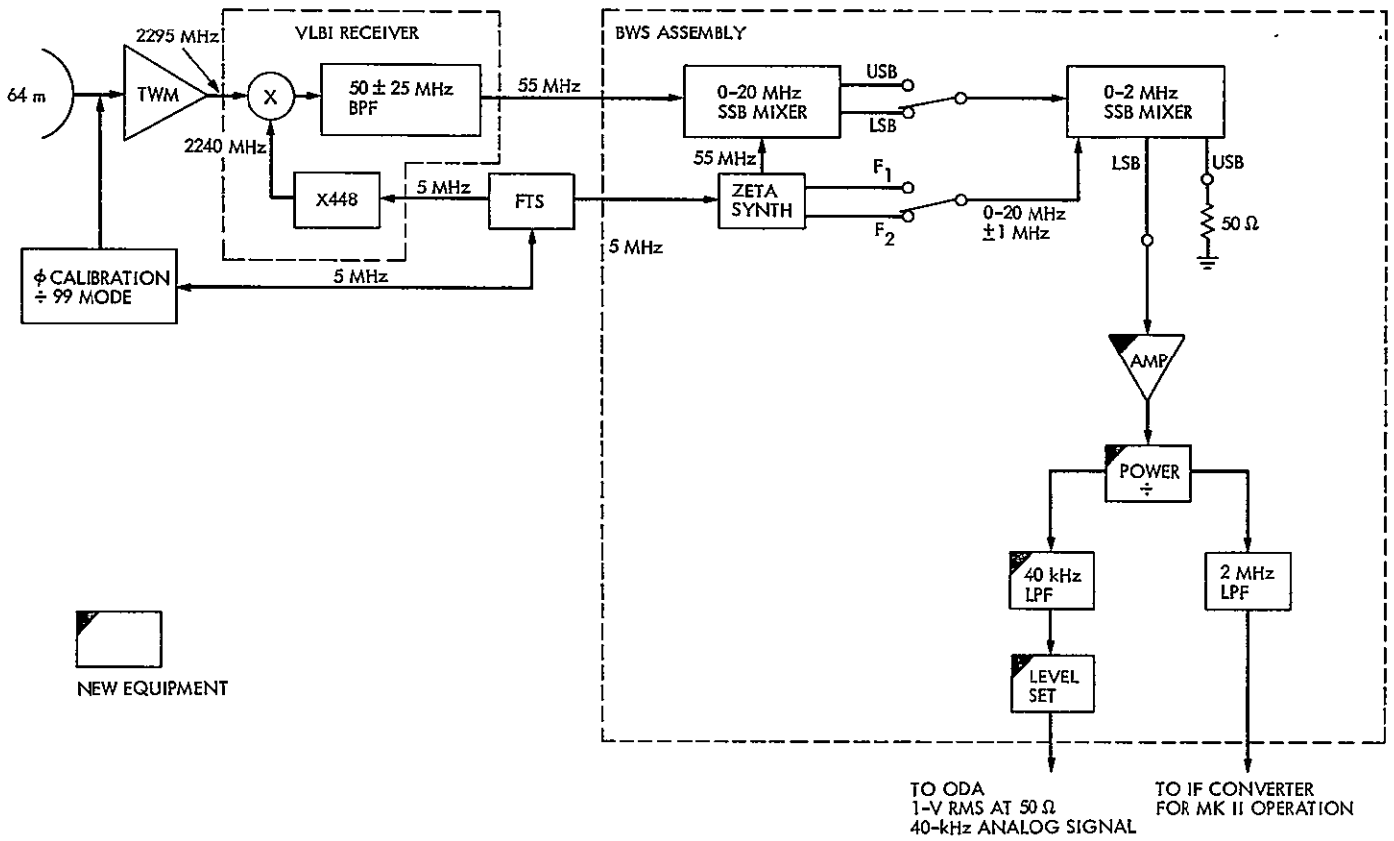

Fig. 10. DSS 14 S-band near real-time VLBI data acquisition system for Voyager demonstration

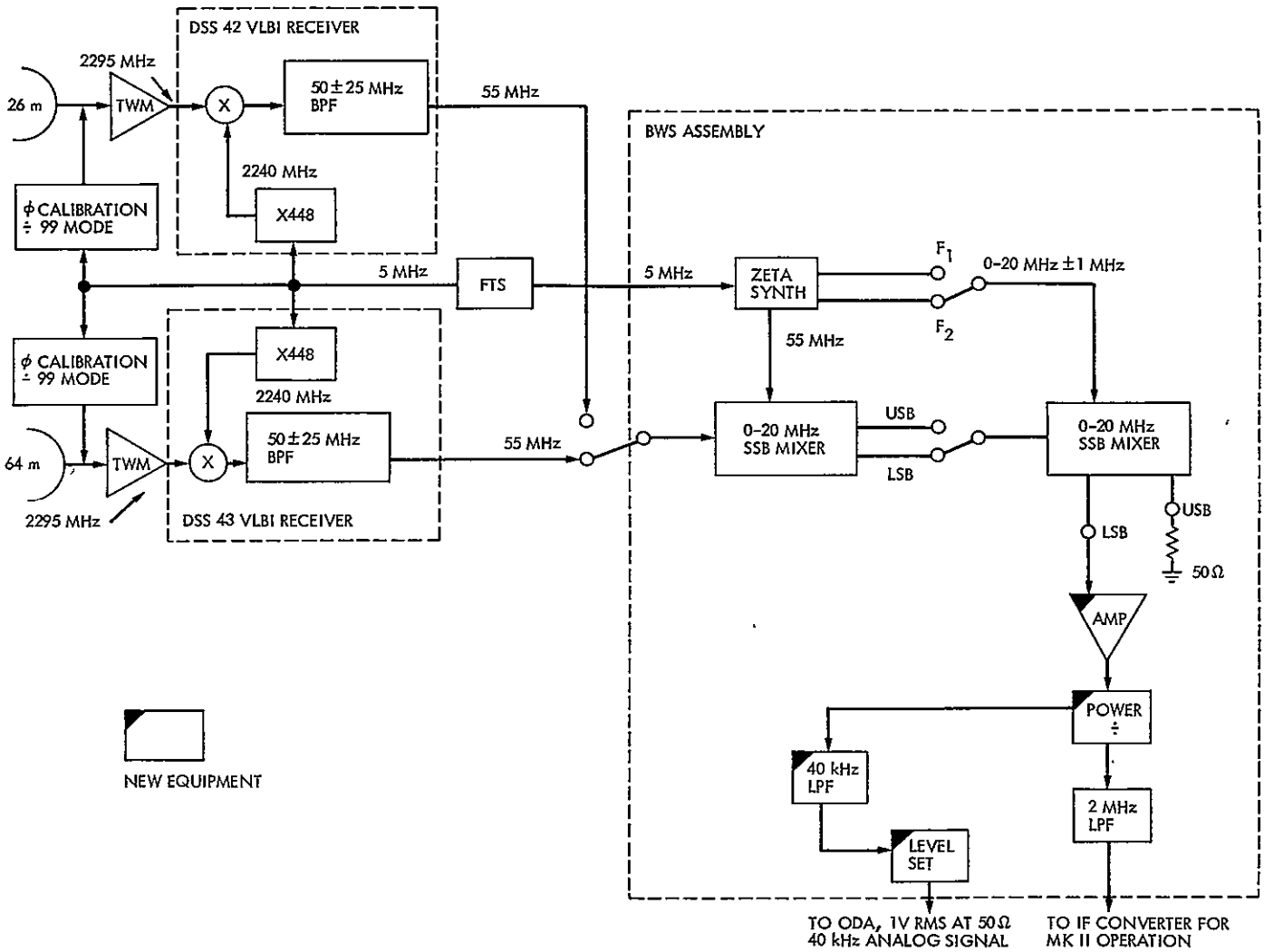

Fig. 11. Conjoint station S-band near real-time VLBI data acquisition system for Voyager demonstration

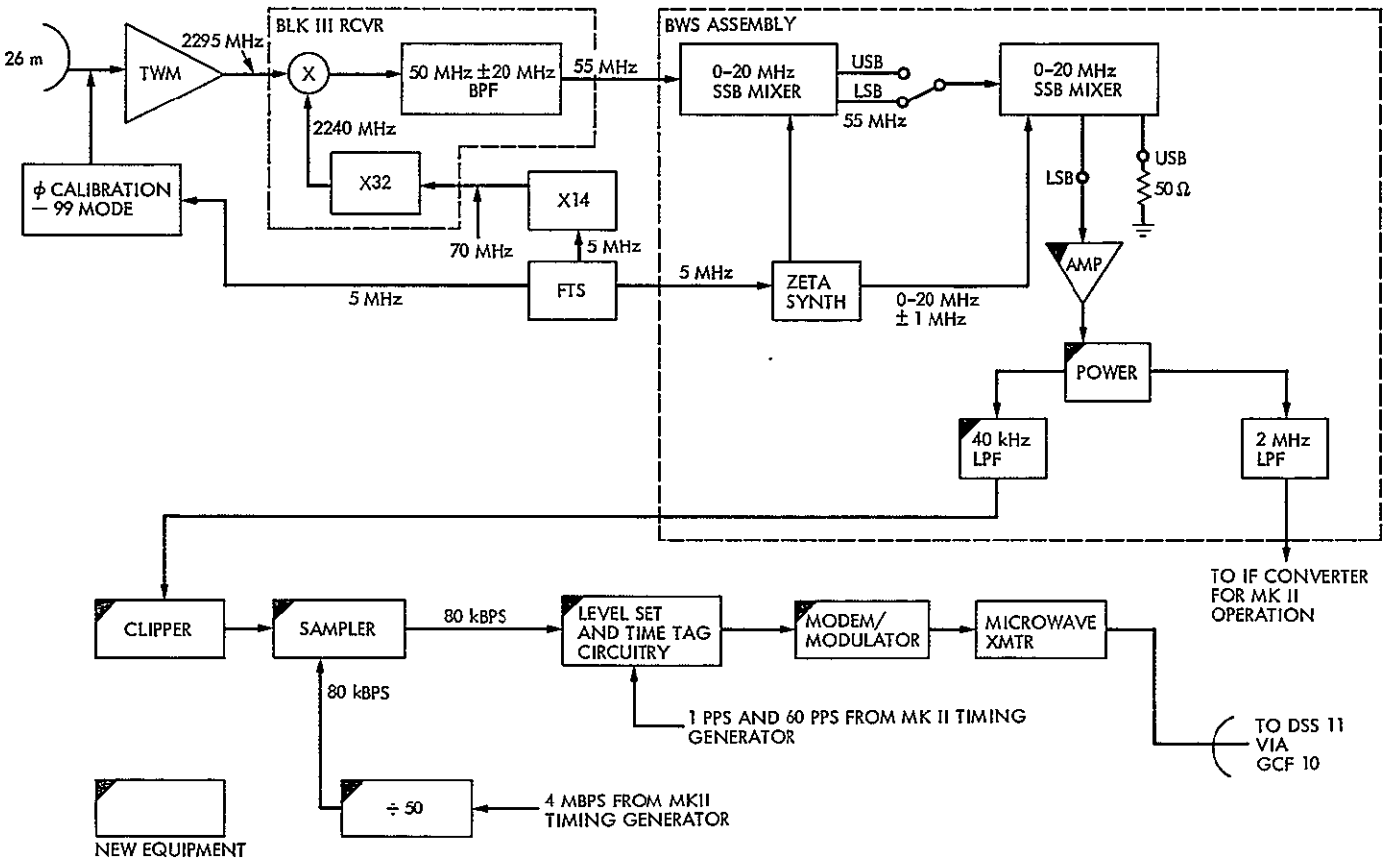

Fig. 12. DSS 13 S-band near real-time VLBI acquisition system for Voyager demonstration

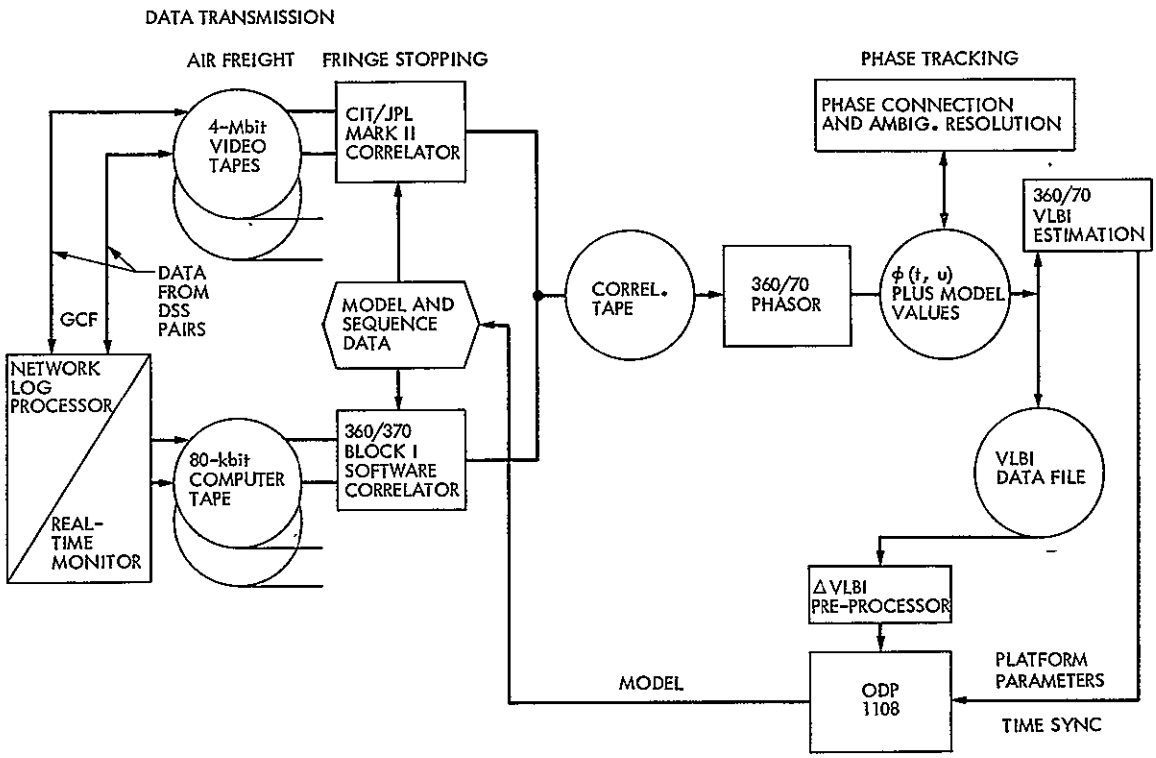

Fig. 13. VLBI navigation R & D processing system

The Lovasz Bound and Some Generalizations

R. J. McEliece, E. R. Rodemich, and H. C. Rumsey, Jr.
Communications Systems Research Section

In 1956, Shannon defined the zero-error capacity of a discrete memoryless channel as the largest rate at which information can be transmitted over the channel with zero error probability. He exhibited one particularly interesting channel with five inputs and outputs whose zero error capacity he could not compute. The problem of computing this capacity remained unsolved until very recently, when Lovasz computed it in an astonishing simple manner. We show that Lovasz' ideas, combined with some of our own, lead to an extremely powerful and general technique, which we phrase in terms of graph theory, for studying combinatorial packing problems. In particular, Delsarte's linear programming bound for cliques in association schemes appear as a special case of the Lovasz bound.

I. Introduction

Let $V = \{v_1, \cdots, v_N\}$ be a finite set with N elements, and let E be a collection of two-element subsets of V. Then the set G consisting of the singletons $\{v_i\}$ from V and the elements of E is called a *graph*[1] *on* V. The elements of V are called the *vertices*, and those of E, the *edges*, of G. Figure 1 depicts a particularly interesting graph for our purposes:

$$G = \{\{0\},\{1\},\{2\},\{3\},\{4\},\{0,1\},\{1,2\},\{2,3\},\{3,4\},\{4,0\}\}$$

Here vertices are represented by points in the plane, and edges by lines joining appropriate pairs of vertices. (For future reference we label this graph C_5.)

A subset $Y = \{y_1, \cdots, y_M\}$ of V is called an *independent set* if none of the pairs $\{y_i, y_j\}$, $i \neq j$, are edges of G. The cardinality of the largest possible independent set in G is denoted by $\alpha(G)$:

$$\alpha(G) = \max \{|Y| : Y \text{ independent set in } G\} \qquad (1)$$

For example, $\alpha(C_5) = 2$, and the set $Y = \{0, 2\}$, circled in Fig. 1, is a maximal independent set.

For any integer $n \geq 2$, we now define the n-th *direct power* of G, denoted by G^n, as follows: the vertex set of G^n is the Cartesian power V^n, i.e., the set of all N^n n-tuples $\mathbf{v} = (v_1, \cdots, v_n)$ from V. The edge set of G^n consists of all pairs $\{\mathbf{v}, \mathbf{v}'\}$ from V^n such that $\{v_i, v_i'\} \in G$ for all i. Note that if A is the incidence matrix of G, i.e., the following $N \times N$ matrix whose rows and columns are indexed by elements of V:

$$A(v, v') = \begin{cases} 1 \text{ if } \{v, v'\} \in G \\ 0 \text{ if not} \end{cases}$$

[1] More accurately, an undirected graph.

then the incidence matrix of G^n is the n-th direct (Kronecker) power of A.

The *capacity* of G, denoted by $\theta(G)$, is now defined as follows:

$$\theta(G) = \sup_n \alpha(G^n)^{1/n} \tag{2}$$

This notion[2] was introduced by Shannon (Ref. 8) in connection with the problem of finding the zero-error capacity of a discrete memoryless channel, and he developed techniques that enabled him to compute the capacities of many, but not all, graphs.

For example, Shannon showed that if there exists a mapping ϕ of V into an independent set of G such that $\{v, v'\} \notin G$ implies $\{\phi(v), \phi(v')\} \notin G$, then $\theta(G) = \alpha(G)$.

Also, he gave a *linear programming* upper bound on $\theta(G)$, as follows. Let P be a probability distribution on V, i.e., $P(v) \geq 0$, $\Sigma \{P(v) : v \in V\} = 1$. P is extended to subsets $X \subset V$ additively: $P(X) = \Sigma \{P(x): x \in X\}$. The subset X is a *clique* in G if $\{x, x'\} \in G$ for all $x, x' \in X$. Then Shannon proved that

$$\theta(G) \leq \lambda^{-1}$$

where

$$\lambda = \min_P \max \{P(X): X \text{ a clique}\} \tag{3}$$

the minimization in Eq. (3) being taken over all possible probability assignments.

These two results enabled Shannon to calculate the capacity of all graphs with five or fewer vertices, with the single exception of the graph C_5 of Figure 1. For C_5 his results yielded only

$$\sqrt{5} \leq \theta(C_5) \leq 5/2 \quad \text{(footnote 3)} \tag{4}$$

[2]Actually, Shannon's definition of capacity is the logarithm of our definition.

[3]The lower bound in Eq. (4) results from the fact that $\alpha(C_5^2) = 5$, a fact also established by Shannon.

Then, twenty-one years later, Lovasz (Ref. 5) established that $\theta(C_5) \leq \sqrt{5}$; this, combined with the lower bound in Eq. (4), shows that $\theta(C_5) = \sqrt{5}$. Let us briefly sketch Lovasz' technique for finding upper bounds on $\theta(G)$.

Lovasz defines an *orthonormal representation* of G as a realization of G in which the vertex set V is a set of vectors in a Euclidean vector space, with the property that $\{\mathbf{v}, \mathbf{v}'\} \in G$ iff $\mathbf{v} \cdot \mathbf{v}' \neq 0$.

If G has such an orthonormal representation, and if $\mathbf{b}$ is any unit vector, Lovasz' bound is as follows:

$$\theta(G) \leq (\min \{(\mathbf{v} \cdot \mathbf{b})^2 : \mathbf{v} \in V\})^{-1} \tag{5}$$

Lovasz applied Eq. (5) to the graph C_5 by considering an "umbrella" with five ribs $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_5\}$ of unit length. If the umbrella is opened to the point where the angle between alternate ribs is 90 deg, then $\{\mathbf{v}_1, \cdots, \mathbf{v}_5\}$ is an orthonormal representation of G in Euclidean 3-space. If the handle $\mathbf{b}$ is also a unit vector, then one easily shows that $\mathbf{b} \cdot \mathbf{v}_i = 5^{-1/4}$ for all i, and hence by Eq. (5), $\theta(C_5) \leq \sqrt{5}$.

Lovasz also derived many other consequences of Eq. (5) that we cannot summarize here. However, let us at least remark that several of the examples in Section IV also appear in Lovasz' paper, and were clearly derived by him earlier. (We will give references to Theorems in Lovasz' paper at the appropriate places in Section IV.)

The present paper arose from an attempt to put Lovasz' results into a general setting. We believe we have succeeded in doing this, but in our development orthonormal representations have entirely disappeared. Nevertheless, the bounds we shall derive (at least the bounds on $\theta(G)$) could all be derived from orthogonality graphs, and so we call these bounds *Lovasz bounds*. (Appendix A contains a proof of the equivalence of our methods and Lovasz'.)

In Section II, we give out derivation of the Lovasz bounds. It will be seen that computing these bounds for a fixed graph G amounts to solving a certain nonlinear programming problem.

In Section III, we demonstrate that if the graph G is highly symmetric (in a sense made precise there), this nonlinear programming problem becomes a *linear* programming [illegible]m.

In Section IV, we apply our general results to several examples. First we give a very simple bound on $\theta(G)$, which applies to any regular graph. Next, we compute the Lovasz bound for two infinite families of graphs, the cyclic graphs C_N

and the quadratic residue graphs Q_p, which are two different generalizations of the graph C_5. We do not succeed in computing the capacity of any of the graphs C_N with odd $N \geqslant 7$, but, for all prime $p \equiv 1 (\mathrm{mod}\ 4)$, we show that $\theta(Q_p) = \sqrt{p}$. Then we consider three special graphs: the Peterson graph, the icosahedron graph, and the dodecahedron graph. Also we consider an especially interesting regular graph on 7 vertices.

As out final example in Section IV, we show that Delsarte's (Ref. 3) linear programming bound for cliques in association schemes follows as a special case of our results.

II. The Lovasz Upper Bounds

In this section and the next, G will denote a fixed graph. We will continually be dealing with vectors and matrices whose components are indexed by the vertex set V of G. If $\mathbf{x}$ is such a vector, and $v \in V$, the v-th component of $\mathbf{x}$ will be denoted by $\mathbf{x}(v)$; if $\mathbf{A}$ is such a matrix, its (v, v')-th component will be denoted by $\mathbf{A}(v,v')$.

We will also be working with the *quadratic forms* associated with such vectors and matrices. If $\mathbf{x}$ is a (column) vector, and $\mathbf{A}$ a symmetric matrix, the quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is defined by

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{(v,v') \in V^2} \mathbf{x}(v)\, \mathbf{x}(v')\, \mathbf{A}(v, v') \tag{6}$$

We will always view $\mathbf{x}^T\mathbf{A}\mathbf{x}$ as a function of the components of the vector $\mathbf{x}$.

The following will no doubt appear quite trivial, and yet it is our main result. The remainder of the paper will be devoted merely to exploring its consequences.

Theorem 1:

Let $\mathbf{A}$ be a symmetric real matrix such that

$$\mathbf{A}(v, v') = 1 \text{ if } v = v'$$

$$\leqslant 0 \text{ if } \{v, v'\} \notin G$$

Then if $\mathbf{u} = (1, 1, \cdots, 1)$ denotes the all-ones vector,

$$\inf \{\mathbf{x}^T\mathbf{A}\mathbf{x} : \mathbf{x} \cdot \mathbf{u} = 1\} \leqslant \alpha(G)^{-1} \tag{7}$$

Proof:

Let $Y \subseteq V$ be a maximal independent set in G, i.e., $|Y| = \alpha(G)$. Define the vector $\mathbf{y}$ by

$$\mathbf{y}(v) = \alpha(G)^{-1} \text{ if } v \in Y$$

$$= 0 \qquad \text{if not}$$

Then clearly $\mathbf{y} \cdot \mathbf{u} = 1$, and by Eq. (6), $\mathbf{y}^T\mathbf{A}\mathbf{y} \leqslant \alpha(G)^{-1}$. This proves Eq. (7).

Let us denote by $\lambda(\mathbf{A})$ the value of the left side of Eq. (7)

$$\lambda(\mathbf{A}) = \inf \{\mathbf{x}^T\mathbf{A}\mathbf{x} : \mathbf{x} \cdot \mathbf{u} = 1\} \tag{8}$$

Theorem 1 gives an upper bound on $\alpha(G)$, viz., $\alpha(G) \leqslant \lambda(\mathbf{A})^{-1}$ provided $\lambda(\mathbf{A}) > 0$. Clearly in order to apply this bound we will need to know more about the function $\lambda(\mathbf{A})$. For future reference we now list some of its more important properties. Throughout we assume $\mathbf{A}$ is real and symmetric. (Proofs of these facts may be found in Appendix B.)

First of all, unless $\mathbf{A}$ is *positive semi-definite* (hereafter abbreviated p.s.d.), i.e., unless $\mathbf{x}^T\mathbf{A}\mathbf{x} \geqslant 0$ for all $\mathbf{x}$, $\lambda(\mathbf{A})$ is negative:

$$\lambda(\mathbf{A}) \geqslant 0 \text{ iff } \mathbf{A} \text{ is p.s.d.} \tag{9}$$

Assuming that $\mathbf{A}$ is p.s.d., let $\xi_1, \cdots, \xi_N$ be a complete orthonormal set of eigenvectors for $\mathbf{A}$, i.e., $\xi_i \cdot \xi_j = \delta_{i,j}$, and $\mathbf{A}\xi_j = \lambda_j \xi_j$. Since $\mathbf{A}$ is p.s.d. the eigenvalues λ_j are all nonnegative. Let $\mathbf{u} = u_1 \xi_1 + \cdots + u_N \xi_N$ be the expansion of $\mathbf{u}$ with respect to this basis.

Then

$$\lambda(\mathbf{A}) = 0 \text{ if } u_j \neq 0, \quad \lambda_j = 0 \text{ for some } j \tag{10}$$

$$\lambda(\mathbf{A}) = (\Sigma\, u_j^2/\lambda_j)^{-1} \text{ otherwise} \tag{11}$$

where the summation in Eq. (11) is extended only over subscripts j with $\lambda_j > 0$. (Alternatively Eq. (11) includes Eq. (10) as a special case if we extend the summation over all j and make the conventions that $u^2/0 = 0$ if $u = 0$, $= \infty$ if $u \neq 0$, and $\infty^{-1} = 0$.)

If u is itself an eigenvector of A with eigenvalue σ, the computation of $\lambda(A)$ is much simpler:

$$\lambda(A) = \sigma/N \text{ if } Au = \sigma u, \text{ and A is p.s.d.} \tag{12}$$

There is a useful dual formulation of the definition (8) for p.s.d. matrices:

$$\lambda(A) = \max \{\lambda : A - \lambda J \text{ is p.s.d.}\} \tag{13}$$

where in Eq. (13) J denotes the matrix of all ones.

Our last auxiliary result about $\lambda(A)$ is that it is *multiplicative* for p.s.d. matrices:

$$\lambda(A \times B) = \lambda(A)\,\lambda(B) \text{ if A and B are p.s.d.} \tag{14}$$

where $A \times B$ denotes the direct product of A and B.

According to Eq. (9) Theorem 1 will only give nontrivial information about $\alpha(G)$ if A is p.s.d. This leads us to define the following two sets of matrices.

Definition 1:

The set $\Omega(G)$ is defined as the set of all matrices $A = (A(v,v'))$ indexed by the vertices of G, satisfying

$$A \text{ is p.s.d.} \tag{15}$$

$$A(v, v) = 1 \text{ for all } v \in V \tag{16}$$

$$A(v, v') \leqslant 0 \text{ if } \{v, v'\} \notin G \tag{17}$$

Definition 2:

Similarly $\Omega_0(G)$ is the set of matrices satisfying Eqs. (15) and (16), with the condition (18) replaced with the stronger condition

$$A(v, v') = 0 \text{ if } \{v,v'\} \notin G \tag{18}$$

The significance of the class $\Omega(G)$ is obvious, in view of Theorem 1 and Eq. (9).

Theorem 2:

$$\alpha(G) \leqslant \lambda(A)^{-1} \text{ for all } A \in \Omega(G)$$

The significance of $\Omega_0(G)$ is given in the next theorem, which is essentially equivalent to the Lovasz bound Eq. (5). (For a proof of this equivalence, see Appendix A.)

Theorem 3:

$$\theta(G) \leqslant \lambda(A)^{-1} \quad \text{for all } A \in \Omega_0(G)$$

Proof:

The key to the proof is the fact that if $A \in \Omega_0(G)$, then the the n-th direct power $A^{[n]} = A \times A \times \cdots \times A$ (n factors) will belong to $\Omega_0(G^n)$.

To see this, observe first that $A^{[n]}$ is p.s.d., being a direct product of p.s.d. matrices. Next, if $v = (v_1, \cdots, v_n)$ and $v' = (v'_1, \cdots, v'_n)$ are vertices in G^n, then by definition of the direct product,

$$A^{[n]}(v, v') = \prod_{j=1}^{n} A(v_j, v_j') \tag{19}$$

It follows immediately that $A^{[n]}(v, v) = 1$ for all $v \in G^n$, since each of the factors on the right side of Eq. (19) will then be 1. If $\{v, v'\} \notin G^n$, there must exist at least one index j such that $\{v_j, v_j'\} \notin G$. Since $A \in \Omega_0(G)$, it follows from Eq. (18) that $A(v_j, v_j') = 0$ and hence that $A^{[n]}(v, v') = 0$ as well. Hence $A^{[n]}$ satisfies Eqs. (16), (17), and (18), and thus lies in $\Omega_0(G^n)$.

Since $\Omega_0(G^n) \subseteq \Omega(G^n)$, we may apply Theorem 2 to the matrix $A^{[n]}$, and conclude that $\alpha(G^n) \leqslant \lambda(A^{[n]})^{-1}$. But from Eq. (14), $\lambda(A^{[n]}) = \lambda(A)^n$. Hence $\alpha(G^n) \leqslant \lambda(A)^{-n}$ for all n, and so from the definition Eq. (2) of $\theta(G)$, we get $\theta(G) \leqslant \lambda(A)^{-1}$.

Motivated by Theorems 2 and 3, we now define the *Lovasz bounds* $\alpha_L(G)$ and $\theta_L(G)$:

$$\alpha_L(G) = \min \{\lambda(A)^{-1} : A \in \Omega(G)\} \tag{20}$$

$$\theta_L(G) = \min \{\lambda(A)^{-1} : A \in \Omega_0(G)\} \tag{21}$$

We have shown in Theorems 2 and 3 that $\alpha(G) \leqslant \alpha_L(G)$, $\theta(G) \leqslant \theta_L(G)$. Unfortunately, we know of no efficient algorithm for computing $\alpha_L(G)$ and $\theta_L(G)$ for an arbitrary graph. However, we will now show that one can use the symmetries of G to simplify the calculations somewhat. In Section III we

will extend these ideas and show that if G is highly symmetric, the bounds $\alpha_L(G)$ and $\theta_L(G)$ can be computed via linear programming.

A *symmetry* of the graph G is a permutation of the vertex set V that leaves the edge set E invariant. Thus if π is a permutation of V, it is a symmetry of G if and only if $\{\pi(v), \pi(v')\} \in E$ whenever $\{v, v'\} \in E$. Notice that a symmetry of G is also a symmetry of the *complementary graph* G', which has vertex set V and edge set E', the set of pairs not in E.

Let P be the group of symmetries of G, and let $E_1, \cdots, E_s$ be the orbits of E under the action of P. Similarly let $E'_1, \cdots, E'_t$ be the orbits of E'. We shall call two edges lying in the same orbit *equivalent edges*.

Now suppose that $A \in \Omega(G)$ or $\Omega_0(G)$. Then it is easy to see that the matrix $\bar{A}$ defined by

$$\bar{A}(v, v') = \frac{1}{|P|} \sum_{\pi \in P} A(\pi(v), \pi(v')) \tag{22}$$

also lies in the same set. Moreover, the matrix $\bar{A}$ has the property that if $\{v_1, v'_1\}$ and $\{v_2, v'_2\}$ are equivalent edges, then $\bar{A}(v_1, v'_1) = \bar{A}(v_2, v'_2)$. Additionally, we can show that $\lambda(\bar{A}) \geqslant \lambda(A)$. For if we denote by $\pi(A)$ the matrix with entries $A(\pi(v), \pi(v'))$, then for any value of λ,

$$\bar{A} - \lambda J = \frac{1}{|P|} \sum_{\pi \in P} (\pi(A) - \lambda J)$$

$$= \frac{1}{|P|} \sum_{\pi \in P} \pi(A - \lambda J)$$

If we let $\lambda = \lambda(A)$ then by Eq. (13) $A - \lambda J$ is p.s.d. and hence so is each $\pi(A - \lambda J)$. Thus $\bar{A} - \lambda J$ is p.s.d. and so by Eq. (13) $\lambda(\bar{A}) \geqslant \lambda(A)$.

Let us denote by B_j, $j = 1, 2, \cdots, s$ the edge incidence matrices for the edge orbits E_j:

$$B_j(v, v') = 1 \text{ if } \{v, v'\} \in E_j$$

$$= 0 \text{ if not}$$

Similarly we define the matrices B'_k, $k = 1, 2, \cdots, t$ as the edge incidence matrices for the edge orbits E'_k. Then, according to the preceding discussion, the matrix $\bar{A}$ can be expressed as a linear combination of these matrices, together with the $N \times N$ identity matrix I:

$$\bar{A} = I + \sum_{j=1}^{s} \mu_j B_j + \sum_{k=1}^{t} \mu'_k B'_k \tag{23}$$

We have thus shown that starting with any matrix A in $\Omega(G)$ (resp. $\Omega_0(G)$), we can construct a matrix $\bar{A}$ of the form Eq. (23) lying in the same class such that $\lambda(\bar{A}) \geqslant \lambda(A)$. What this means is that in the computation of the bounds $\alpha_L(G)$ and $\theta_L(G)$, we can safely restrict ourselves to matrices of the form of Eq. (23). More formally, we define

$$\overline{\Omega}(G) = \text{p.s.d. matrices of the form of Eq. (23)}$$

$$\text{with } \mu'_k \leqslant 0 \text{ for } k = 1, 2, \cdots, t$$

$$\overline{\Omega}_0(G) = \text{p.s.d. matrices of the form of Eq. (23)}$$

$$\text{with } \mu'_k = 0 \text{ for } k = 1, 2, \cdots, t$$

We have then the following computationally simpler definition of the Lovasz bounds:

$$\alpha_L(G) = \min \{\lambda(A)^{-1} : A \in \overline{\Omega}(G)\}$$

$$\theta_L(G) = \min \{\lambda(A)^{-1} : A \in \overline{\Omega}_0(G)\}$$

III. A Linear Programming Bound for α (G) and θ (G)

In this section we will show that if the graph G is sufficiently symmetric, the computation of the bounds $\alpha_L(G)$ and $\theta_L(G)$ can be greatly simplified.

The degree of symmetry we require is that the *incidence matrices* $\{B_j\}$, $\{B'_k\}$ in Eq. (23) *commute with each other*. That this is in fact a statement about the symmetry group of G can be seen as follows.

Suppose P is the symmetry group of G. With each $\pi \in P$ we associate the corresponding permutation matrix π^*:

$$\pi^*(v, v') = 1 \text{ if } \pi(v) = v'$$

$$= 0 \text{ if not}$$

Naturally the edge orbits $\{E_j\}$, $\{E'_k\}$ are left invariant by the symmetries $\pi \in P$; in terms of the corresponding incidence matrices, this can be expressed as

$$\pi^* \mathbf{B}_j = \mathbf{B}_j \pi^*, \quad \text{all } j = 1, 2, \cdots, s, \pi \in P \tag{24}$$

$$\pi^* \mathbf{B}'_k = \mathbf{B}'_k \pi^*, \text{ all } k = 1, 2, \cdots, t, \pi \in P$$

Now let P^* denote the group of all permutation matrices corresponding to the permutations in P, and let $Z(P^*)$ be the *centralizer ring* of P^*, i.e., the set of all matrices that commute with all $\pi^* \in P^*$. According to Eq. (24), the matrices $\{\mathbf{B}_j\}$, $\{\mathbf{B}'_j\}$ all belong to $Z(P^*)$.

If the ring $Z(P^*)$ were known to be commutative, then it would follow immediately that the matrices $\mathbf{B}_j, \mathbf{B}'_k$ commute with each other. Fortunately, this frequently turns out to be the case. Indeed it can be shown (Ref. 10, Chapter 5) that if P is transitive, then $Z(P^*)$ is commutative if and only if the complex representation of P afforded by the matrix group P^* decomposes into a sum of inequivalent irreducible representations. In particular, if P contains a transitive abelian subgroup, or if for any pair (v, v') of distinct vertices there is an element of P that exchanges v and v', this condition will be satisfied.

Motivated by the preceding discussion, we now place our results in the following general setting.

Let V be a finite set containing N elements, and let $\{E_1, E_2, \cdots, E_n\}$ be a partition of the collection E of all two-element subsets of V. For each $j = 1, 2, \cdots, n$ let $\mathbf{A}_j$ be the incidence matrix for E_j:

$$\mathbf{A}_j(v, v') = 1 \text{ if } \{v, v'\} \in E_j$$

$$= 0 \text{ if not}$$

Let $\mathbf{A}_0$ denote the $N \times N$ identity matrix. Assume that the matrices $\{\mathbf{A}_j : j = 0, 1, \cdots, n\}$ commute with each other. In summary, the assumptions are that the $\mathbf{A}_j$'s are (0, 1) matrices satisfying

$$\mathbf{A}_0 = I, \sum_{j=0}^{n} \mathbf{A}_j = J \tag{25}$$

$$\text{Each } \mathbf{A}_j \text{ is symmetric} \tag{26}$$

$$\mathbf{A}_j \mathbf{A}_k = \mathbf{A}_k \mathbf{A}_j \quad \text{for all } j, k = 0, 1, \cdots, n \tag{27}$$

If C is a fixed subset of $\{1, 2, \cdots, n\}$, let G_C be the graph with vertex set V and edge set

$$E_C = \bigcup_{j \in C} E_j \tag{28}$$

Our goal is to give "linear programming" upper bounds on $\alpha(G_C)$ and $\theta(G_C)$ (Theorems 4 and 5, below). To state these results, however, we need some preliminary discussion.

Notice that because of Eqs. (25) and (27), each matrix $\mathbf{A}_j$ commutes with $\mathbf{J}$, the all ones matrix, and hence the $n + 2$ matrices $\mathbf{J}, \mathbf{A}_0, \cdots, \mathbf{A}_n$ all commute with each other. Since these matrices are moreover symmetric, and hence diagonalizable, it follows from a known theorem of linear algebra (see Ref. 4, Chapter 6, Theorem 4), that there exists a set $\{\xi_m\}_{m=1}^{N}$ of linearly independent *simultaneous eigenvectors* for these $n + 2$ matrices.

In particular the ξ_m's are eigenvectors for J:

$$\mathbf{J}\xi_m = \lambda_m \xi_m, \quad m = 1, 2, \cdots, N \tag{29}$$

$\{\lambda_m\}$ being the set of eigenvalues for $\mathbf{J}$. But $\mathbf{J}$ has only the eigenvalues $\{0, N\}$, and a simple calculation shows that if $\mathbf{J}\xi = N\xi$, then ξ must be a scalar multiple of $\mathbf{u}$, the all ones vector. Thus we may assume that $\xi_1 = \mathbf{u}$.

Now for each j, m, define the eigenvalues $\lambda_{j,m}$ by

$$\mathbf{A}_j \xi_m = \lambda_{j,m} \xi_m \quad j = 0, 1, \cdots, n \tag{30}$$

$$m = 1, 2, \cdots, N$$

We come now to our "linear programming" bounds for $\alpha(G)$ and $\theta(G)$.

Theorem 4:

Let $\mu_1, \mu_2, \cdots, \mu_n$ be real numbers such that

$$\mu_j \leqslant 0 \quad \text{if } j \notin C \tag{31}$$

and

$$1 + \sum_{j=1}^{n} \mu_j \lambda_{j,m} \geqslant 0, \quad m = 1, 2, \cdots, N \tag{32}$$

Then

$$\alpha(G_C) \leqslant N \Big/ \left(1 + \sum_{j=1}^{n} \mu_j \lambda_{j,1}\right)$$

Theorem 5:

Let $\mu_1, \cdots, \mu_n$ satisfy Eq. (32), and also

$$\mu_j = 0 \text{ if } j \notin C \tag{33}$$

Then

$$\theta(G_C) \leqslant N \Big/ \left(1 + \sum_{j=1}^{n} \mu_j \lambda_{j,1}\right)$$

Proofs:

For the given constants $\{\mu_j\}$, define

$$\mathrm{A} = \mathrm{I} + \sum_{j=1}^{n} \mu_j \mathrm{A}_j$$

Clearly the vectors $\{\xi_m\}$ are eigenvectors for A, since

$$\mathrm{A}\,\xi_m = \xi_m + \sum_{j=1}^{n} \mu_j \mathrm{A}_j \xi_m$$

$$= \left(1 + \sum_{j=1}^{m} \mu_j \lambda_{j,m}\right) \xi_m$$

Furthermore, the hypothesis Eq. (32) ensures that the eigenvalues $\{1 + \Sigma\, \mu_j\, \lambda_{j,m}\}$ of A are nonegative, and hence that A is p.s.d. The conditions (31) and (33) now imply that the matrix A belongs to $\Omega\,(G_C)$ or $\Omega_0(G_C)$. Thus by Theorems 2 and 3, we get $\alpha(G_C) \leqslant \lambda(\mathrm{A})^{-1}$, $\theta(G_C) \leqslant \lambda(\mathrm{A})^{-1}$.

To compute $\lambda(\mathrm{A})$ observe that $\xi_1 = \mathrm{u}$ is an eigenvector for A, with corresponding eigenvalue

$$1 + \sum_{j=1}^{n} \mu_j \lambda_{j,1}$$

and so by Eq. (12), $\lambda(\mathrm{A}) = (1 + \Sigma\, \mu_j\, \lambda_{j,1})/N$. Theorems 4 and 5 now follow.

To get the best possible bounds of the kind given in Theorems 4 and 5, we are essentially required to maximize the linear function

$$\sum_{1}^{n} \mu_j \lambda_{j,1}$$

subject to the linear constraints (31) or (33), and (32). This is a linear programming problem (once the eigenvectors and eigenvalues of the matrices A_j are known); and hence we have succeeded in showing that *the Lovasz bounds $\alpha_L(G)$ and $\theta_L(G)$ can be computed via linear programming, provided the incidence matrices B_j, B'_k of the edge orbits E_j, E'_k commute.*

IV. Some Applications

In this section we will describe a few of the many possible applications of the preceding results. In particular we will obtain Lovasz' original result on the capacity of C_5, and Delsarte's linear programming bound for cliques in association schemes.

A. Example 1. Regular Graphs

A graph G is said to be *regular* if the number of edges containing a given vertex v is a constant r, independent of v, called the *valence* of G.

A consequence of the regularity of G is that the incidence matrix **B** corresponding to the edge set E commutes with J: **JB = BJ**. Obviously **B** also commutes with the identity matrix **I**, and hence also with $\mathbf{B}' = \mathbf{J} - \mathbf{I} - \mathbf{B}$, which is the edge incidence matrix of the complementary graph. We are thus in a position to apply Theorem 5. Omitting the straightforward details, the result is

$$\theta(G) \leqslant \frac{N}{1 + r/|\lambda_{\min}|} \tag{34}$$

where $\lambda_{\min}$ is the smallest eigenvalue of **B** (which is necessarily negative, unless E is empty.)

Furthermore, if in addition the group of symmetries of G permutes the edges transitively, it follows from our results that $\theta_L(G)$ is equal to the right side of Eq. (34), i.e., Eq. (34) is the best possible bound of this type. (The bound Eq. (34) is equivalent to Lovasz' Theorem 9.)

B. Example 2. The Graphs C_N

Denote by C_N the cyclic graph on N vertices, i.e, $V = \{0, 1, \cdots, N-1\}$, $E = \{\{i, i+1\} : i = 0, 1, \cdots, N-1\}$, with indices taken mod. N. These graphs are all regular, and indeed the cyclic group of order N permutes the edges transitively, so we may compute $\theta_L(C_N)$ by the formula (34).

To find the eigenvalues of the incidence matrix **B** in this case, observe that the vectors $\mathbf{x}(\zeta) = (1; \zeta, \cdots, \zeta^{N-1})$, where ζ is any complex N-th root of unity, form an independent set of eigenvectors for **B**, and indeed

$$\mathbf{B}\mathbf{x}(\zeta) = (\zeta + \zeta^{-1})\mathbf{x}(\zeta)$$

Hence the eigenvalues of **B** are $\{(\zeta + \zeta^{-1})\} = \{2 \cos(2\pi k/N) : k = 0, 1, \cdots, \lfloor N/2 \rfloor\}$. The least member of this set is clearly -2 if N is even, and $2\cos(\pi - \pi/N) = -2\cos(\pi/N)$ if N is odd and $\geqslant 3$. Thus by Eq. (34).

$$\theta(C_N) \leqslant N/2, N \text{ even}$$

$$\leqslant N/(1 + (\cos \pi/N)^{-1}), \quad N \text{ odd}, \geqslant 3 \tag{35}$$

For N even, or $N = 3$, this bound is sharp; but these results are quite elementary and were already known to Shannon.

For odd $N \geqslant 5$, however, the bounds are nontrivial. (They appear as Corollary 5 in Lovasz.) With $N = 5$, for example, we have $\theta(C_5) \leqslant \sqrt{5}$.

For odd $N \geqslant 7$, the upper and lower bounds on $\theta(C_N)$ do not agree. Here is a table of the upper bound (35) vs the best known lower bounds (Ref. 1) for odd $7 \leqslant N \leqslant 19$:

$$7^{3/5} = 3.21410 \leqslant \theta(C_7) \leqslant 3.31767$$

$$81^{1/3} = 4.32675 \leqslant \theta(C_9) \leqslant 4.36009$$

$$148^{1/3} = 5.28957 \leqslant \theta(C_{11}) \leqslant 5.38630$$

$$247^{1/3} = 6.27431 \leqslant \theta(C_{13}) \leqslant 6.40417$$

$$380^{1/3} = 7.24316 \leqslant \theta(C_{15}) \leqslant 7.41715$$

$$4913^{1/4} = 8.37214 \leqslant \theta(C_{17}) \leqslant 8.42701$$

$$7666^{1/4} = 9.35712 \leqslant \theta(C_{19}) \leqslant 9.43477$$

C. Example 3. The Quadratic Residue Graphs

Let $p \equiv 1 \pmod 4$ be a prime. The graph Q_p has vertex set $V = \{0, 1, \cdots, p-1\}$, and edge set $E = \{\{v, v'\} : v - v' \text{ is a quadratic residue (q.r.) mod } p\}$. (Note that Q_5 is isomorphic to the pentagonal graph C_5.) Q_p is regular with valence $(p-1)/2$. The edge-incidence matrix is given by

$$\mathbf{B}_p(v - v') = 1 \text{ if } v - v' \text{ is a q.r. } (\text{mod } p)$$

$$0 \text{ if not}$$

One easily verifies that the p vectors $\mathbf{x}(\xi) = (1, \xi, \cdots, \xi^{p-1})$, where ξ is any complex p-th root of unity, are eigenvectors for $\mathbf{B}_p$, and that the eigenvalue associated with $\mathbf{x}(\xi)$ is $\Sigma\{\xi^a : a \text{ is a q.r.}\}$. It is well known (see Ref. 9, Section 11, for example) that these sums assume only the three distinct values $(p-1)/2$, $(-1 \pm \sqrt{p})/2$. Hence the least eigenvalue of $\mathbf{B}_p$ is $(-1 - \sqrt{p})/2$, and Eq. (34) yields $\theta(Q_p) \leqslant \sqrt{p}$. On the other hand, if b is a fixed quadratic nonresidue (mod p), the p ordered pairs (v, bv), $v \in V$ form an independent set in Q_p^2, and hence $\alpha(Q_p^2) \geqslant p$. These two inequalities establish the fact that $\theta(Q_p) = \sqrt{p}$, for all $p \equiv 1 \pmod 4$. Because of this result, it is clear that the graphs Q_p form a more satisfactory generalization of the pentagon of Fig. 1 than the graphs C_N. (These graphs do not

appear in Lovasz' paper. But he does show in Theorem 12 that if G is self-complementary, and if the symmetry group of G is transitive on the vertics, then $\theta(G) = \sqrt{|V|}$. This example is thus an explicit case of Lovasz' Theorem 12.)

D. Example 4. Some Miscellaneous Edge-Transitive Graphs

Here we will apply the bound (34) to three particularly interesting graphs. In each case there is only one equivalence class of edges, so that the bounds obtained are all equal to $\theta_L(G)$. In each case $\theta_L(G)$ is strictly less than any bound that could be obtained by Shannon's techniques.

The Peterson Graph (Fig. 2) is a regular graph with $N = 10$, $r = 3$. The minimum eigenvalue here turns out to be -2, so (34) yields $\theta(G) \leqslant 4$. On the other hand $\alpha(G) = 4$ (note the four circled vertices in Fig. 2), and so $\theta(G) = 4$. (This result is a special case of Lovasz' Theorem 13.)

The Icosahedron Graph (Fig. 3) has $N = 12$, $r = 5$; its vertices and edges are formed from those of the regular icosahedron. Here the minimum eigenvalue is $-\sqrt{5}$, and so from (34), $\theta(G) \leqslant 3(\sqrt{5} - 1) = 3.7082$. On the other hand $\alpha(G) = 3$, so we have $3 \leqslant \theta(G) \leqslant 3.7082$.

The Dodecahedron Graph (Fig. 4) is the graph of the regular dodecahedron, with $N = 20$, $r = 3$. Here $\lambda_{min} = -\sqrt{5}$ also; hence (34) gives $\theta(G) \leqslant 15\sqrt{5} - 25 = 8.5410$. On the other hand, $\alpha(G) = 8$, as shown. Thus $8 \leqslant \theta(G) \leqslant 8.5410$.

E. Example 5. A Special Graph on 7 Vertices

Consider the graph depicted in Fig. 5. This graph is regular with $N = 7$, $r = 4$, and minimum eigenvalue $= 4 \cos 6\pi/7 \cos 2\pi/7 = -2.2470$, and hence by (34), $\theta(G) \leqslant 2.5178$. However, under the action of the symmetries of G, there are two equivalence classes of edges: those of type $\{i, i+1\}$, and those of type $\{i, i+2\}$, modulo 7. In this case the bound of (34) is strictly larger than the Lovasz bound $\theta_L(G)$; in order to compute $\theta_L(G)$, we must apply Theorem 5 directly.

Thus let A_0 denote the 7×7 identity matrix: A_1, the incidence matrix for edges of type $\{i, i+1\}$; A_2, for edges of type $\{i, i+2\}$; and A_3, for edges of type $\{i, i+3\}$ (which are not edges of G.)

One easily verifies that the matrices $\{A_0, A_1, A_2, A_3\}$ satisfy conditions (25) through (27). If $C = \{1,2\}$ the graph G_C is the graph of Fig. 5. Also, the 7 vectors of the form $x(\zeta) = (1, \zeta, \cdots, \zeta^6)$, where ζ is a complex 7-th root of unity, form a set of common eigenvectors for the A's:

$$A_0\, x(\zeta) = x(\zeta)$$

$$A_1\, x(\zeta) = (\zeta + \zeta^{-1})\, x(\zeta)$$

$$A_2\, x(\zeta) = (\zeta^2 + \zeta^{-2})\, x(\zeta)$$

$$A_3\, x(\zeta) = (\zeta^3 + \zeta^{-3})\, x(\zeta)$$

Thus according to Theorem 5, if μ_1 and μ_2 satisfy $1 + \mu_1(\zeta + \zeta^{-1}) + \mu_2(\zeta^2 + \zeta^{-2}) \geqslant 0$ for all 7-th roots of unity ζ, then $\theta(G) \leqslant 7/(1 + 2\mu_1 + 2\mu_2)$. To get the best possible such bound we must maximize the function $\mu_1 + \mu_2$ subject to the above set of inequalities. This is easily done by hand[4], and we get the largest possible value with $\mu_1 = 0.8020$, $\mu_2 = 0.3569$. The resulting bound is $\theta(G) \leqslant 2.1098$. (This graph is the complement of the cyclic graph C_7; we denote this by writing $G = C_7'$. Lovasz' Theorem 8 asserts that if G is any graph with a vertex-transitive symmetry group, then $\theta_L(G)\, \theta_L(G') = N$. It thus follows from Example 2 that $\theta_L(C_N') = 1 + (\cos \pi/N)^{-1}$ for odd N. We have included this alternate derivation only to illustrate a nontrivial example of our linear programming approach.)

F. Example 6. Delsarte's Linear Programming Bounds

Delsarte (Ref. 3) obtained linear programming bounds for cliques in association schemes. Here we sketch a demonstration that these bounds are subsumed under our Theorem 4. (Recently Schrijver, (Ref. 7), has obtained similar results.)

Let V be a finite set, and let $R = \{R_0, R_1, \cdots, R_n\}$ be a family of $n+1$ subsets of the Cartesian square V^2. The R_j are *relations* on V and can be described by their incidence matrices

$$A_j(v, v') = 1 \text{ if } (v, v') \in R_j$$

$$= 0 \text{ if not}$$

[4] There are really only three inequalities to consider, viz. those with $\zeta = \exp(2\pi ik/7)$, $k = 1, 2, 3$.

The pair (V, R) is called a (symmetric) *association scheme* if the following conditions are satisfied:

(1) R is a partition of V^2, and R_0 is the diagonal, i.e., $R_0 = \{(v, v) : v \in V\}$.

(2) The relations $\{R_j\}$ are symmetric, i.e., $(v, v') \in R_j$ implies $(v', v) \in R_j$.

(3) There exist numbers $p_{i,j}^{(k)} = p_{j,i}^{(k)}$ such that for all $i, j = 0, 1, \cdots n$,

$$A_i A_j = \sum_{k=0}^{n} p_{i,j}^{(k)} A_k$$

If M is a subset of $\{0, 1, \cdots, n\}$ with $0 \in M$, a nonempty subset $Y \subseteq V$ is called an *M-clique with respect to R* if it satisfies

$$R_j \cap Y^2 = \phi \text{ for all } j \notin M$$

Delsarte (Ref. 3) gives an upper bound on the number of points in an M-clique, which is the value of a certain linear program.

But we can equally apply our Theorem 4 to the same problem, for the matrices $\{A_j\}$ of the association scheme certainly satisfy conditions (25) through (27) (note that since $p_{i,j}^{(k)} = p_{j,i}^{(k)}$, condition (3) is considerably stronger than (27).) If we let $C = \{j : j \notin M\}$, then an M-clique as defined above is an independent set in G_C, and so the upper bound of Theorem 4 is an upper bound on the cardinality of any M-clique in G_C. One can in fact show that this is the same bound as Delsarte's. Hence Theorem 4 is more general than Delsarte's bound, since it applies to many cases that are not association schemes.

Appendix A

Equivalence of Theorem 3 with Lovasz' Bound

Given an orthonormal representation of the graph G, define the matrix A by

$$A(v, v') = v \cdot v'$$

Clearly $A \in \Omega_0(G)$. If b is a unit vector and if $\lambda = \min \{(v \cdot b)^2 : v \in V\}$, then $A - \lambda J$ is p.s.d. This is because we can write $A - \lambda J = B + C$, where the matrices B and C are defined by

$$B(v, v') = (v - (v \cdot b)b) \cdot (v' - (v' \cdot b)b)$$

$$C(v, v') = (v \cdot b)(v' \cdot b) - \lambda$$

B is p.s.d., since it is the matrix of inner products of a set of vectors. C is also p.s.d., since if $\{x(v) : v \in V\}$ is any set of real numbers,

$$x^T C x = (\sum_v x(v)(v \cdot b))^2 - \lambda(\sum x(v))^2$$

$$\geqslant 0$$

since $(v \cdot b) \geqslant \sqrt{\lambda}$ for all v. Thus $A - \lambda J$, being the sum of two p.s.d. matrices, is also p.s.d.

Since $A \in \Omega_0(G)$ and $A - \lambda J$ is p.s.d., it now follows from (13) that $\lambda(A) \geqslant \lambda$ and hence from Theorem 3 that $\theta(G) \leqslant \lambda^{-1} = (\min \{(v \cdot b)^2 : v \in V\})^{-1}$. Thus Theorem 3 implies the Lovasz bound (1.5).

Conversely, if $A \in \Omega_0(G)$, let $\lambda = \lambda(A)$. Then $A - \lambda J$ is p.s.d., and from a known theorem (Ref. 2), Chapter 9), there exists a matrix B such that $B^T B = A - \lambda J$. Letting $\{w(v) : v \in V\}$ denote the column vectors of B, we have

$$w(v) \cdot w(v') = A(v, v') - \lambda$$

$$= 1 - \lambda \text{ if } v = v'$$

$$= -\lambda \text{ if } \{v, v'\} \notin G.$$

Now, let t be a vector orthogonal to all the $w(v)$'s with $|t|^2 = \lambda$ (increase the dimension of the underlying space, if necessary), and define

$$x(v) = w(v) + t$$

The $x(v)$'s are unit vectors, since

$$x(v) \cdot x(v') = A(v, v')$$

The orthogonality graph defined by the x's is thus a subgraph (same vertex set, a subset of edges) G' of G.

Furthermore, if we define the unit vector b:

$$b = \frac{t}{|t|}$$

we have $x(v) \cdot b = |t| = \sqrt{\lambda}$ for all $v \in V$. Hence by Lovasz' bond (5) $\theta(G') \leqslant \lambda^{-1} = \lambda(A)^{-1}$. But clearly $\theta(G) \leqslant \theta(G')$, and so Lovasz' result implies our Theorem 3.

Appendix B

Proof of Assertions (9) through (14)

Recall that A is a real symmetric $N \times N$ matrix, and that

$$\lambda(A) = \inf \{x^T Ax : x \cdot u = 1\} \tag{B-1}$$

where $u = (1, 1, \cdots, 1)$ is the all ones vector. According to the Principal Axis Theorem, (Ref. 2), (Chapters 9 and 10), there exists a set $\{\xi_1, \cdots, \xi_N\}$ of N orthogonal eigenvectors of A:

$$\xi_i \cdot \xi_j = 1 \text{ if } i = j \tag{B-2}$$

$$= 0 \text{ if } i \neq j$$

$$A\xi_j = \lambda_j \xi_j, \ = 1, 2, \cdots, N \tag{B-3}$$

Thus, if $x = x_1 \xi_1 + \cdots + x_N \xi_N$ and $y = y_1 \xi_1 + \cdots + y_N \xi_N$, we have:

$$x \cdot y = \sum_{j=1}^{N} x_j y_j \tag{B-4}$$

$$x^T Ay = \sum_{j=1}^{N} \lambda_j x_j y_j \tag{B-5}$$

If one of the eigenvalues λ_j is negative, we can construct a vector x with $x \cdot u = 1$ and $x^T Ax < 0$, as follows. Let $v = v_1 \xi_1 + \cdots + v_N \xi_N$ be a fixed vector with $v \cdot u = 1$, and define for any real β

$$x = \frac{\beta \xi_j + v}{\beta u_j + 1} \tag{B-6}$$

where $u = u_1 \xi_1 + \cdots + u_N \xi_N$ is the expansion of u. Clearly $x \cdot u = 1$, and from (B-5) we compute

$$x^T Ax = \frac{1}{(\beta u_j + 1)^2} \left\{ \lambda_j \beta^2 + 2\beta \lambda_j v_j + \sum_{i=1}^{N} \lambda_j v_j^2 \right\}$$

Clearly this will be negative if β is large enough, since the expression in brackets is then dominated by the term $-|\lambda_j| \beta^2$. This proves (9).

Thus we assume A is p.s.d., i.e., that the eigenvalues $\{\lambda_j\}$ are all nonnegative. If for some index j we have $\lambda_j = 0$ but $u_j \neq 0$, and if we set $x = u_j^{-1} \xi_j$, then $x \cdot u = 1$, $x^T Ax = 0$. This proves (10).

On the other hand if $u_j = 0$ whenever $\lambda_j = 0$, we get by Schwarz' inequality

$$(\Sigma x_j u_j)^2 \leqslant (\Sigma \lambda_j x_j^2)(\Sigma \lambda_j^{-1} u_j^2) \tag{B-7}$$

where the summation is extended only over indices for which $\lambda_j \geqslant 0$. Since by (B-4) and (B-5) $\Sigma x_j u_j = x \cdot u$, and $\Sigma \lambda_j x_j^2 = x^T Ax$, (B-7) immediately implies that if $x \cdot u = 1$, then $x^T Ax \geqslant (\Sigma \lambda_j^{-1} u_j^2)^{-1} = \lambda$. On the other hand, by choosing $x_i = u_i \lambda_i^{-1} \lambda$ for all i, we get $x \cdot u = 1$ and $x^T Ax = \lambda$. This proves (11).

If u is itself an eigenvector for A, with eigenvalue σ, then in the expansion $u = u_1 \xi_1 + \cdots + u_N \xi_N$, u_j must be zero unless $\lambda_j = \sigma$. Hence $\lambda(A) = (\sigma^{-1} \Sigma u_j^2)^{-1} = \sigma/N$, since $\Sigma u_j^2 = u \cdot u = N$. This proves (12).

To prove (13), observe that

$$x^T(A - \lambda J)x = x^T Ax - \lambda(x \cdot u)^2$$

$$= \Sigma \lambda_j x_j^2 - \lambda(\Sigma x_j u_j)^2 \tag{B-8}$$

Comparing this to (B-7), we see that this expression will be nonnegative for all x if and only if $\lambda \leqslant (\Sigma \lambda_j^{-1} u_j^2)^{-1}$, i.e. $\lambda \leqslant \lambda(A)$. This proves (13).

Finally we turn to (14). Suppose then that $\xi_1, \cdots, \xi_N$ are principal axes for the matrix A, with corresponding eigenvalues $\{\lambda_j\}$, and that $\eta_1, \cdots, \eta_M$ are principal axes for B, with eigenvalues $\{\mu_k\}$. Suppose further that $u^{(A)} = u_1^{(A)} \xi_1 + \ldots + u_N^{(A)} \xi_N$, $u^{(B)} = u_1^{(B)} \eta_1 + \cdots + u_M^{(B)} \eta_M$ are the expansions of all ones vectors, with respect to these two bases.

It follows from known results (see Ref. 6, Section VII) that the MN vectors $\xi_j \times \eta_k$ are principal axes for the matrix $A \times B$, with associated eigenvalues $\lambda_j \mu_k$. Furthermore, the expansion of the MN – dimensional all-ones vector with respect to the basis $\{\xi_j \times \eta_k\}$ is clearly

$$u = \sum_{j,k} u_j^{(A)} u_k^{(B)} (\xi_j \times \eta_k)$$

Thus, according to (11),

$$\lambda(A \times B) = \left(\sum_{j,k} (\lambda_j \mu_k)^{-1} \left[u_j^{(A)} u_k^{(B)} \right]^2 \right)^{-1}$$

$$= \lambda(A)\, \lambda(B),$$

establishing (14).

References

1. Baumert, L. D., et. al., "A Combinatorial Packing Problem," pp. 97-108 in *Computers in Algebra and Number Theory*, SIAM-AMS Proceedings Vol. IV. Providence: American Mathematical Society, 1971.
2. Birkhoff, G., and MacLane, S., *A Survey of Modern Algebra*. New York: MacMillan, 1960.
3. Delsarte, P., "An Algebraic Approach to the Association Schemes of Coding Theory," Philips Research Reports Supplement 1973, No. 10.
4. Hoffman, K., and Kunze, R., *Linear Algebra*. Englewood Cliffs: Prentice – Hall, Inc., 1961.
5. Lovasz, L., "On the Shannon Capacity of a Graph," *IEEE Trans. Inform. Theory*, to appear.
6. MacDuffee, C., *The Theory of Matrices*. New York: Chelsea, 1956.
7. Schrijver, A., "A Comparison of the Bounds of Delsarte and Lovasz," *IEEE Trans. Inform. Theory*, to appear.
8. Shannon, C., "The Zero Error Capacity of a Noisy Channel," *IRE Trans. Inform. Theory*, Vol. 1T-2, pp. 8-19, Sept. 1956.
9. Weyl, H., *Algebraic Theory of Numbers* (Annals of Math. Studies No. 1). Princeton: Princeton University Press, 1940.
10. Wielandt, M. *Finite Permutation Groups*. New York: Academic Press, 1964.

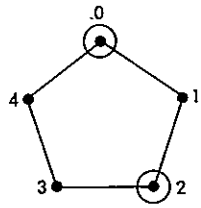

Fig. 1. The graph C_5; α (C_5) = 2 (this is also the graph Q_5 of Example 3 in Section IV)

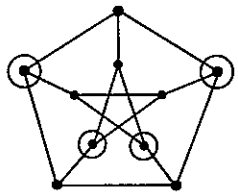

Fig. 2. The Peterson Graph

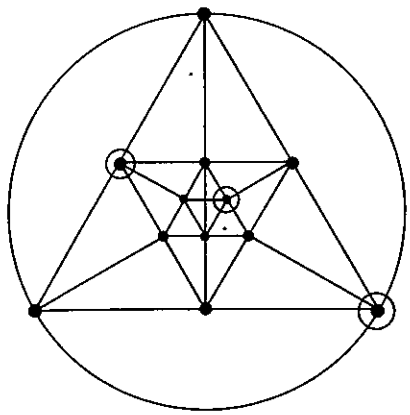

Fig. 3. The Icosahedron Graph

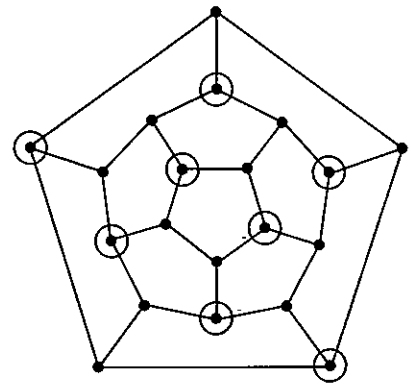

Fig. 4. The Dodecahedron Graph

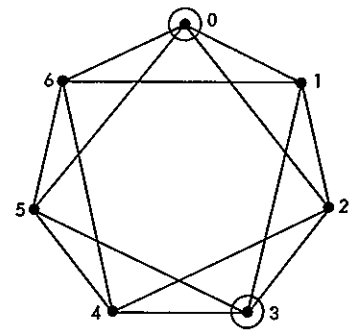

Fig. 5. A regular graph on seven vertices

An Algorithm for Generating an *m*-ary Summation Tree

M. Sievers
Communications Systems Research Section

An algorithm is presented for generating an m-ary summation tree. The algorithm is completely general and may be applied to any length input string. For an N length sequence summed in groups of m_ℓ at each level ℓ, a maximum of 3L - 2 storage is required where

$$\prod_{\ell=1}^{L} m_\ell = N$$

A special case of the general m-ary tree where all m_ℓ are equal will be used to smooth data in a radio-frequency interference experiment. The maximum storage required when $m_\ell = m$ for all ℓ reduces to the closed form $3 \log_m N - 2$.

I. Introduction

A study is in development that will determine the magnitude and nature of radio-frequency interference (RFI) in the vicinity of the Goldstone tracking complex. Spectra collected in the RFI experiments will be smoothed by a binary summation tree. This tree is a special case of the more general *m*-ary summation tree.

An *m*-ary summation tree is the tree formed by initially dividing an input string of length N into groups of length m_1 each. There will be $g_1 = \lfloor N/m_1 \rfloor$ groups. The notation $\lfloor X \rfloor$ denotes "the integer part of X." The m_1 elements in each group are summed to form a group sum. The resulting group sums form the first level in the tree.

The g_1 first level group sums are added in m_2-tuples to form the group sums of the second level. Clearly, the number of groups at the second level will be $\lfloor g_1/m_2 \rfloor$. In general, the number of groups at level L will be $\lfloor g_{L-1}/m_L \rfloor$. The process is shown pictorially in Fig. 1 for $N = 16$, $m_1 = 2$, $m_2 = 4$, and $m_3 = 2$.

M_ℓ-tuple additions are performed until either a single result is formed representing the sum of all inputs, or the tree

is incomplete. An incomplete tree is one in which at some level ℓ, g_ℓ groups are present where g_ℓ modulo m_ℓ is not zero. If the tree is incomplete, the final sum (sum of all inputs) is not formed since there will be one or more levels in which it was not possible to form an m_ℓ group.

In order for the final sum to be calculated, it must be possible to form m_ℓ groups at each level in the tree. This implies that there must be an L such that

$$\prod_{\ell=1}^{L} m_\ell = N \tag{1}$$

The value of L is the depth of the tree.

In the remainder of this article, only complete trees will be considered. Incomplete trees are of no practical value in the RFI experiments.

It is worth noting that in generating the m-ary summation tree, at most $N-1$ additions are required. This is equal to the minimum number of additions needed to add a string of N numbers. The tree structure, therefore, does not add overhead to the number of additions required to sum the N length input string.

For small N it might be reasonable to store each input word and summation result in memory. However, for large N this is not practical even if the summations are done "in place." The algorithm presented in the next section requires only $3L-2$ storage, where L is the depth of the m-ary tree.

II. m-ary Summation Tree Algorithm

The m-ary summation tree algorithm presented below makes the following assumptions.

(1) Input data are available serially.

(2) An accumulator is present for holding partial sums.

(3) When a group sum is completed it is immediately output.

(4) Once a level ℓ sum is formed, the level $\ell-1$ addends from which it was formed are no longer needed.

It is easy to show that only L data storage is necessary to generate the summation tree. To start, it is clear that only one storage word is needed to generate the first level in the tree. This is because once the first group sum is completed in the accumulator, it is moved to the storage location. The accumulator is then free to accumulate the next group sum. After this second group sum is available, it is added to memory and the result placed back in memory.

A second level group sum is produced when first level storage holds the sum of $m_2 - 1$ first level group sums and the accumulator contains an additional first level group sum. The level two group sum is obtained by adding the accumulator to level one storage. The first level storage is then free to collect the next $m_2 - 1$ first level groups sums.

The level two group sum is left in the accumulator after its formation. This situation is identical to the situation in the first level when a first level group sum is held in the accumulator. Therefore, it follows that only one storage element will be needed at the second level.

In general, any time m_ℓ group sums are available at the $\ell-1^{st}$ level ($m_\ell - 1$ held in $\ell-1$ memory and one in the accumulator), these are added to form a level ℓ group sum. This sum is left in the accumulator, and again the situation reduces to the level one situation. Therefore, only one storage element is needed at each level in the tree, except the last.

No storage is needed at level L, because when the level L sum is formed in the accumulator, it is immediately output. Since there are L levels, there will be need of $L-1$ storage locations in the tree. Including the accumulator, a total of L data storage is required.

An algorithm has been designed to implement the recursion described above. Figure 2 shows the flow chart of the algorithm.

The algorithm makes use of group counters and level counters. Group counters keep a count of the number of group sums stored in a given level storage. Level counters count the number of times a word of a given level has been formed. Since there will be $L-1$ storage cells, and since there is no need to count the number of level L sums formed (this count will be zero until the final sum is available at which time it becomes one), $L-1$ group and level counters are necessary.

The counters are used to keep track of which original input elements have been added to form a given level sum. Labeling the original inputs, $1, 2, 3, \ldots N$, then the c^{th} sum formed at the j^{th} level will be the sum:

$$S = \sum_{i=1+(c-1)r_j}^{cr_j} \tag{2}$$

where

$$r_j = \prod_{\ell=1}^{j} m_\ell$$

The vector of group counters (GC) can be used to indicate the total number of inputs processed. GC is a multiple radix number where the radix vector R is composed of the elements r_j. If each group counter gc_ℓ is interpreted as a radix r_ℓ digit, then the number of inputs processed at any time is:

$$\sum_{\ell=1}^{L} r_\ell \, gc_\ell \tag{3}$$

The total storage required for this algorithm is the sum of the storage needed to generate the tree plus the number of counters. From the discussion in this section, L storage is needed to generate the tree, and $2(L-1)$ counters are required from Eq. (1). This total equals $3L-2$ where

$$\prod_{\ell=1}^{L} m_\ell = N \tag{4}$$

III. Special Case: $m_\ell = m$ for all ℓ

The case when all level group sizes are the same is a special case of the general m-ary tree that greatly simplifies Eqs. (1) to (4). This case is also much simpler to implement than the general case. This makes it more attractive to build hardware for as will be done in the RFI experiments for $m = 2$.

Formation of m size groups at each level leads to the closed form

$$L = \log_m N \tag{5}$$

for Eq. 1. Since in this case $r_i = m^i$, Eq. (2) simplies to

$$\sum_{i=1+(c-1)m^j}^{cm^j} \tag{6}$$

for the c^{th} sum at the j^{th} level. Eq. (3) for the total number of inputs processed reduces to

$$\sum_{\ell=1}^{\log_m N-1} m^\ell gc_\ell \tag{7}$$

Finally, Eq. (4), the total storage required, becomes

$$3 \log_m N - 2 \tag{8}$$

IV. Sample Program

A portion of a FORTRAN program that implements the algorithm of Fig. 2 along with sample output data is shown in Fig. 3. The program was written for the special case where all $m_\ell = 2$. The input length N is 256 with input (1) = 1, input (2) = 2, etc. Vector IDATA is the input data vector, IGC is the group count vector, ICNT is the level count vector, and M is the group size. IGC, ICNT and M have been initialized to 0, 0, and 2, respectively, outside the portion of the program segment shown.

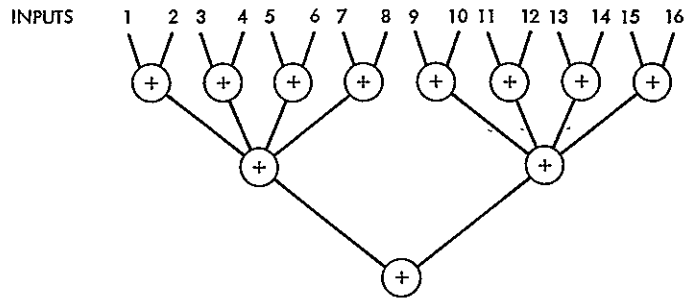

Fig. 1. *M*-ary summary tree for $N = 16$, $M_1 = 2$, $M_2 = 4$, and $M_3 = 2$

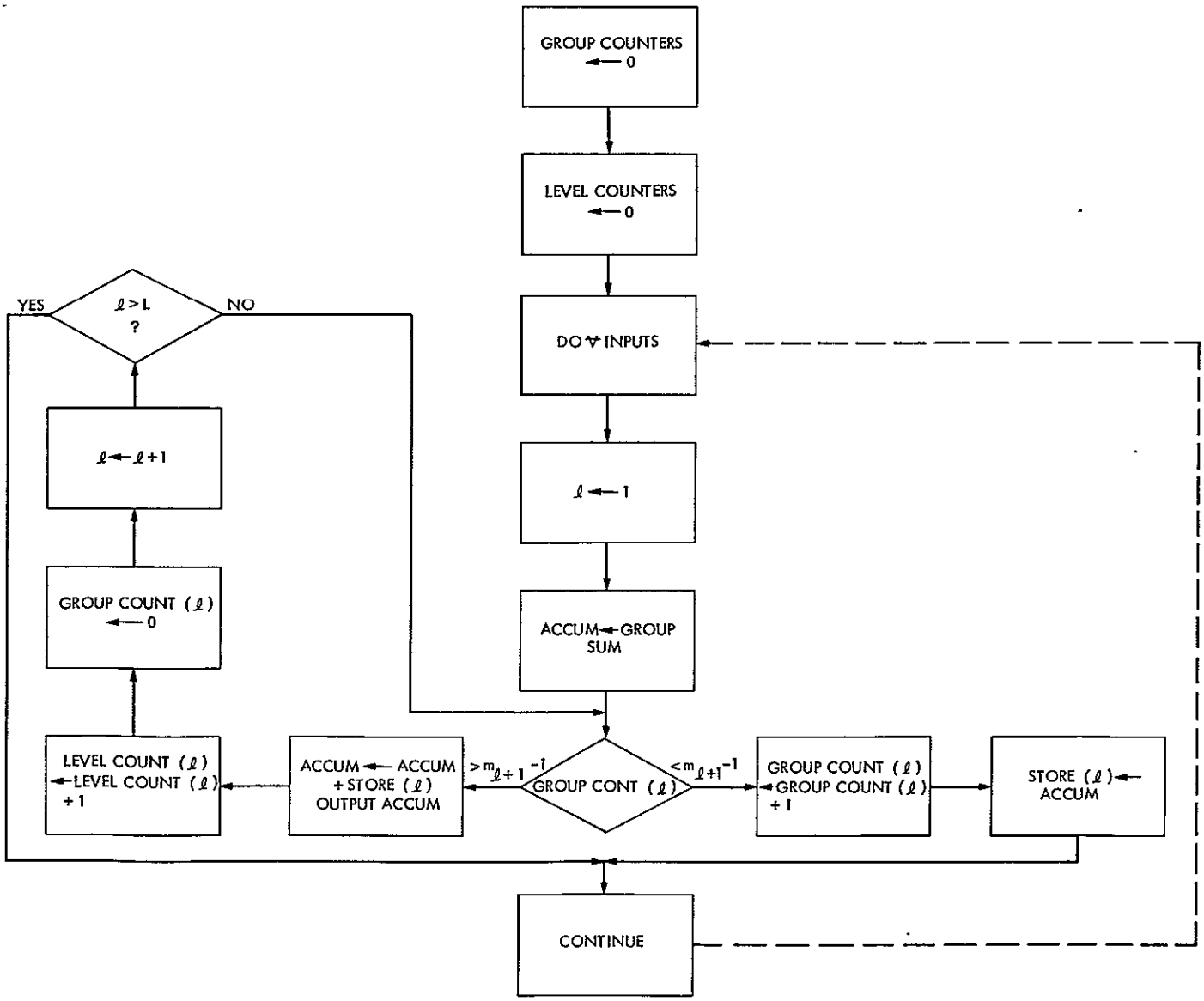

Fig. 2. Algorithm

```
 1                ISTOP = 256-M + 1
 2                DO 30 J = 1, ISTOP, M
 3                L = 1
 4                IACCUM = 0
 5                INDEX = J + M -1
 6                DO 25 JJ = J, INDEX
 7                WRITE (3, 60)JJ, IDATA (JJ)
 8                IACCUM = IACCUM + IDATA (JJ)
 9          25    CONTINUE
10                WRITE (3, 65) IACCUM
11          50    IF (IGC(L). LT. M-1) GO TO 40
12                IACCUM = IACCUM + ISTORE (L)
13                ICNT (L) = ICNT (L) +1
14                IBEG = M** (L + 1)* (ICNT (L) -1) + 1
15                IEND = M** (L + 1)* ICNT (L)
16                WRITE (3, 80) IBEG, IEND, IACCUM, ICNT (L),' L
17                IGC (L) = 0
18                ISTORE (L) = 0
19                L = L + 1
20                IF (L .EQ. 9) GO TO 30
21                GO TO 50
22          40    IGC (L) = IGC (L) + 1
23                ISTORE (L) = IACCUM + ISTORE (L)
24          30    CONTINUE
```

```
INPUT ( 1 )       1
INPUT ( 2 )       2
SUM OF INPUTS     3

INPUT ( 3 )       3
INPUT ( 4 )       4
SUM OF INPUTS     7

          SUM FROM 1 to 4 = 10

INP UT ( 5 )      5
INPUT ( 6 )       6
SUM OF INPUTS    11

INPUT ( 7 )       7
INPUT ( 8 )       8
SUM OF INPUTS    15

          SUM FROM 5 TO 8 = 26

          SUM FROM 1 TO 8 = 36
```

Fig. 3. FORTRAN Program sample output, $M = 2$, $N = 256$

On Estimating the Phase of a Periodic Waveform in Additive Gaussian Noise — Part I

L. L. Rauch
Communications Systems Research Section

Motivated by recent advances in technology, a new look is taken at the problem of estimating the phase of a periodic waveform in additive gaussian noise. The maximum a posteriori probability criterion with signal space interpretation is used to obtain the structures of optimum and some suboptimum phase estimators for the following cases: (1) known constant frequency and unknown constant phase with an a priori distribution; (2) unknown constant frequency and phase with a joint a priori distribution; (3) frequency a parameterized function of time with a joint a priori distribution on parameters and phase; (4) frequency a gaussian random process. (Part I introduces the general problem and treats case 1).

I. Introduction

Many of the algorithms, such as the phase-locked loop, currently used for phase estimation were originated against the background of analog technology some 25 years ago. While most of the algorithms have proved to be readily implementable in terms of current digital technology, this does not necessarily mean these are the most desirable algorithms with current (or near future) technology. In some cases, the structure of the optimal phase estimator is not complicated and it is worth considering against the background of new technology, whether some of the optimal or related suboptimal structures may be more desirable than some of the older algorithms. The purpose of this article is to look at some of the optimal and suboptimal structures.

II. Problem Statement

Following the approach of Ref. 1, we choose for the modulator function

$$y(f_c t + x_0(t))$$

where $y(\cdot)$ is periodic with unit period, f_c is the nominal frequency of the periodic waveform and $x_0(t)$ is the phase to be estimated on the basis of the received signal

$$z(t) = y(f_c t + x_0(t)) + n(t)$$

where $n(t)$ is a gaussian random process with mean 0 and known autocorrelation. In the case where the frequency is constant and known equal to f_c, then $x_0(t) = x_0$ is a constant to be estimated by a certain functional on $z(t)$ (random variable). In the case where the frequency is constant and unknown, then $x(t) = f_d t + x_0$ is a ramp function with parameters f_d and x_0 to be estimated in the form of certain functionals on $z(t)$. In the case where $x_0(t)$ is a gaussian random process with known autocorrelation function, then

the estimator for $x(t)$ is a certain operator on $z(t)$, which is of course a random process.

Let us assume the channel for $z(t)$ is bandlimited with bandlimit $F/2$ and that the gaussian noise $n(t)$ is white over this bandwidth. Then to proceed with the analysis we shall deal with all waveforms in terms of their unique sample representations at the sampling rate F. Thus,

$$\left.\begin{aligned} t_i &= i/F \\ z_i &= z(i/F) \\ n_i &= n(i/F) \\ y_i &= y(i/F) \\ x_i &= x(i/F) \end{aligned}\right\} \tag{1}$$

where the index i runs over the integers. For $x_0(t)$, a gaussian random process, the multivariant a priori probability density for the phase sequence $\{x_i\}$, $A_x \leqslant i \leqslant B_x$ is given by

$$f_x(\{x_i\}) = (2\pi)^{-\frac{B_x - A_x + 1}{2}} |R_{xij}|^{-1/2}$$

$$\exp\left\{-1/2 \sum_{i=A_x}^{B_x} \sum_{j=A_x}^{B_x} R_{xij}^{-1} x_i x_j\right\} \tag{2}$$

where $R_{xij} = R_{xx}(i/F, j/F)$ is the covariance matrix, $|R_{xij}|$ is its determinant and R_{xij}^{-1} is its inverse. In the case of unknown constant frequency and phase, we drop the gaussian assumption and the joint a priori probability density for f_0 and x_0 is given by

$$f_{fx}(f_0, x_0) = f_f(f_0) \cdot f_x(x_0) \tag{3}$$

where we have assumed f_0 and x_0 are independent random variables for the usual application and $f_f(\cdot)$ and $f_x(\cdot)$ are arbitrary. In the case of known constant frequency f_c and unknown phase x_0, the a priori probability density

$$f_x(x_0) \tag{4}$$

may also be chosen arbitrarily.

Similarly to Eq. (2) the probability density function for the noise sequence $\{n_i\}$ is given by

$$f_n(\{n_i\}) = (2\pi)^{-\frac{B_z - A_z + 1}{2}} |R_{nij}|^{-1/2}$$

$$\exp\left\{-1/2 \sum_{i=A_z}^{B_z} \sum_{j=A_z}^{B_z} R_{nij}^{-1} n_i n_j\right\} \tag{5}$$

Because of the assumptions regarding noise spectrum and sampling rate, the members of the noise sequence are uncorrelated.

This gives

$$R_{nij} = \sigma_n^2 \delta_{ij} \tag{6}$$

$$R_{nij}^{-1} = \frac{1}{\sigma_n^2} \delta_{ij} \tag{7}$$

$$|R_{nij}| = \sigma_n^{2(B_z - A_z + 1)} \tag{8}$$

where δ_{ij} is the Kronecker delta and we have used the assumption that $n(t)$ is wide sense stationary. Substituting Eqs. (6), (7), and (8) into Eq. (5) gives

$$f_n(\{n_i\}) = (2\pi)^{-\frac{B_z - A_z + 1}{2}} \sigma^{-(B_z - A_z + 1)}$$

$$\exp\left\{-\frac{1}{2\sigma_n^2} \sum_{i=A_z}^{B_z} n_i^2\right\} \tag{9}$$

Under the assumption that $\{x_i\}$ and $\{n_i\}$ are independent random processes

$$f_{z|x}(\{z_i\} \mid \{x_i\}) = f_n(\{z_i - y(f_c t + x_i)\}) \tag{10}$$

and

$$f_{zx}(\{z_i\}, \{x_i\}) = f_n(\{z_i - y(f_c t + x_i)\}) f_x(\{x_i\}) \tag{11}$$

Then

$$f_{x|z}(\{x_i\} \mid \{z_i\}) = \frac{f_n(\{z_i - y(f_c t + x_i)\}) f_x\{x_i\}}{f_z(\{z_i\})} \qquad (12)$$

Now the denominator of Eq. (12) is not a function of $\{x_i\}$, it is merely a factor that normalizes the joint density of Eq. (11) with respect to $\{x_i\}$. Thus, when estimating $\{x_i\}$ on the basis of a received $\{z_i\}$, we may treat $f_z(\{z_i\})$ as a constant.

Substituting Eqs. (2) and (9) in Eq. (12) gives

$$f_{x|z}(\{x_i\}|\{z_i\}) = C_1(\{z_i\}) \exp(-1/2) \left[\sum_{i=A_x}^{B_x} \sum_{j=A_x}^{B_x} R^{-1}_{xij} x_i x_j + \frac{1}{\sigma_n^2} \sum_{k=A_z}^{B_z} (z_k - y(f_c t_k + x_k))^2 \right] \qquad (13a)$$

where $C_1(\{z_i\})$ includes the constant coefficients from Eqs. (2) and (9). For the case of constant unknown frequency and phase we use Eq. (3) instead of Eq. (2) which gives, in place of Eq. (13a),

$$f_{fx|z}(f, x \mid \{z_i\}) = C_2(\{z_i\}) \exp\left[-\frac{1}{2\sigma_n^2} \sum_{k=A_z}^{B_z} (z_k - y(ft_k + x))^2 + \ln f_f(f) + \ln f_x(x) \right] \qquad (13b)$$

and, for the case of constant known frequency and unknown phase,

$$f_{x|z}(x|\{z_i\}) = C_2(\{z_i\}) \exp\left[\frac{-1}{2\sigma_n^2} \sum_{k=A_z}^{B_z} (z_k - y(f_c t_k + x))^2 + \ln f_x(x) \right] \qquad (13c)$$

The conditional probability density functions (13) contain all needed information for estimation of the phase (and/or frequency). The maximum a posteriori probability (MAP) estimator is simply the mode of the distribution (13), given the observed $\{z_i\}$. The minimum mean square error (MMSE) estimator is the mean of the distribution and the minimum mean absolute error (MMAE) estimator is the median of the distribution, etc. It is possible for a distribution to fail to have a unique mode or median.

III. Estimation of Unknown Phase with Known Constant Frequency

As the sampling rate F and noise bandwidth $F/2$ increase, while holding constant the noise spectral density $S_{nn} = \sigma_n^2/F$, the initial time $t_1 = A_z/F$ and the final time $t_2 = B_z/F$, the summation in Eq. (13c) becomes well approximated by an integral and we can write

$$f_{x|z}(x|[z(\cdot)]) = C_3[z(t)] \exp\left[\frac{-1}{2S_{nn}} \int_{t_1}^{t_2} (z(\tau) - y(f_c \tau + x))^2 \, d\tau + \ln f_x(x) \right] \qquad (14)$$

where the constant (with respect to x) coefficient C_3 is now a functional on $z(t)$. (As F becomes large, the coefficient for $f_{x,z}(\cdot,\cdot)$ becomes small, but the coefficient for $f_{x|z}(\cdot|\cdot)$ does not.) If the argument of the exponential function in Eq. (14) is

an even function of x about its maximum value, then the value of x at the maximum is not only the MAP estimator $\hat{x}_0$, but also the MMSE and MMAE estimator. We will pursue the MAP estimator and observe when it is also the optimum estimator under other criteria.

From Eq. (14) we see that the MAP estimator, if it exists, is the value of x that maximizes

$$-\frac{1}{N_0}\int_{t_1}^{t_2}(z(\tau)-y(f_c\tau+x))^2\,d\tau+\ln f_x(x) \tag{15}$$

where N_0 is the one-sided noise spectral density. Expanding the integrand gives

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

$$-\frac{1}{N_0}\int_{t_1}^{t_2}y^2(f_ct+x)\,d\tau+\ln f_x(x) \tag{16}$$

The first integral is not a function of x and the third integral is not a function of x if $t_2 - t_1 = n/f_c$, n a positive integer, which we assume henceforth. Thus the MAP estimator $\hat{x}_0$ is the value of x that maximizes

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

Real functions such as $z(t)$ and $y(f_ct+x)$ on the interval $t_1 \leqslant t \leqslant t_2$ are elements of a signal space (essentially a Hilbert space) with the usual inner product

$$\langle\vec{x},\vec{y}\rangle=\int_{t_1}^{t_2}x(t)\,y(t)\,dt \tag{18}$$

and norm

$$\|\vec{x}\|=\sqrt{\langle\vec{x},\vec{x}\rangle} \tag{19}$$

In terms of the vector space notation, Eq. (15) can be written

$$-\|\vec{z}-\vec{y}(x)\|^2+N_0\ln f_x(x) \tag{20}$$

and Eq. (17) can be written

$$2\langle\vec{z},\vec{y}(x)\rangle+N_0\ln f_x(x) \tag{21}$$

where $\|\vec{z}\|$ and $\|\vec{y}(x)\|$ are not functions of x. If x_0 has a uniform a priori distribution

$$f_x(x_0)=\begin{cases}1, & -1/2\leqslant x_0\leqslant 1/2\\ 0, & \text{elsewhere}\end{cases} \tag{22}$$

as is usually the case for an initial estimate, then we see from Eq. (20) that $\hat{x}_0$ should be chosen to minimize the distance between $\vec{z}$ and $\vec{y}(\hat{x}_0)$ or, from Eq. (21), to minimize the angle between $\vec{z}$ and $\vec{y}(\hat{x}_0)$ (maximize the cosine of the angle where

$$\cos\theta=\frac{\langle\vec{z},\vec{y}(\hat{x}_0)\rangle}{\|\vec{z}\|\cdot\|\vec{y}\|}\Bigg) \tag{23}$$

If $y(\cdot)$ is differentiable for all values of its argument then a necessary condition for maximizing

$$\langle\vec{z},\vec{y}(x)\rangle=\int_{t_1}^{t_2}z(\tau)\,y\,(f_c\tau+x)\,d\tau \tag{24}$$

is

$$\langle\vec{z},\vec{y}'(x)\rangle=\int_{t_1}^{t_2}z(\tau)\,y'(f_c\tau+x)\,d\tau=0 \tag{25}$$

Schemes that maximize Eq. (24) are often called direct estimators and those that solve Eq. (25) are often called indirect estimators. Indirect estimation must be used with caution. In the first place, the integrand of Eq. (25) must contain the derivative of $y(\cdot)$ and not merely $y(\cdot)$ shifted one quarter period, as in the special case of a sinusoid. Also Eq. (25) is a necessary condition and may well be satisfied by values of x that do not maximize Eq. (24). For example, Eq. (24) may have a number of relative maxima only one of which is the absolute maximum. In this case, Eq. (25) has multiple solutions.

If $|y(\cdot)|$ is a bounded variation function, as we shall assume, then the tip of the vector $\vec{y}(x)$ describes a continuous closed curve on the surface of a hypersphere of radius $||\vec{y}||$ about the origin in the signal space as x traverses its unit interval from $-1/2$ to $+1/2$. This curve may not have a derivative everywhere and it spans a linear subspace L_y that is generally not finite dimensional. The MAP phase direct estimation problem is now visualized by choosing a point $\vec{y}(x_0)$ on this curve (the phase to be estimated), moving away a distance in some direction to $\vec{z}$, determined by the additive noise vector[1] $\vec{n}$

$$\vec{z} = \vec{y}(x_0) + \vec{n} \tag{26}$$

and then seeking the point $\hat{x}_0$ on the curve $\vec{y}(x)$, parameterized by x, which is closest to $\vec{z}$ or its orthogonal projection onto L_y. In equation form this amounts to choosing $\hat{x}_0$ to minimize

$$||\vec{y}(x_0) + \vec{n} - \vec{y}(\hat{x}_0)||$$

(see Eq. (20) with $f_x(\cdot)$ constant). This is the same as choosing the point $\vec{y}(\hat{x}_0)$ on the curve to minimize the angle between $\vec{z}$ (or its orthogonal projection onto L_y) and $\vec{y}(\hat{x}_0)$ (see Eq. (21)). Since the path described by $\vec{y}(x)$ on the surface of the hypersphere can be rather arbitrary, depending on the choice of $y(\cdot)$, it can be appreciated that the indirect method of estimation that looks for points on the curve for which $||\vec{z} - \vec{y}(\hat{x}_0)||$ or $\langle \vec{z}, \vec{y}(\hat{x}_0) \rangle$ is stationary must be used with caution. The point $\hat{x}_0$ that maximizes $\langle \vec{z}, \vec{y}(x) \rangle$ may not even be stationary and there may be stationary points that do not give the maximum value of $\langle \vec{z}, \vec{y}(x) \rangle$. The vector $\vec{y}'(x)$, if it exists, is tangent to the curve and the solution of Eq. (25) involves finding points on the curve where its tangent is orthogonal to z (or its orthogonal projection onto L_y).

The indirect method tends to be computationally more desirable than the direct method (driving something to zero tends to be simpler than driving something to its global maximum). For this reason, many phase estimation devices are based on or inspired by the indirect method. A well-known example inspired by the indirect method is the phased-locked loop.

As we shall see in Section V, when $y(\cdot)$ is sinusoidal all of the perils of the indirect method disappear and other good things happen. But for as simple a waveform as the square wave there are problems as discussed in Section VI.

[1] We need only consider the orthogonal projection of $\vec{n}$ onto the linear subspace L_y spanned $\vec{y}(x)$.

IV. Discrete Phase Estimation

In case there are only a finite number of possible phases, as in the case of subcarrier or symbol synchronization in a fully synchronous system, the indirect method cannot be used if advantage is to be taken of the a priori information. The possible phases $x_{01}, x_{02}, \cdots, x_{0n}$ are used to obtain n values of Eq. (17), with $f_x(\cdot)$ replaced by the discrete probabilities, and the largest value of Eq. (17) identified (not more than two values of Eq. (17) need be stored at any time). This use of the direct method does not require that $y(\cdot)$ be differentiable, etc.

V. Estimation Of The Phase Of A Sinusoid

If $y(\eta) = A_c \cos 2\pi\,\eta$ then the curve described by $\vec{y}(x)$ in signal space is a circle of radius $A_c\sqrt{T/2}$ about the origin and lying in a certain two-dimensional linear subspace L_y (a plane through the origin). The orientation of the plane of the circle with respect to a fixed basis depends on the frequency f_c. The direct method of estimation amounts to finding the point $\vec{y}(\hat{x}_0)$ on the circle closest to the received signal $\vec{z} = \vec{y}(x_0) + \vec{n}$ or its orthogonal projection onto L_y and the indirect method amounts to finding a stationary point where the tangent of the circle is orthogonal to $\vec{z}$ or its orthogonal projection onto L_y. If $\vec{z}$ is not orthogonal to the plane of the circle (a singular condition) then there are always exactly two stationary points and the one corresponding to minimum $||\vec{z} - \vec{y}(\hat{x}_0)||$ is easily selected. The fact that $\vec{y}(x)$ lies in a two-dimensional linear subspace gives the further simplification that $||\vec{z} - \vec{y}(x)||$ and $\langle \vec{z}, \vec{y}(x) \rangle$ can be evaluated in terms of two scalar functions of $\vec{z}$ (functionals of $z(t)$) and two trigonometric functions of x. In fact Eq. (25) can be solved explicitly for x_0 in terms of the two scalar functions of $\vec{z}$ (the coordinates of $\vec{z}$ with respect to a basis). Also, due to the symmetry of the circular path of $\vec{y}(x)$, $||\vec{z} - \vec{y}(x)||$ is an even function about its minimum for any $\vec{z}$ and thus, by Eq. (14),

$$f_{x|z}(x|\vec{z}) = C_3(\vec{z}) \exp\left[\frac{-1}{2S_{nn}} ||\vec{z} - \vec{y}(x)||^2 + \ln f(x)\right] \tag{27}$$

is an even function about its maximum provided the a priori distribution $f(x)$ is uniform. In this case, the MAP estimator is also the MMSE, MMAE, etc. estimator.

Let us return to Eq. (17) to obtain the indirect estimator relation by differentiating

$$\frac{2}{N_0} \int_{t_1}^{t_2} z(\tau) A_c \cos 2\pi\,(f_c \tau + x)\, d\tau + \ln f_x(x) \tag{28}$$

to obtain

$$\frac{2 \cdot 2\pi A_c}{N_0} \int_{t_1}^{t_2} z(\tau) \sin 2\pi (f_c \tau + \hat{x}_0)\, d\tau = \frac{f_x'(\hat{x}_0)}{f(\hat{x}_0)} \quad (29)$$

Define

$$I[z(\cdot)] = \int_{t_1}^{t_2} z(\tau) \cos 2\pi f_c \tau \, d\tau = \frac{\langle \vec{z}, \vec{y}(0) \rangle}{\sqrt{2}\, \|\vec{y}\|} \quad (30)$$

$$Q[z(\cdot)] = \int_{t_1}^{t_2} z(\tau) \sin 2\pi f_c \tau \, d\tau = -\frac{\langle \vec{z}, \vec{y}'(0) \rangle}{\sqrt{2}\, \|\vec{y}'\|} \quad (31)$$

Then Eq. (29) may be written

$$Q \cos 2\pi \hat{x}_0 + I \sin 2\pi \hat{x}_0 = \frac{N_0}{4\pi A_c} \frac{f_x'(\hat{x}_0)}{f_x(\hat{x}_0)} \quad (32)$$

If the signal-to-noise ratio is very large or if the a priori distribution is uniform the right member of Eq. (32) vanishes and we have the familiar result

$$2\pi \hat{x}_0 = -\arctan \frac{Q}{I} \quad (33)$$

With current digital technology Eq. (33) can be easily mechanized and x_0 can be obtained on a four-quadrant basis.

It is interesting to consider the functional form of $f_x'(x)/f_x(x)$ for several a priori distributions. In the case of the gaussian distribution

$$\frac{f_x'(x)}{f_x(x)} = -\frac{x - m}{\sigma^2} \quad (34)$$

where m is the mean and σ^2 the variance. The gaussian is the only distribution giving a linear result. Substituting Eq. (34) into the right member of Eq. (29) gives something like a phase-locked loop except that $\hat{x}_0$ is constant over the interval $t_1 \leqslant t \leqslant t_2$ instead of a function of time.

The a posteriori density function for the estimator of Eq. (33) is given by (Ref. 2)[2]:

$$f_{x|\hat{x},r}(x_0|\hat{x}_0, r) = C\, e^{\alpha r \cos 2\pi(x_0 - \hat{x}_0)} \quad \text{(footnote 3)} \quad (35)$$

where C is the normalizing constant and

$$\alpha^2 = \frac{A_c^2 T}{N_0} \quad (36)$$

$$r^2 = 4\frac{I^2 + Q^2}{N_0 T} \quad (37)$$

where $T = t_2 - t_1$.

The distribution (35) can also be viewed as the a priori distribution for a second estimate $\hat{x}_{02}$ conditioned on the first estimate $\hat{x}_{01}$ ($= \hat{x}_0$). We can then write

$$f_{x_2|x_1,r_1}(\hat{x}_{02}|\hat{x}_{01}, r_1) = C\, e^{\alpha_1 r_1 \cos 2\pi(\hat{x}_{02} - \hat{x}_{01})} \quad (38)$$

where we assume I_2 and Q_2 are independent of I_1 and Q_1 as in the case where the time intervals for the successive estimations are nonoverlapping. From Eq. (38),

$$\frac{f'_{x_2|x_1,r_1}(\hat{x}_{02}|\hat{x}_{01}, r_1)}{f_{x_2|x_1,r_1}(\hat{x}_{02}|\hat{x}_{01}, r_1)} = -2\alpha_1 r_1 \sin 2\pi(\hat{x}_{02} - \hat{x}_{01}) \quad (39)$$

---

[2] The distribution Eq. (35) results from averaging Eq. (14) over all $z(t)$ subject to the condition of the observable r. If the condition r is removed then

$$f_{x|\hat{x}}(x_0|\hat{x}_0) = e^{-\frac{\alpha^2}{2}} + \sqrt{2\pi}\,\alpha \cos 2\pi(\hat{x}_0 - x_0)\, \Phi(\alpha \cos 2\pi(\hat{x}_0 - x_0))\, e^{-\frac{\alpha^2}{2}\sin^2 2\pi(\hat{x}_0 - x_0)}$$

which is less desirable.

[3] $C\, e^{\alpha r \cos 2\pi\epsilon}$ is simply the probability density function of the estimation error ϵ, conditioned on r.

which can be substituted in the right member of Eq. (32) to give

$$Q_2 \cos 2\pi\hat{x}_{02} + I_2 \sin 2\pi\hat{x}_{02} = -\frac{N_0 \alpha_1 r_1}{2A_c} \sin 2\pi(\hat{x}_{02} - \hat{x}_{01}) \tag{40}$$

Expanding the sine term in the right member and solving for x_{02} gives

$$\tan 2\pi\hat{x}_{02} = -\frac{Q_2 + \frac{N_0 \alpha_1 r_1}{2A_c} \sin 2\pi\hat{x}_{01}}{I_2 + \frac{N_0 \alpha_1 r_1}{2A_c} \cos 2\pi\hat{x}_{01}} \tag{41}$$

This is actually a bayesian recursive estimator based on the new observables Q_2 and I_2 and the previous estimate $\hat{x}_{01}$ and the previous r_1. A little calculation will show that

$$\frac{N_0 \alpha_1 r_1}{2A_c} \sin 2\pi\hat{x}_{01} = Q_1, \quad \frac{N_0 \alpha_1 r_1}{2A_c} \cos 2\pi\hat{x}_{01} = I_1 \tag{42}$$

and so Eq. (41) can be written

$$\tan 2\pi\hat{x}_{02} = -\frac{Q_2 + Q_1}{I_2 + I_1} \tag{43}$$

which is as it should be since the optimum recursive estimator should give the same result as a one-time optimum estimator using all of the observables. Clearly the error distribution for $\hat{x}_{02}$ has the same form [Eq. (35)] as that for $\hat{x}_{01}$ and thus the form of Eq. (35) is a reproducing distribution for this recursive baysian estimator. At each step of the estimator, $f_{x|z}(x_{0n}|\, z(\cdot))$ is an even function of x_{0n} about its maximum [see Eq. (14)] and so the MAP recursive estimator is also MMSE, MMAE, etc.

The mechanization of an estimator suggested by Eq. (43) and Eq. (35) seems straightforward. Successive pairs Q_i, I_i for successive equal increments of time of duration T are stored in a shift register and the contents summed to give

$$Q = \sum_{i=1}^{n} Q_i, \quad I = \sum_{i=1}^{n} I_i \tag{44}$$

where n is increasing with time. The phase estimate at any time is given by

$$\tan 2\pi\hat{x}_0 = -\frac{Q}{I} \tag{45}$$

and the error of the estimate is described by Eq. (35) where

$$\alpha r = \frac{2A_c}{N_0} \sqrt{(I^2 + Q^2)} \tag{46}$$

The mean square error of the estimate (or any other moment) is an easily calculated function of αr. Thus if a specified value of the MSE is designated as the "operating threshold," then when the generally increasing αr reaches the corresponding value, the two shift registers can start dumping so that Eq. (44) is replaced by

$$Q = \sum_{n-N}^{n} Q_i, \quad I = \sum_{n-N}^{n} I_i \tag{47}$$

to give a running or aperture sum of the Q_i, I_i. The calculation of αr from Eqs. (46) and (47) is continued for each value of n and, if it should decrease significantly due to the run of values of Q_i^2 and I_i^2 (r is a random variable), then the value of N in Eq. (47) can be automatically increased.

This has the advantage of following slow changes in x_0 with an aperture filter of length NT with N controlled to maintain the quality of the estimate at a prescribed level. This is something like a phase-locked loop of variable bandwidth with a guaranteed limit on noise MSE. However, unlike the phase-locked loop, there is not a phase initial condition to affect acquisition time.

The aperture averaging scheme is a suboptimal way of handling the estimation of a nonconstant phase. If a variable phase $x(t)$ can be modelled by a parameterized random process such as

$$x(t) = f_d t + x_0$$

with f_d and x_0 random variables, or by a more general one, then an optimum estimator can be formulated as discussed in Part II.

VI. Estimation of the Phase of Nonsinusoids

Returning to Eqs. (20) and (21), we see that to evaluate $\|\vec{z}_y - \vec{y}(x)\|$ or $\langle \vec{z}_y, \vec{y}(x) \rangle$ as functions of $\vec{z}_y$ we must have the orthogonal projection $\vec{z}_y$ of $\vec{z}$ onto the linear subspace L_y spanned by $\vec{y}(x)$, $-1/2 \leqslant x \leqslant 1/2$ expressed in terms of coordinates $z_1, z_2, \cdots$ of $\vec{z}$ with respect to some orthonormal basis $\vec{v}_1, \vec{v}_2, \cdots$ for L_y. For the sinusoid considered in the previous section, L_y is two dimensional, and orthonormal basis vectors $\vec{v}_1$ and $\vec{v}_2$ are given by [see Eqs. (30) and (31)]:

$$v_1(t) = \sqrt{\frac{2}{T}} \cos 2\pi f_c t \tag{48}$$

$$v_2(t) = \sqrt{\frac{2}{T}} \sin 2\pi f_c t \tag{49}$$

and the coordinates z_1 and z_2 of $\vec{z}$ are given by

$$z_1 = \langle \vec{z}, \vec{v}_1 \rangle = \sqrt{\frac{2}{T}}\, I\,[z(\cdot)] \tag{50}$$

$$z_2 = \langle \vec{z}, \vec{v}_2 \rangle = \sqrt{\frac{2}{T}}\, Q\,[z(\cdot)] \tag{51}$$

These coordinates of $\vec{z}$ are a minimum set of observables required to determine $\hat{x}_0$ as in Eq. (33) or Eq. (41). When L_y has $n \geqslant 3$ dimensions, the explicit solution for $\hat{x}_0$ in terms of the n observables (coordinates of $\vec{z}$) becomes more difficult.

A convenient orthonormal basis for a general periodic $y(\eta)$ is the fourier basis $\sqrt{2/T} \cos 2\pi j\eta$, $\sqrt{2/T} \sin 2\pi j\eta$, $j = 1, 2, \cdots \infty$. Clearly any $y(\cdot)$ with a discontinuity or a discontinuity in any derivative requires a basis of infinite dimension. In such a case, the minimum set of observables required to determine $\hat{x}_0$ is, in effect, the complete function $z(t)$ from t_1 to t_2; that is, the infinite set of coordinates of $\vec{z}$ with respect to the basis above. In the case of large noise where the magnitude of the orthogonal projection of $\vec{n}$ onto L_y is of the same order as $\|\vec{y}\|$, other periodic waveforms $y(\cdot)$ cannot be much better than the sinusoid for phase estimation. This is because the difference between the maximum and minimum values of $\|\vec{z}_y - \vec{y}(x)\|$ tend to be about as large for the sinusoid as for any other $y(\cdot)$. The sinusoid describes a great circle on the surface of the hypersphere of radius $\|\vec{y}\|$ and the path of any other $\vec{y}(x)$ is also confined to the surface of the same hypersphere.

However, for the case of small noise where $\vec{n}_y$ usually does not take $\vec{z}$ far from the path $\vec{y}(x)$ on the hypersphere and the error due to the noise is approximately the component of $\vec{n}$ along $\vec{y}'(x_0)$, the error in the estimate can be reduced simply by increasing the length of the closed path described by $\vec{y}(x)$ so that a given noise displacement along $\vec{y}'(x_0)$ corresponds to a smaller increment of x. This can only be done by increasing the dimension of L_y and of the resulting one-less dimension of the surface of the hypersphere on which the path described by $\vec{y}(x)$ lies. If s is the arc length along the path, then we want to increase

$$\frac{ds}{dx} = \|\vec{y}'(x)\| \tag{52}$$

in order to reduce the error of phase estimation for the small noise case. For the sinusoidal case of the previous section,

$$\frac{ds}{dx} = 2\pi A_c \sqrt{\frac{T}{2}} \tag{53}$$

and the variance of the zero mean noise component along s is just $S_{nn} = N_0/2$. Thus

$$\sigma^2(\hat{x}_0) = \frac{N_0/2}{\left(\frac{ds}{dx}\right)^2} = \frac{1}{4\pi^2} \cdot \frac{N_0}{A_c^2 T} \tag{54}$$

a well known result.

For the case of

$$y(\eta) = \sum_{j=1}^{n} (a_j \cos 2\pi j\eta + b_j \sin 2\pi j\eta) \tag{55}$$

where

$$(1/2) \sum_{j=1}^{n} (a_j^2 + b_j^2) = p_y \text{ is fixed}$$

we have, by Eq. (52),

$$\left(\frac{ds}{dx}\right)^2 = 2\pi^2 T \sum_{j=1}^{n} j^2 (a_j^2 + b_j^2) \tag{56}$$

As a simple example of a nonsinusoidal $y(\cdot)$ let us consider

$$y(\eta) = \sqrt{0.9}\,A_c \left(\cos 2\pi\eta + \frac{1}{3}\cos 6\pi\eta\right) \tag{57}$$

which has the same power as $A_c \cos 2\pi\eta$ and which is the fundamental and next higher harmonic of a square wave. Here

$$\left(\frac{ds}{dx}\right)^2 = 0.9 \times 4\pi^2 A_c^2 T \tag{58}$$

and

$$\sigma^2(\hat{x}_0) = \frac{1}{0.9 \times 8\pi^2} \cdot \frac{N_0}{A_c^2 T} \tag{59}$$

for small noise. This is a 2.6-db reduction of phase error with respect to the sinusoid result of Eq. (54). We next consider phase estimation for this example in the presence of large noise. The L_y for Eq. (57) has four dimensions and $\vec{y}(x)$ traces out a smooth (all derivatives exist) closed path on the three-dimensional surface of a four-dimensional hypersphere for $-1/2 \leqslant x \leqslant 1/2$. The path is rather more interesting than the circle considered in the previous section. For a typical large noise $\vec{z}$ the norm $||\vec{z}_y - \vec{y}(x)||$ to be minimized as a function of x has six stationary points – three maxima and three minima. If we define

$$I_1 = \int_{t_1}^{t_2} z(\tau) \cos 2\pi f_c \tau \, d\tau \tag{60a}$$

$$Q_1 = \int_{t_1}^{t_2} z(\tau) \sin 2\pi f_c \tau \, d\tau \tag{60b}$$

$$I_3 = \int_{t_1}^{t_2} z(\tau) \cos 6\pi f_c \tau \, d\tau \tag{60c}$$

$$Q_3 = \int_{t_1}^{t_2} z(\tau) \sin 6\pi f_c \tau \, d\tau \tag{60d}$$

then

$$\frac{\langle \vec{z}_y, \vec{y}(x) \rangle}{\sqrt{0.9}\,A_c} = I_1 \cos 2\pi x - Q_1 \sin 2\pi x + \frac{I_3}{3} \cos 6\pi x - \frac{Q_3}{3} \sin 6\pi x \tag{61}$$

and

$$\frac{\langle \vec{z}_y, \vec{y}'(x) \rangle}{\sqrt{0.9}\,A_c} = -2\pi \left[Q_1 \cos 2\pi x + I_1 \sin 2\pi x + Q_3 \cos 6\pi x + I_3 \sin 6\pi x \right] \tag{62}$$

For the case $I_1 = Q_1 = I_3 = Q_3 = 1$ the indirect estimator $\langle \vec{z}_y, \vec{y}'(x) \rangle = 0$ has six solutions at approximately $x = -0.31$, -0.25, $-.06$, 0.19, 0.25, 0.44 at which are located respectively a max, min, max, min, max, min of $\langle \vec{z}_y, \vec{y}(x) \rangle$. The respective values of $\langle \vec{z}_y, \vec{y}(x) \rangle / \sqrt{0.9}\,A_c$ are 0.72, 0.67, 1.74, -0.72, -0.67, -1.74. Of the six values of x provided by the indirect estimator only one, $\hat{x}_0 = -0.06$ is the MAP estimate of x_0. In this case $\langle \vec{z}_y, \vec{y}(x) \rangle$ is not an even function with respect to $x = \hat{x}_0$ and therefore Eq. (14) (with $f_x(\cdot)$ uniform) is not an even function with respect to its maximum at $x = \hat{x}_0$. Thus it is not clear whether the MAP estimator is MMSE, MMAE, etc. As we have seen, it is not difficult to find $\sigma^2(\hat{x}_0)$ for small noise (in this case $\hat{x}_0$ is approaching the MMSE estimator), but the probability distribution for $\hat{x}_0$ does not appear to be known for the general case.

For small noise, the indirect MAP estimator $\langle \vec{z}_y, \vec{y}'(x) \rangle = 0$ may have six solutions, but if so, there will be two closely spaced pairs that are easily distinguished and rejected while the other two are as in the case of a sinusoid. The small noise performance of the MAP estimator above in Eq. (59) is considerably better (7-db) than that of the suboptimal estimator $\langle \vec{z}, \vec{y}(x - 1/4) \rangle = 0$, which is sometimes used.

We turn next to estimation of the phase of a square wave

$$y(\eta) = \text{Cos } 2\pi\eta \quad \text{(footnote 4)} \tag{63}$$

[4] The Notation Cos and Sin refers to unit-amplitude square waves in phase with cos and sin.

Here L_y is infinite dimensional, and the closed path on the hypersphere surface described by $\vec{y}(x)$ is the sum of circular motions in mutually orthogonal planes in accordance with the fourier series expansion of Eq. (63). In this case, the maximum of $\langle \vec{z}, \vec{y}(x) \rangle$ [see Eq. (21)] is generally not stationary and an indirect MAP estimator is not appropriate. The direct estimator for a uniform a priori distribution on x_0 seeks the value of x that maximizes

$$\langle \vec{z}_y, \vec{y}(x) \rangle = \int_{t_1}^{t_2} z(\tau)\ \mathrm{Cos}\ 2\pi(f_c\tau + x)\, d\tau$$

$$= \frac{4}{\pi} \sum_{j=1}^{\infty} \frac{1}{2j-1} \int_{t_1}^{t_2} z(\tau) \cos 2\pi(2j-1)$$

$$\times (f_c\tau + x)\, d\tau \qquad (64)$$

$$= \frac{4}{\pi} \sum_{j=1}^{\infty} \frac{I_{2j-1}}{2j-1} \cos 2\pi(2j-1)x$$

$$- \frac{Q_{2j-1}}{2j-1} \sin 2\pi(2j-1)x$$

where

$$I_k = \int_{t_1}^{t_2} z(\tau) \cos 2\pi k f_c \tau\, d\tau \qquad (65a)$$

$$Q_k = \int_{t_1}^{t_2} z(\tau) \sin 2\pi k f_c \tau\, d\tau \qquad (65b)$$

much as in Eq. (60), where the I_k, Q_k are the infinite (or very large) set of observables needed. Since

$$z(t) = A_c\ \mathrm{Cos}\ 2\pi(f_c t + x_0) + n(t) \qquad (66)$$

I_k and Q_k are independent gaussian random variables with means

$$E[I_k] = \frac{2A_c T}{\pi k} \cos 2\pi k x_0 \qquad (67a)$$

$$E[Q_k] = -\frac{2A_c T}{\pi k} \sin 2\pi k x_0 \qquad (67b)$$

and variances

$$\sigma^2[I_k] = \sigma^2[Q_k] = \frac{N_0 T}{4} \qquad (68)$$

where $T = t_2 - t_1$. Another way to look at this estimation problem is to use Eq. (66) to write

$$\langle \vec{z}_y, \vec{y}(x) \rangle = A_c \int_{t_1}^{t_2} \mathrm{Cos}\ 2\pi(f_c\tau + x_0)\ \mathrm{Cos}\ 2\pi(f_c\tau + x)\, d\tau$$

$$+ \int_{t_1}^{t_2} n(\tau)\ \mathrm{Cos}\ 2\pi(f_c\tau + x)\, d\tau \qquad (69)$$

The first integral is an even periodic triangular function of $x - x_0$ of period unity, peak values $\pm A_c T$, and slopes $\pm 4A_c T$. The last integral is a periodic random process $N_z(x)$ that, when added to the triangular function, is the function of x to be maximized for the MAP estimator. Thus

$$\langle \vec{z}_y, \vec{y}(x) \rangle = A_c T\ \mathrm{Cos}\ 2\pi(x - x_0) + N_z(x) \quad \text{(footnote 5)} \qquad (70)$$

Now

$$R_{NN}(x_1, x_2) = E\left[\int_{t_1}^{t_2} \int_{t_1}^{t_2} n(\tau_1)\, n(\tau_2)\ \mathrm{Cos}\ 2\pi(f_c\tau_1 + x_1) \right.$$

$$\left. \times\ \mathrm{Cos}\ 2\pi(f_c\tau_2 + x_2)\, d\tau_2 d\tau_1 \right]$$

$$= \frac{N_0}{2} \int_{t_1}^{t_2} \int_{t_1}^{t_2} \delta(\tau_1 - \tau_2)\ \mathrm{Cos}\ 2\pi(f_c\tau_1 + x_1)$$

[5] The notation Cos and Sin refers to unit-amplitude triangular waves in phase with cos and sin.

$$\times \text{Cos}\, 2\pi(f_c\tau_2 + x_2)\, d\tau_2 d\tau_1$$

$$= \frac{N_0}{2} \int_{t_1}^{t_2} \text{Cos}\, 2\pi(f_c\tau_1 + x_1)$$

$$\times \text{Cos}\, 2\pi(f_c\tau_1 + x_2)\, d\tau_1$$

$$= \frac{N_0 T}{2} \text{Cos}\, 2\pi(x_1 - x_2) \tag{71}$$

Thus the autocorrelation of $N_z(x)$ is a periodic triangular function of unit period. This means the power spectral density $S_{NN}(f)$ consists of a sum of delta functions in accordance with the amplitudes and frequencies of the sinusoidal components of the triangular autocorrelation (71). This is in agreement with Eqs. (64), (65), and (67).

The noise performance evaluation of the MAP phase estimator for the square wave requires that we obtain the probability distribution of the value of x that maximizes Eq. (70) where $N_z(x)$ is gaussian with autocorrelation (71). It is clear that the unconditional density will be an even function of x about x_0. However, the conditional density (with uniform a priori x)

$$f_{x|z}(x|\vec{z}) = C_4(\vec{z}) \exp \frac{2}{N_0} \left[A_c T \,\text{Cos}\, 2\pi(x - x_0) + N_z(x) \right] \tag{72}$$

is generally not an even function of x about x_0 and so the MAP estimator is not necessarily the MMSE, etc. estimator.

The small noise performance of the MAP phase estimator for the square wave is arbitrarily good in the sense that as a square wave is approximated by including more and more of its harmonics, the length of the closed path on the hypersphere surface increases without limit and the ds/dx of Eq. (52) approaches infinity. However, in this case "small noise" is noise whose component of RMS length $\sqrt{N_0/2}$ along $\vec{y}'(x_0)$ is small compared to the distance Δs along which the curve $\vec{y}(x)$ is approximately straight. This distance Δs is a fraction of the amplitude of the highest harmonic in $y(\cdot)$ and becomes arbitrarily small as the square wave is better and better approximated. Thus, the ideal square wave is a special case having no "linearized small noise" result for the MAP phase estimator. Some previous work (Ref. 3) on phase estimation of square waves is in agreement with this insight.

It appears that although square waves are easily generated and manipulated in electronic circuits, they are more difficult to deal with than sinusoids when estimating phase. This is generally true for any $\vec{y}(x)$ which spans a linear subspace of more than two dimensions. To find the value of x that maximizes $f_{x|z}(x|\vec{z})$, $\vec{z}$ must be observed in L_y and not in some linear subspace of smaller dimensions; that is, the number of observables must equal the dimension of L_y. If we use a smaller number of observables, we are dealing with an orthogonal projection of $\vec{y}(x)$ and $\vec{z}$ onto some linear subspace $\overline{L}_y$ of L_y. Then choosing x to minimize the distance between $\vec{y}(x)$ and $\vec{z}$ in $\overline{L}_y$ generally does not minimize the distance in L_y, and so the result is a suboptimal estimate. For the small noise case, the variance of the component of $\vec{n}$ along the projection of $\vec{y}'(x_0)$ onto $\overline{L}_y$ is the same as the variance of the component of $\vec{n}$ along $\vec{y}'(x_0)$ (in L_y).

But if $\bar{s}$ is the arclength of the projection onto $\overline{L}_y$, then generally

$$\frac{d\bar{s}}{dx} < \frac{ds}{dx} \tag{73}$$

and by Eq. (54) the variance of the error in $\hat{x}_0$ is increased.

For example, in the case of a square wave where the observables chosen are

$$\overline{I} = \int_{t_1}^{t_2} z(\tau)\, \text{Cos}\, 2\pi\tau\, d\tau \tag{74a}$$

$$\overline{Q} = \int_{t_1}^{t_2} z(\tau)\, \text{Sin}\, 2\pi\tau\, d\tau \tag{74b}$$

instead of Eq. (65), as in the idealization of the sequential ranging system, the above observations apply. L_y is infinite dimensional and $\overline{L}_y$ is two dimensional. Here we have $\vec{y}(x)$ given by

$$y(f_c t + x) = \frac{A_c}{\sqrt{2}} \text{Cos}\, 2\pi(f_c t + x) \tag{75}$$

with

$$\vec{v}_1 = \frac{\vec{y}(0)}{\|\vec{y}\|} \tag{76a}$$

$$\vec{v}_2 = \frac{\vec{y}(-1/4)}{\|\vec{y}\|} \tag{76b}$$

as orthonormal basis vectors for $\bar{L}_y$. The small noise performance easily follows from

$$\left(\frac{ds}{dx}\right)^2 = \left(\frac{dy_1}{dx}\right)^2 + \left(\frac{dy_2}{dx}\right)^2 = 8A_c^2 T \sqrt{(\cos^2 2\pi x + \sin^2 2\pi x)}$$

$$= 16A_c^2 T, x \neq 0, \pm 1/4, 1/2 \tag{77}$$

Substituting this in the middle member of Eq. (54) gives

$$\sigma^2(\hat{x}_0) = \frac{N_0}{32A_c^2 T} \tag{78}$$

Thus the small noise performance for this case is 0.9 db worse than for a sinusoid of the same power and frequency (in the presence of the same noise).

The performance of the suboptimal estimator using the observables in Eq. (74) has been completely evaluated in Refs. 4, 5, and 6 including the singular behaviour at $x = 0$, $\pm 1/4$, $1/2$ in the small noise case. It is interesting to observe that optimum estimation of phase based on the incomplete set of observables in Eq. (74) requires a knowledge of the amplitude of the square wave in $z(t)$ – a piece of information that is not needed by the MAP estimator that maximizes Eqs. (69) or (70).

If we return to the full infinite dimensional basis for the square wave we see that the radius of the circle described by $\vec{y}(x)$ in the plane corresponding to each harmonic decreases inversely as the frequency of harmonic while the variance of the projection of $\vec{n}$ on any plane is N_0. Thus for a given value of $A_c^2 T/N_0$ there will be a highest harmonic in whose plane the signal to noise ratio is greater than unity. The observables corresponding to higher harmonics will be mostly noise and therefore should be more lightly weighted. This leads to the question of what is the useful number of observables to use as a function of $A_c^2 T/N_0$. An almost equivalent question is what is the useful number of harmonics in $y(\cdot)$ as a function of $A_c^2 T/N_0$. This has been in effect answered in the context of an indirect estimator in Ref. 3. There the optimum local reference waveform is essentially the derivative $\vec{y}'(x)$ of the $\vec{y}(x)$ having the useful number of harmonics.

The work reported above has benefited significantly from conversations with S. A. Butman and J. R. Lesh.

References

1. Rauch, L. L., *Improved Demodulation and Predetection Recording,* USAF Research Report, Contract No. AF-33(616)-5796; February 28, 1961.

2. Davenport, W. B., and Root, W. L., *An Introduction to the Theory of Random Signals and Noise,* Section 8-6, McGraw-Hill; 1958.

3. Layland, J. W., "On the Optimum Correlation Function for a First-Order Tracking Loop," *Space Program Summary 37-50,* Vol. III, pages 284-287. Jet Propulsion Laboratory, Pasadena, Calif., April 30, 1968.

4. Goldstein, R. M., "Ranging with Sequential Components", *Space Program Summary 37-52,* Vol II, pages 46-69. Jet Propulsion Laboratory, Pasadena, Calif., July 31, 1968.

5. Butman, S., "On Estimating the Phase of a Square Wave in White Noise," *Space Program Summary 37-53,* Vol. II, pages 200-209. Jet Propulsion Laboratory, Pasadena, Calif., October 31, 1968.

6. Butman, S., "The Effect of Amplitude Uncertainty on Estimating Phase of a Square Wave," *Space Program Summary 37-58.* Jet Propulsion Laboratory, Pasadena, Calif., August 31, 1969.

Economic Evaluation of DSS 13 Unattended Operations Demonstration

D. S. Remer
Harvey Mudd College

I. Eisenberger
Communications Systems Research Section

G. Lorden
California Institute of Technology

This article presents the goals and data collection requirements to be used for the economic and performance evaluation indices and life-cycle cost parameters for the upcoming operations demonstration of an automated Deep Space Station run unattended and controlled remotely from JPL. These evaluation indices will compare the remote operation of telemetry at DSS 13 with the cost and performance of a comparable manned operation at DSS 11. A description is presented of the data that needs to be collected, how the data will be analyzed, and what can and cannot be learned from this operations demonstration.

I. Introduction

In the last decade there has been an increasing emphasis on improving the cost effectiveness of DSN use of NASA resources for tracking and data acquisition. To this end, the DSN has been introducing automation into the stations to reduce manpower and to improve network productivity. For example, the crew size at DSS 12 has gone from 15 people in 1967 to 3 in 1977. The change in crew size is shown below for the last decade:

Year	DSS 12 Crew Size
1967	15
1972	8
1975	4
1977	3

The next step in this process is to attempt to run a completely automated station, i.e., unattended and controlled remotely from JPL in Pasadena. Accordingly, during the last half of 1978, an unattended operations demonstration test will be performed at DSS 13.

There are three objectives of this demonstration. First, to see if unattended remote operation can be accomplished. Second, to collect operations data from which an evaluation of unattended operation can be performed, and third to provide a single point, remote, unattended control of DSS 13 to accomplish Voyager spacecraft telemetry reception and transmission to the Network Operations Control Center (NOCC) via DSS 12 during the DSS 12 downtime for S/X upgrade from a 26-meter to a 34-meter antenna.

This article will be concerned with the second objective, namely, what data do we collect, how do we analyze the data,

and what can and cannot be learned from this demonstration test.

The overall objective will be to compare the unattended remote operation for telemetry at DSS 13 with the cost and performance of a comparable operation at DSS 11. This comparison will hopefully give us valuable insight into the advantages and disadvantages of automated, remote operation compared to our present method of operation.

In Section II, the detailed goals are outlined for the test. Also the indices and figures of merit used for comparing DSS 13 performance with DSS 11 are introduced. In Section III, the data requirements from DSS 13 and DSS 11 are outlined. In Section IV, some of the limitations on extrapolation of the test results are explored. Finally, in Section V we summarize the salient features of the report.

II. Goals for Unattended Operations Demonstration

A. Background

One of the primary DSN program goals of the demonstration is to collect an operations data base from which the values of life cycle cost (LCC) parameters can be computed. However, in the original program goal these LCC parameters were not defined. The major purpose of this report is to summarize our work on developing these parameters so that the DSN program goal can be reached.

B. Comparison Indices

Described below are eight comparison indices that we will use to evaluate the cost and performance of telemetry activities at DSS 13 and DSS 11. The comparison is limited to the unattended remote operation of telemetry at DSS 13 with a comparable operation at DSS 11.

The first two comparison indices are efficiency and productivity. These indices are to be determined for both DSS 13 and DSS 11 in telemetry mode. These indices are referenced in Tables 1 and 3.

I-A (1) Efficiency ratio = $\frac{\text{EUH/unit time}}{\text{SOH/unit time}}$

where

SOH = station operating hours, those hours when a station is required to be available to conduct DSN activities. (SOH are usually set at 40, 80, 120, 160, or 168 hours per week.)

EUH = end user hours, those station operating hours where spacecraft data, test, or training information is obtained for the end user.

I-A (2) Productivity ratio = $\frac{\text{EUH/unit time}}{\text{M\&O MH/unit time}}$

where

MH = manhours

M&O MH = the sum of the manhours spent on operations, corrective maintenance, preventive maintenance and training.

The next three indices are to be used for comparing the maintenance and operations costs at each station per end user hour. There is a separate index for operations and another index for maintenance because it is expected that unattended operation will lower operating manpower costs, but may increase maintenance costs because of the extra equipment required.

I-A (3) End user's hourly M&O cost = $\frac{\text{M\&O cost \$/unit time}}{\text{EUH/unit time}}$.

I-A (4) End user's hourly operations cost = $\frac{\text{Operations cost \$/unit time}}{\text{EUH/unit time}}$.

I-A (5) End user's hourly maintenance cost = $\frac{\text{Maintenance cost \$/unit time}}{\text{EUH/unit time}}$.

The next three indices are to be used for comparing maintenance and operations costs at DSS 13 and DSS 11 per station operating hour.

I-A (6) Station hourly M&O cost = $\frac{\text{M\&O cost \$/unit time}}{\text{SOH/unit time}}$.

I-A (7) Station hourly operations cost = $\frac{\text{Operations cost \$/unit time}}{\text{SOH/unit time}}$.

I-A (8) Station hourly maintenance cost = $\frac{\text{Maintenance cost \$/unit time}}{\text{SOH/unit time}}$.

C. Automated vs Nonautomated Station Ratios

In addition to the previous eight comparison indices, we have developed eleven ratios to compare automated with

nonautomated station operation. The first ratio compares operating manhours per station operating hour.

I-B (1) Operating MH ratio $= \dfrac{\text{operations MH/SOH(A)}}{\text{operations MH/SOH(NA)}}$

where

A = automated station (DSS 13)

NA = Nonautomated station (DSS 11 or 12).

The next four comparison ratios, given below, are designed to compare corrective maintenance at DSS 13 with that at DSS 11. Corrective maintenance is divided into two categories, according to whether station downtime is or is not a consequence of the necessity for maintenance action. Also corrective maintenance is compared per unit time, e.g., week, month, year, and per station operating hour.

I-B (2) Corrective (DT) MMH ratio (1) =

$$\frac{\text{corrective (DT)MMH/unit time (A)}}{\text{corrective (DT)MMH/unit time (NA)}}$$

where

MMH = maintenance manhours

DT means that downtime resulted from the necessity for maintenance action.

I-B (3) Corrective ($\overline{\text{DT}}$) MMH ratio (1) =

$$\frac{\text{corrective } (\overline{\text{DT}}) \text{ MMH/unit time (A)}}{\text{corrective } (\overline{\text{DT}}) \text{ MMH/unit time (NA)}}$$

where

$\overline{\text{DT}}$ means that no downtime was necessary

I-B (4) Corrective (DT) MMH ratio (2) =

$$\frac{\text{corrective (DT) MMH/SOH (A)}}{\text{corrective (DT) MMH/SOH (NA)}}.$$

I-B (5) Corrective ($\overline{\text{DT}}$) MMH ratio (2) =

$$\frac{\text{corrective } (\overline{\text{DT}}) \text{ MMH/SOH (A)}}{\text{corrective } (\overline{\text{DT}}) \text{ MMH/SOH (NA)}}.$$

Automation is expected to reduce pre- and postcalibration time per track. The following efficiency ratio will show this potential improvement if it occurs.

I-B (6) Pre- and postcal efficiency ratio =

$$\frac{\text{pre- and postcal SOH/track (A)}}{\text{pre- and postcal SOH/track (NA)}}.$$

Another very significant variable is downtime hours, i.e., time during which the end user suffers a loss of data as a result of human error or equipment failure. At the present time, it is not clear whether we will have more or less downtime with automation. With an automated station, human operating errors should be reduced, but reaction time to repair equipment may be increased. The following ratio will give us an indication of the downtime associated with automated and nonautomated equipment in this test.

I-B (7) Downtime ratio $= \dfrac{\text{downtime hours/SOH (A)}}{\text{downtime hours/SOH (NA)}}$.

The following four cost ratios will be important life cycle cost parameters.

I-B (8) Operations cost ratio $= \dfrac{\text{operations cost/SOH (A)}}{\text{operations cost/SOH (NA)}}$.

I-B (9) Corrective MC ratio =

$$\frac{\text{corrective MC/unit time (A)}}{\text{corrective MC/unit time (NA)}}$$

where

MC = maintenance cost.

I-B (10) M&O cost ratio (1) $= \dfrac{\text{M\&O cost/unit time (A)}}{\text{M\&O cost/unit time (NA)}}$.

I-B (11) M&O cost ratio (2) $= \dfrac{\text{M\&O cost/SOH (A)}}{\text{M\&O cost/SOH (NA)}}$.

All of the above 8 indices and 11 ratios are summarized in Table 1.

D. Other Goals:

In addition to the previously described quantitative goals, there are several very important qualitative goals. These are:

II- (1) To determine advantages and disadvantages of operating in an unattended mode.

II- (2) To list some trouble areas that may require design or operating changes or that may provide inputs to a future automated station design.

II- (3) To try to determine the change in effectiveness of unattended operation due to learning and how rapid this change is likely to progress.

II- (4) To determine the difference in quality of the telemetry received by DSS 13 in unattended operation and that received by DSS 11 or DSS 12 in attended operation.

II- (5) To try to determine the effect on the availability of the telemetry reception system at DSS 13 of adding control equipment.

III. Data Required to Achieve Goals

Twenty-five data types, listed in Table 2, must be collected to achieve the above goals.

Data categories 1-11 are variables appearing in the comparison indices discussed in Section II. Data types 12 and 15 will permit direct comparisons of the quantity and quality of telemetry data received.

For the extra equipment required in DSS 13 to do unattended operation, it is important to have adequate data to evaluate failure rates, mean-time-to-repair, and man-hours needed for repairs. This information must be compiled separately from the corresponding data for the conventional equipment. This accounts for data categories 16–19. Comparison of the results with DSS 11 is provided for by data category 25.

The remaining data types, 20–24, correspond directly with the goals for qualitative evaluation discussed in Subsection IID.

IV. Limitations

The limitations on the conclusions to be drawn from the demonstration are principally caused by the following factors:

(1) DSS 13 is not a standard DSN station, nor are its personnel or hours of operation comparable.

(2) The method of automation used in the demonstration, adding control equipment to existing subsystems, is different from the integration of operating and control equipment that would actually be used to implement unattended operation in the DSN.

(3) The cost to design and implement integrated operating and control equipment from demonstration data cannot be estimated.

(4) Some effects of automation (e.g., on wear or reliability of subsystems) may not show up at all in the demonstration period, or may be masked by normal statistical fluctuations.

(5) Sometimes it will be impossible to distinguish if downtime was caused by standard subsystems, or control equipment added for unattended operation, or the interface between the standard subsystems and the new control equipment.

(6) Segregating an "equivalent" system at DSS 11 to compare with DSS 13 and allocating operators time to that equivalent system will be difficult and may introduce significant variances in the results.

(7) The results from this demonstration test are limited to the effectiveness of unattended operation for the telemetry mode of operation only.

V. Summary

To permit evaluation of the unattended operation demonstration at DSS 13, a data base will be constructed during this demonstration test. The types of data included are those needed to calculate selected indices of performance and costs, using as a basis of comparison the telemetry operations of DSS 11 during the same period.

Additionally, subjective impressions will be recorded during the demonstration in an effort to gain insight into the problems and potential benefits of unattended operation.

Although there are significant limitations upon the extrapolation of the data base to project the costs and benefits of DSN automation, we can expect to learn a great deal from the demonstration.

Table 1. Summary – Indices

		X DSS 13	Y DSS 11 or 12	Ratio X/Y
	Index: I–A			
1.	Efficiency ratio = $\frac{\text{EUH/unit time}}{\text{SOH/unit time}}$			
2.	Productivity ratio = $\frac{\text{EUH/unit time}}{\text{M\&O MH/unit time}}$			
3.	End user's hourly M&O cost = $\frac{\text{M\&O cost \$/unit time}}{\text{EUH/unit time}}$			
4.	End user's hourly OC = $\frac{\text{OC \$/unit time}}{\text{EUH/unit time}}$			
5.	End user's hourly MC = $\frac{\text{MC \$/unit time}}{\text{EUH/unit time}}$			
6.	Station hourly M&O cost = $\frac{\text{M\&O cost \$/unit time}}{\text{SOH/unit time}}$			
7.	Station hourly OC = $\frac{\text{OC \$/unit time}}{\text{SOH/unit time}}$			
8	Station hourly MC = $\frac{\text{MC \$/unit time}}{\text{SOH/unit time}}$			
	Index: I–B			
1.	Operations MH/SOH			
2.	Corrective (DT) MMH/unit time			
3.	Corrective ($\overline{DT}$) MMH/unit time			
4.	Corrective (DT) MMH/SOH			
5.	Corrective ($\overline{DT}$) MMH/SOH			
6.	Pre- and postcal SOH/track			
7.	Downtime hours/SOH			
8.	Operations cost/SOH			
9.	Corrective MC/unit time			
10.	M&O cost/unit time			
11.	M&O cost/SOH			

Table 2. Data required to achieve overall goals of DSS 13 unattended test demonstration

I. Data required from DSS 13 and DSS 11 or DSS 12 in telemetry mode:

1. End user hours.
2. Station operating hours.
3. Operations manhours.
4. Preventive maintenance manhours.
5. Corrective (DT) maintenance manhours.
6. Corrective ($\overline{DT}$) maintenance manhours.
7. Pre- and postcal station operating hours/track.
8. Downtime hours during operation.
9. Training manhours.
10. Cost per operating manhour.
11. Cost per maintenance manhour.
12. Quality of telemetry data received.
13. Cost of materials used to operate equipment.
14. Cost of materials used to maintain equipment.
15. Hours of telemetry data received, arranged by spacecraft.

II. Additional data required from DSS 13:

16. Corrective maintenance manhours for automated equipment.
17. Corrective maintenance manhours for automating equipment.
18. Failure history and time to repair for automated equipment.
19. Failure history and time to repair for automating equipment.
20. An evaluation of the differences in ease of operation between unattended and attended operations. This information should come from all personnel associated with the operation of DSS 13.
21. A record of all unusual occurrences at DSS 13 whose cause cannot be immediately ascertained, presumably because of remote control.
22. A list of advantages and disadvantages of operating in an unattended mode as they become apparent to operating and supervisory personnel.
23. A list of trouble areas that may be due to the design of an automating piece of equipment or its interface with the equipment it controls.
24. A periodic evaluation of the changes in effectiveness of unattended operation due to learning as the test progresses.

III. Additional data required from DSS 11 or DSS 12:

25. Failure history and time to repair for equipment used for telemetry.

Definitions:

Station operating hours denotes the number of hours that a station is required to be available to conduct DSN activities (usually set at 40, 80, 120, 160, or 168 hours per week).

End user hours denotes the number of station operating hours in which spacecraft data, test, or training information is obtained for the end user.

DT means that downtime resulted from the necessity for maintenance action.

$\overline{DT}$ means no downtime was necessary.

Table 3. Data requirements to achieve each specific goal of DSS 13 unattended test demonstration

Goal indices	Data requirement indices
I-A1	I1, I2
A2	I1, I3, I4, I5, I6, I9
A3	I1, I3, I4, I5, I6, I9, I10, I11, I13, I14
A4	I1, I3, I9, I10, I13
A5	I1, I4, I5, I6, I11, I14
A6	I2, I3, I4, I5, I6, I9, I10, I11, I13, I14
A7	I2, I3, I9, I10, I13
A8	I2, I4, I5, I6, I11, I14
B1	I2, I3
B2	I5
B3	I6
B4	I2, I5
B5	I2, I6
B6	I7
B7	I2, I8
B8	I2, I3, I10
B9	I5, I6, I11
B10	I3, I4, I5, I6, I9, I10, I11, I13, I14
B11	I2, I3, I4, I5, I6, I9, I10, I11, I13, I14
II-1	II 20, 22
2	II 23
3	II 24
4	I 12
5	II 16, II 17, II 18, II 19, III 25

A New Hybrid Algorithm for Computing a Fast Discrete Fourier Transform

I. S. Reed
University of Southern California

T. K. Truong
Tracking and Data Acquisition Engineering

For certain long transform lengths, Winograd's algorithm for computing the discrete Fourier transform (DFT) is extended considerably. This is accomplished by performing the cyclic convolution, required by Winograd's method, with the Mersenne-prime number-theoretic transform developed originally by Rader. This new algorithm requires fewer multiplications than either the standard fast Fourier transform (FFT) or Winograd's more conventional algorithm.

I. Introduction

Several authors (Refs. 1 through 13) have shown that transforms over finite fields or rings can be used to compute circular convolutions without round-off error. Recently, Winograd (Ref. 14) developed a new class of algorithms which depend heavily on the computation of a cyclic convolution for computing the conventional DFT. This new algorithm, for a few hundred transform points, requires substantially fewer multiplications than the conventional FFT algorithm.

C. M. Rader (Ref. 3) defined a special class of finite Fourier-like transforms over $GF(q)$, where $q = 2^p - 1$ is a Mersenne prime for $p = 2,3,5,7,13,17,19,31,61 \cdots$ These number-theoretic transforms are used and specialized here to transform lengths of p points. The advantage of this transform over others is that it can be accomplished simply by circular shifts, i.e., no multiplications are needed (Ref. 3).

In this paper, it is shown that Winograd's algorithm can be combined with the above-mentioned number-theoretic transform over $GF(q)$ to yield a new algorithm for computing the discrete Fourier transform (DFT). By this means, a fast method for accurately computing the DFT of a sequence of real and complex numbers of very long transform lengths is obtained.

II. Cyclic Convolution

The following algorithm for the cyclic convolution of two sequences is based on ideas due to Winograd (Ref. 14). Let R be the field of rationals. Also let $X(u) = x_0 + x_1 u + x_2 u^2 + \cdots + x_n u^{n-1}$, $Y(u) = y_0 + y_1 u + y_2 u^2 + \cdots + y_n u^{n-1}$ be two polynomials over R. The product $T(u) = X(u) \cdot Y(u)$ can be computed by

$$T(u) = X(u) \cdot Y(u) \bmod \prod_{i=0}^{2n-2} (u - \alpha_i) \qquad (1)$$

where $\alpha_i \epsilon R$. It is shown in (Ref. 14) that a minimum of $2n - 1$ multiplications are needed to compute Eq. (1).

It is readily shown that the cyclic convolution of $X(u)$ and $Y(u)$ is the set of coefficients of the polynomial,

$$T(u) = X(u) \cdot Y(u) \bmod (u^n - 1)$$

Let the polynomial $u^n - 1$ be factored into irreducible relatively prime factors, i.e.,

$$u^n - 1 = \prod_{i=1}^{k} g_i(u)$$

where

$$(g_i(u),\ g_j(u)) \neq 1 \text{ for } i \neq j$$

Then $T(u) \bmod g_i(u)$ for $i = 1,2,\cdots,k$ can be computed, using Eq. (1). Finally, the Chinese remainder theorem is used to evaluate $T(u)$ from these residues. The above summarizes Winograd's method for performing a cyclic convolution.

The following theorem, due to Winograd (Ref. 15), will be needed.

Theorem 1: Let a and b be relatively prime positive integers and A be the cyclic $ab \times ab$ matrix, given by

$$\mathrm{A}(x, y) = f\,(x + y \bmod a \cdot b),\ 0 \leqslant x, y < ab$$

If π is a permutation of the set of integers $\{0,1,\cdots,ab-1\}$, let

$$\mathrm{B}(x, y) = \mathrm{A}(\pi(x), \pi(y))$$

Then there exists a permutation π such that, if B is partitioned into $b \times b$ submatrices, each submatrix is cyclic and the submatrices form an $a \times a$ cyclic matrix.

It was shown in (Refs. 15 and 16) that the number of multiplications needed to perform a circular convolution of 2,3,4,5,6, and 8 points is 2,4,5,10,8, and 14 multiplications, respectively. To compute the cyclic convolution of two longer sequences of integers, a p-point transform over $GF(q)$ will be utilized here. Since the latter transform can be evaluated without multiplications (Ref. 3), it can be used with considerable advantage to compute the cyclic convolution of two p-point real number sequences. Hence, for the transform over $GF(q)$, the number of integer multiplications needed to perform a circular convolution is precisely p, excluding the multiplications by p^{-1} in the inverse transform.

III. The DFT When the Transform Length d is a Prime $d = q'$

The DFT is defined by

$$A_j = \sum_{i=0}^{d-1} a_i w^{ij}$$

where w is a d-th root of unity. Let

$$A_0 = \sum_{i=0}^{d-1} a_i \tag{2a}$$

and

$$A_j = a_0 + B_j \quad \text{for } j = 1, 2, \cdots, d-1$$

where

$$B_j = \sum_{i=1}^{d-1} a_i w^{ij}$$

That is, let

$$\bar{\mathrm{B}} = \mathrm{W}\bar{\mathrm{a}} \tag{2b}$$

where W is the $(d-1) \times (d-1)$ matrix (w^{ij}), and $\bar{\mathrm{a}}$, $\bar{\mathrm{B}}$ are the column matrices (a_i) and (B_k), respectively. If $d = q'$ is a prime, then by (Ref. 13), one can find an element α in $GF(q')$ that generates its cyclic multiplicative subgroup of $q' - 1$ elements. Using the element α, a cyclic permutation of the elements of $GF(q')$ can be defined by

$$\sigma = \begin{pmatrix} 1, 2, \cdots, q'-2, q'-1 \\ \alpha, \alpha^2, \cdots, \alpha^{q'-2}, \alpha^{q'-1} \end{pmatrix} \tag{2c}$$

With this permutation, one can permute the indices of $\overline{B}$, $\overline{a}$, W defined in Eq. (2b) so that the matrix $\widetilde{W} = (w^{\sigma(i)\sigma(j)})_{i,\, j\neq 0}$, is cyclic. That is,

$$B_{\sigma(j)} = \sum_{i=1}^{q'-1} a_{\sigma(i)}\, w^{\sigma(i)\sigma(j)}$$

$$= \sum_{i=1}^{q'-1} a_{\sigma(i)}\, w^{\alpha^{i+j}}$$

$$= \sum_{i=1}^{q'-1} a_{\sigma(i)}\, w^{\sigma(i+j)} \quad \text{for } j = 1, 2, \cdots, q'-1 \qquad (3)$$

Thus, $B_{\sigma(j)}$ is a cyclic convolution of $a_{\sigma(i)}$ and $w^{\sigma(i)}$ for $j = 1, 2, \cdots, q'-1$.

Let $q'-1 = p_1 \cdot p_2 \cdots p_r$, where $(p_i, p_j) = 1$ for $i \neq j$. If one lets $a_1 = p_1 \cdot p_2 \cdots p_{r-1}$ and $b_1 = p_r$, by Theorem 1 the cyclic matrix W can be partitioned into $b_1^2 = p_r^2$ cyclic matrices each of block size $a_1 \times a_1$. Next let $a_1 = a_2 \times b_2$, where $a_2 = p_1 \cdots p_{r-2}$ and $b_2 = p_{r-1}$. If a_2 is not a prime, then each $a_1 \times a_1$ cyclic matrix can be partitioned into b_2^2 cyclic matrices of block size $a_2 \times a_2$. In general, $a_i = a_{i+1} \cdot b_{i+1}$, where b_{i+1}^2 is a prime. If $a_{i+1} \neq 1$, then each $a_i \times a_i$ cyclic matrix can be partitioned into b_{i+1}^2 cyclic matrices of block size $a_{i+1} \times a_{i+1}$. Otherwise, the procedure terminates. If the number of multiplications used to compute the cyclic convolution of p_i points is m_i for $i = 1, 2, \cdots, r$, then Winograd has shown in Ref. 14 that the number of multiplications for computing a q'-point DFT is equal to $N = m_1 \cdot m_2 \cdots m_r$.

For most applications, the two Mersenne primes $2^{31} - 1$ and $2^{61} - 1$ will provide enough bit accuracy and dynamic range for computing the DFT. For these primes, we choose the prime q' to have the form

$$q' = 1 + (a \cdot 2^n) \cdot p \quad \text{for } n = 1,2,3$$

where $p = 31$ or 61 and $a = 3$ or 5. Such values for the prime q' are 367, 373, 733, 1831, 1861, and 2441.

If $d = q'$ is the transform length of the DFT, then, by Theorem 1, there exists a permutation of rows and columns so that cyclic matrix W can be partitioned into blocks of $p \times p$ cyclic matrices, such that the blocks from a $(2^n \cdot a) \times (2^n \cdot a)$ cyclic matrix. This cyclic matrix can be reduced further by Winograd's method. First $q'-1 = 2^n \cdot a \cdot p$ is an even number, and $w^{2^n \cdot ap} = w^{-1}$ where w is the d-th root of unity in the field of complex numbers. For such a case, Winograd showed that the elements in the $p \times p$ cyclic matrices finally required by the transform are either all real or imaginary numbers. To show this, consider the case $n = 1$. For this case, $q'-1 = 2 \cdot a \cdot p$. The permutation in Eq. (2c) is given by

$$\sigma = \begin{pmatrix} 1, 2 \cdots ap-1, ap, ap+1 \cdots 2ap \\ \alpha, \alpha^2 \cdots \alpha^{ap-1}, \alpha^{ap}, \alpha^{ap+1} \cdots \alpha^{2ap} \end{pmatrix}$$

where α is a generator of the multiplicative subgroup consisting of $q'-1 = 2ap$ elements in $GF(q')$. Applying the above permutation to Eq. (2b) and using the fact that $\alpha^{ap} \equiv -1 \bmod q'$, one obtains the cyclic matrix equation in terms of w as follows:

$$\begin{bmatrix} b_{\sigma(1)} \\ b_{\sigma(2)} \\ \cdot \\ \cdot \\ \cdot \\ b_{\sigma(2ap)} \end{bmatrix} = \begin{bmatrix} w^{\alpha^2} & w^{\alpha^3} & w^{\alpha^4} & \cdots & w^{-1} & w^{-\alpha} & w^{-\alpha^2} & \cdots & w^1 & w^{\alpha} \\ w^{\alpha^3} & w^{\alpha^4} & \cdots & w^{-1} & w^{-\alpha} & w^{-\alpha^2} & \cdots & w^1 & w^{\alpha} & w^{\alpha^2} \\ \cdot & & & & & & & & & \\ \cdot & & & & & & & & & \\ \cdot & & & & & & & & & \\ w^{\alpha^1} & w^{\alpha^2} & \cdots & & & w^{-1} & w^{-\alpha} & \cdots & w^1 & \end{bmatrix} \begin{bmatrix} a_{\sigma(1)} \\ a_{\sigma(2)} \\ \\ \\ \\ a_{\sigma(2ap)} \end{bmatrix} \qquad (4)$$

Let $\varphi_0 = b_{\sigma(1)}, \varphi_1 = b_{\sigma(2)}, \cdots, \varphi_{2ap-1} = b_{\sigma(2ap)}, x_0 = w^{\alpha 1}, x_1 = w^{\alpha 2}, \cdots, x_{ap} = w^{-1}, \cdots, x_{2ap} = w^1, y_0 = a_{\sigma(1)}, y_1 = a_{\sigma(2)}, \cdots, y_{2ap-1} = a_{\sigma(2ap)}$. Then Eq. (4) becomes

$$\begin{bmatrix} \varphi_0 \\ \varphi_1 \\ \cdot \\ \cdot \\ \cdot \\ \varphi_{m-2} \\ \varphi_{m-1} \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_{m/2} & \cdots & x_m & x_0 \\ x_2 & x_3 & \cdots & x_{m/2+1} & \cdots & x_0 & x_1 \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ x_m & x_0 & x_1 & \cdots & x_{m/2-1} & \cdots & x_{m-1} \\ x_0 & x_1 & x_2 & \cdots & x_{m/2} & \cdots & x_m \end{bmatrix} \begin{bmatrix} y_0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_{m-1} \end{bmatrix} \tag{5}$$

where $m = 2ap$.

By Theorem 1, the above cyclic $2 \times ap$ matrix equation can be partitioned into blocks of $ap \times ap$ cyclic matrices, so that the blocks form a 2×2 cyclic matrix. To illustrate this, note first that 2 and $a \cdot p$ are relatively prime. Thus, the Chinese remainder theorem, an isomorphism

$$k \rightarrow (k_1, k_2)$$

exists betwen an integer k modulo m and the pairs of integers k_1 and k_2 modulo 2 and $a \cdot p$, respectively. This relationship between k and (k_1, k_2) is

$$k = k_1 M_1^{-1} + k_2 M_2^{-1} \bmod m$$

where M_1^{-1} and M_2^{-1} satisfy the congruences $a \cdot pM_1^{-1} \equiv 1 \bmod 2$ and $2M_2^{-1} \equiv 1 \bmod a \cdot p$, respectively.

Let the variables $y_k = y_{(k_1, k_2)}$, $x_k = x_{(k_1, k_2)}$ and $\varphi_k = \varphi_{(k_1, k_2)}$ be rearranged in such a manner that when the first component k of the index pair (k_1, k_2) is 0, component k_2 is in ascending order, and when component k_1 is set to 1, component k_2 also is in ascending order. The variables $x_{(k_1, k_2)}$ for Eq. (5) are then rearranged in the order

$$x_{(0,0)}, x_{(0,1)}, x_{(0,2)}, \cdots, x_{(0,ap-1)}, x_{(1,0)}, x_{(1,1)}, x_{(1,2)}, \cdots, x_{(1,ap-1)}$$

C-3

If such a rearrangement is made also on the variables y_k, x_k, and φ_k, respectively, the cyclic convolution of Eq. (5) has the form

$$\begin{bmatrix} \varphi_{(0,0)} \\ \varphi_{(0,1)} \\ \vdots \\ \varphi_{(0,ap-1)} \\ \varphi_{(1,0)} \\ \vdots \\ \varphi_{(1,ap-1)} \end{bmatrix} = \begin{bmatrix} x_{(1,1)}, x_{(1,2)} \cdots, x_{(1,0)}, x_{(0,1)} x_{(0,2)} \cdots x_{(0,0)} \\ \vdots \\ x_{(1,ap-1)} x_{(1,0)} \cdots x_{(1,ap-2)} x_{(0,ap-1)} x_{(0,0)} \cdots x_{(0,ap-2)} \\ x_{(1,0)} \quad x_{(1,1)} \cdots x_{(1,ap-1)} x_{(0,0)} \quad x_{(0,1)} \cdots x_{(0,ap-1)} \\ x_{(0,1)} \quad x_{(0,2)} \cdots x_{(0,0)} \quad x_{(1,1)} \quad x_{(1,2)} \cdots x_{(1,0)} \\ \vdots \\ x_{(0,ap-1)} x_{(0,0)} \cdots x_{(0,ap-2)} x_{(1,ap-1)} x_{(1,0)} \cdots x_{(1,ap-2)} \\ x_{(0,0)} \quad x_{(0,1)} \cdots x_{(0,ap-1)} x_{(1,0)} \quad x_{(1,1)} \cdots x_{(1,ap-1)} \end{bmatrix} \begin{bmatrix} y_{(0,0)} \\ y_{(0,1)} \\ \vdots \\ y_{(0,ap-1)} \\ y_{(1,0)} \\ \vdots \\ y_{(1,ap-1)} \end{bmatrix} \tag{6}$$

Observe that the matrix Eq. (6) can be further reduced to block form as follows

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} A & B \\ B & A \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{7}$$

where

$$x_1 = \begin{bmatrix} \varphi_{(0,0)} \\ \varphi_{(0,1)} \\ \vdots \\ \varphi_{(0,ap-1)} \end{bmatrix}, \qquad x_2 = \begin{bmatrix} \varphi_{(1,0)} \\ \varphi_{(1,1)} \\ \vdots \\ \varphi_{(1,ap-1)} \end{bmatrix}$$

$$y_1 = \begin{bmatrix} y_{(0,0)} \\ y_{(0,1)} \\ \cdot \\ \cdot \\ \cdot \\ y_{(0,ap-1)} \end{bmatrix}, \quad y_2 = \begin{bmatrix} y_{(1,0)} \\ y_{(1,1)} \\ \cdot \\ \cdot \\ \cdot \\ y_{(1,ap-1)} \end{bmatrix}$$

$$A = \begin{bmatrix} x_{(1,1)} & x_{(1,2)} \cdots x_{(1,0)} \\ \cdot & \\ \cdot & \\ \cdot & \\ \cdot & \\ \cdot & \\ \cdot & \\ x_{(1,ap-1)} & x_{(1,0)} \cdots x_{(1,ap-2)} \\ x_{(1,0)} & x_{(1,1)} \cdots x_{(1,ap-1)} \end{bmatrix}$$

and

$$B = \begin{bmatrix} x_{(0,1)} & x_{(0,2)} \cdots x_{(0,0)} \\ \cdot & \\ \cdot & \\ \cdot & \\ \cdot & \\ \cdot & \\ \cdot & \\ x_{(0,ap-1)} & x_{(0,0)} \cdots x_{(0,ap-2)} \\ x_{(0,0)} & x_{(0,1)} \cdots x_{(0,ap-1)} \end{bmatrix}$$

Since

$$x_{(0,ap-1)} = w^{\alpha^{(0,ap-1)+(1,1)}} = w^{\alpha^{(1,0)}} = w^{-1}$$

then

$$x_{(1,j)} = w^{\alpha(1,j)+(1,1)} = w^{\alpha(0,j+1)} = w^{\alpha(1,0)+(1,j+1)}$$

$$= w^{-\alpha(1,j+1)} = w^{-\alpha(0,j)+(1,1)} = x^*_{(0,j)}$$

for $j = 0, 1, \cdots, ap - 1$ where $*$ denotes complex conjugation. Thus, in (7), the cyclic matrix A is the complex conjugate of the cyclic matrix B, i.e.,

$$\mathrm{A} = \mathrm{B}^* \tag{8}$$

The matrix Eq. (7) can be obtained by computing the set of coefficients of

$$T(u) \equiv (\mathrm{B} + Au) \cdot (y_2 + y_1 u) \bmod (u^2 - 1) \tag{9}$$

where $u^2 - 1 \equiv (u - 1)(u + 1)$ and $u - 1$ and $u + 1$ are relatively prime polynomials.

Taking the congruences of $T(u)$ in Eq. (9) modulo $u - 1$ and $u + 1$, respectively,

$$T_1(u) \equiv (\mathrm{B} + \mathrm{A}) \cdot (y_2 + y_1) \bmod u - 1 \tag{10a}$$

and

$$T_2(u) \equiv (\mathrm{B} - \mathrm{A}) \cdot (y_2 - y_1) \bmod u + 1 \tag{10b}$$

By the Chinese remainder theorem, $T(u)$ can be reconstituted from Eqs. (10a) and (10b) as follows:

$$T(u) = 2^{-1}[(\mathrm{B} + \mathrm{A}) \cdot (y_2 + y_1) - (\mathrm{B} - \mathrm{A})(y_2 - y_1)$$

$$+ ((\mathrm{B} + \mathrm{A}) \cdot (y_2 + y_1) + (\mathrm{B} - \mathrm{A}) \cdot (y_2 - y_1))\, u]$$

$$= \mathrm{x}_1 + \mathrm{x}_2\, u$$

This is reexpressed in matrix form as

$$\begin{bmatrix} \mathrm{x}_1 \\ \\ \mathrm{x}_2 \end{bmatrix} = \begin{bmatrix} (\mathrm{B} + \mathrm{A}) \cdot (y_2 + y_1) + (\mathrm{A} - \mathrm{B}) \cdot (y_2 - y_1) \\ \\ (\mathrm{A} + \mathrm{B}) \cdot (y_2 + y_1) - (\mathrm{A} - \mathrm{B}) \cdot (y_2 - y_1) \end{bmatrix} \tag{11}$$

By Eq. (8), the elements of the cyclic matrices (B + A) and (A - B) in Eq. (11) are evidently real and imaginary numbers. Since $(a, p) = 1$, again by Theorem 1, the cyclic matrices (B + A) and (A - B) can be partitioned into blocks of $p \times p$ cyclic matrices such that the blocks are $a \times a$ cyclic matrices. Thus, the elements of these $p \times p$ cyclic matrix blocks are either real numbers or imaginary numbers, never complex numbers. Hence, if the input datum is real, then a multiplication by an element in such a $p \times p$ cyclic matrix requires only one real multiplication. If the input datum is a complex number, then a multiplication by an element in such a $p \times p$ cyclic matrix requires two real multiplications.

Using a procedure precisely similar to that used above for $n = 1$, it can be shown that the elements in the required $p \times p$ cyclic matrices of the $2^n \cdot ap$ cyclic matrix for $n = 2, 3$ are also either real numbers or imaginary numbers. It was pointed out in the last section that a transform of length p over $GF(q)$ can be used to compute the cyclic convolution of p real number points. The number of multiplications needed to perform this convolution is p. If one combines this with the number of multiplications needed for Winograd's algorithm for the prime q', the total number of multiplications required to perform a DFT of $d = q'$ real or complex number points can be computed. The results are shown in Table 1.

It has been shown that Winograd's algorithm can be combined with a transform over $GF(q)$ to yield a new rather fast hybrid algorithm for computing the DFT of real and complex values. In this algorithm, it is necessary to compute the cyclic convolution of p real number points. This cyclic convolution of two p-point sequences of real number points is given by

$$c_k = \sum_{n=0}^{p-1} e_n f_{(k-n)} \quad \text{for } k = 0,1,2,\cdots,p-1 \tag{12}$$

where $c_k, e_n, f_n \in GF(q)$ and $(k - n)$ denotes the residue of $k - n \bmod p$. To compute this convolution, the components of the truncated real number e_n and f_n must be converted first to integers a_n and b_n with dynamic ranges, A and B, respectively. In Refs. 6 and 9, it was shown that a sufficient dynamic range constraint for A and B is

$$A \leqslant \frac{q-1}{2Bp} \tag{13a}$$

If $A = B$, Eq. (13a) reduces to

$$A \leqslant \left[\sqrt{\frac{q-1}{2p}} \right] \tag{13b}$$

where $[x]$ denotes the greatest integer less than x.

If the circular convolution of a_n and b_n is denoted by c'_k for $k = 0,1,2,\cdots,p-1$, then, using the procedure described in the example of Ref. 7, c'_k can be obtained by using fast transforms over $GF(q)$. c_k in Eq. (12) can be obtained by scaling back c'_k to the scale of the original real numbers for $k = 0,1,2,\cdots,p-1$. Evidently, the only error made in this computation of c'_k is the truncation error.

The dynamic range constraint, A, of the input sequence given in Eq. (13b) is generally very pessimistic. It was shown in Ref. 17 that for integer convolutions, one can lessen the severity of the dynamic range constraint (13) and still maintain c_k in the interval $\pm(q-1)/2$ with a small probability of overflow.

To illustrate this new hybrid algorithm, consider the following example.

Example: Consider the DFT for $d = 7$ points. Let the input function be defined by

$$a_n = 1 \quad \text{for} \quad n = 0, 2$$

$$= 0 \quad \text{for} \quad n = 1, 3, 4, 5, 6$$

By Eq. (2a), this transform is

$$A_0 = \sum_{i=0}^{6} a_i = 2 + \hat{i}\,0 \tag{14a}$$

and

$$A_j = a_0 + b_j \quad \text{for } j = 1,2, \cdots, 6 \tag{14b}$$

where

$$b_j = \sum_{i=1}^{6-1} a_i w^{ij}, \quad w = e^{i2\pi/7}$$

For $d = 7$, the permutation σ is given by

$$\sigma = \begin{pmatrix} 1, & 2, & 3, & 4, & 5, & 6 \\ 3, & 2, & 6, & 4, & 5, & 1 \end{pmatrix}$$

Applying the above permutation to Eq. (14b), one obtains $\bar{\bar{B}} = \bar{\bar{W}}\,\bar{a}$ as

$$\begin{pmatrix} b_3 \\ b_2 \\ b_6 \\ b_4 \\ b_5 \\ b_1 \end{pmatrix} = \begin{pmatrix} w^2 & w^6 & w^4 & w^5 & w^1 & w^3 \\ w^6 & w^4 & w^5 & w^1 & w^3 & w^2 \\ w^4 & w^5 & w^1 & w^3 & w^2 & w^6 \\ w^5 & w^1 & w^3 & w^2 & w^6 & w^4 \\ w^1 & w^3 & w^2 & w^6 & w^4 & w^5 \\ w^3 & w^2 & w^6 & w^4 & w^5 & w^1 \end{pmatrix} \begin{pmatrix} a_3 \\ a_2 \\ a_6 \\ a_4 \\ a_5 \\ a_1 \end{pmatrix}$$

By Theorem 1, there exists a permutation π of rows and columns so that the above cyclic matrix can be partitioned into 2×2 block matrix of 3×3 cyclic blocks as follows:

$$\begin{pmatrix} b_3 \\ b_5 \\ b_6 \\ b_4 \\ b_2 \\ b_1 \end{pmatrix} = \begin{pmatrix} w^2 & w^1 & w^4 & w^5 & w^6 & w^3 \\ w^1 & w^4 & w^2 & w^6 & w^3 & w^5 \\ w^4 & w^2 & w^1 & w^3 & w^5 & w^6 \\ w^5 & w^6 & w^3 & w^2 & w^1 & w^4 \\ w^6 & w^3 & w^5 & w^1 & w^4 & w^2 \\ w^3 & w^5 & w^6 & w^4 & w^2 & w^1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \tag{15}$$

This matrix equation has the block form,

$$\begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{D} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix}$$

$$= 2^{-1} \begin{pmatrix} (\mathbf{C}+\mathbf{D})(\mathbf{Z}_1+\mathbf{Z}_2) + (\mathbf{C}-\mathbf{D})(\mathbf{Z}_1-\mathbf{Z}_2) \\ (\mathbf{C}+\mathbf{D})(\mathbf{Z}_1+\mathbf{Z}_2) - (\mathbf{C}-\mathbf{D})(\mathbf{Z}_1-\mathbf{Z}_2) \end{pmatrix}$$

$$= 2^{-1} \begin{pmatrix} \mathbf{E}+\mathbf{F} \\ \mathbf{E}-\mathbf{F} \end{pmatrix} \tag{16}$$

Since **C** and **D** are 3×3 cyclic matrices, it is evident that the matrices $\mathbf{C}+\mathbf{D}$ and $\mathbf{C}-\mathbf{D}$ are also 3×3 cyclic matrices. (Note that for a 6×6 cyclic matrix in Eq. (15), the powers of w in **E** and **F** in Eq. (16) are real numbers and imaginary numbers, respectively). In Eq. (16), **E** is

$$\mathbf{E} = \begin{pmatrix} e_0 \\ e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} -0.445, & 1.247, & -1.802 \\ 1.247, & -1.802, & -0.445 \\ -1.802, & -0.445, & 1.247 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \tag{17}$$

where approximately $1/2\ \mathrm{Re}(w^2 + w^5) = -0.445$, $1/2\ \mathrm{Re}(w^1 + w^6) = 1.247$, etc. Let $a_0 = -1.802$, $a_1 = -0.445$, $a_2 = 1.247$ and $y_0 = 0$, $y_1 = 1$, $y_2 = 0$. Then the matrix equation defined in Eq. (17) can be obtained by computing the convolution of the two sequences a_n and y_n. This requires using a transform over $GF(q)$. To avoid overflow, one needs to choose $q = 7$ so that the integer components of a_n, y_n lie in the interval $\pm(7-1)/2$.

By Ref. 7, the sequence of a_n is converted first to a sequence of integers x_n in the dynamic range $A = 2$. Since 2 is a third root of unity, the transform over $GF(7)$ of x_n is

$$X_k = \sum_{n=0}^{3-1} x_n \cdot 2^{nk} = -1 + 2^{2k} \quad \text{for } k = 0, 1, 2$$

Thus $X_0 = 0, X_1 = 3, X_2 = 1$.

Similarly, the transform over $GF(q)$ of sequence y_n is

$$Y_k = \sum_{n=0}^{3-1} y_n \cdot 2^{nk} = 2^k \quad \text{for } k = 0, 1, 2$$

That is, $Y_0 = 1$, $Y_1 = 2$, $Y_2 = 4$. Define $E_k = X_k \cdot Y_k$, i.e., $E_0 = 0$, $E_1 = 6$, $E_2 = 4$. These are the only integer multiplications needed to perform this DFT. The inverse transform of E_k is

$$e_n = 3^{-1} \sum_{k=0}^{3-1} E_k \cdot 2^{-nk} \quad \text{for } n = 0, 1, 2$$

or

$$e_0 = 1, e_1 = -1, e_2 = 0$$

In a similar fashion, matrix F, given in Eq. (16), can also be obtained as $f_0 = -\hat{i}, f_1 = \hat{i} \cdot 0, f_2 = -\hat{i}$. Thus, by Eq. (16), one obtains $b_1 = 1/2\,\hat{i}$, $b_2 = -1/2$, $b_3 = (1 - \hat{i})/2$, $b_4 = (1 + \hat{i})/2$, $b_5 = -1/2$, $b_6 = -\hat{i}/2$. Hence, finally $A_0 = 2 + \hat{i}\,0$, $A_1 = 1 + 1/2\hat{i}$, $A_2 = 1/2 + \hat{i}0$, $A_3 = 1/2(3 - \hat{i})$, $A_4 = 1/2(3 + \hat{i})$, $A_5 = 1/2 + \hat{i}0$, $A_6 = 1 - 1/2\hat{i}$. For this example, the dynamic range of $GF(7)$ is inadequate. Also there is a large truncation error due to the course approximation used for the roots of unity. Evidently, the DFT in this example has an accuracy of precisely two binary digits, including the sign bit. This example, though only illustrative, suggests that the large finite fields suggested above have more than adequate dynamic range to compute the DFT with small truncation error.

IV. Transforms of Very Long Sequences

To compute the DFT of much longer sequences than considered in the last section, let $d = d_1 \cdot d_2 \cdots d_r$, where $(d_i, d_j) = 1$ for $i \neq j$. By using the Chinese remainder theorem Ref. 18, it is shown by Winograd in Ref. 14 that the DFT matrix W can be transformed into the direct product of $W_1, W_2, \cdots, W_r$, where W_i is the matrix of a d_i-point DFT. Assume the number of multiplications used to perform the d_i-point DFT for $i = 1,2, \cdots, r$ is m_i. Then, the number of multiplications for computing a d-point DFT is $N = m_1 \cdot m_2 \cdots m_r$. To illustrate this, see Winograd's example for computing a 12-point DFT, given in Ref. 15. By the same procedure used in the computation of this example, the number of integer multiplications needed to perform the transforms of longer sequences of complex numbers can be obtained by using Table 1 above and Table I in Ref. 14. These numbers are given in Table 2. The present algorithm and conventional FFT algorithm (Ref. 19) are compared in Table 2 by giving the number of real multiplications needed to perform these algorithms. The number of real multiplications needed to perform a transform of a few thousand points is given in Table II of Ref. 14.

Acknowledgements

The authors wish to thank Dr. N. A. Renzetti, Manager of Tracking and Data Acquisition Engineering, and the members of the Advanced Engineering Group in that organization at the Jet Propulsion Laboratory for their earrly support, suggestions, and encouragement of the research that led to this paper.

This work was supported in part by NASA Contract No. NAS 7-100, and also in part by the U.S. Air Force Office of Scientific Research under Grant AFOSR-75-2798.

References

1. Pollard, J. M., "The Fast Fourier Transform in a Finite Field," *Math. Comput.*, 1971, Vol. 25, pp. 365–374.
2. Schonhage, A., and Strassen, V., "Schnelle Multiplication Grosser Zahlen," *Computing*, 1971, Vol. 7, pp. 281–292.
3. Rader, C. M., "Discrete Convolution via Mersenne Transforms," *IEEE Trans. Comp.*, 1972, Vol. C-21, pp. 1269–1273.
4. Agarwal, R. C., and Burrus, C. S., "Number Theoretic Transforms to Implement Fast Digital Convolution," *Proc. IEEE*, 1975, Vol. 63, pp. 550–560.
5. Agarwal, R. C., and Burrus, C. S., "Fast Convolution Using Fermat Number Transforms with Applications to Digital Filtering," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-22, No. 2, April 1974.
6. Reed, I. S., and Truong, T. K., "The Use of Finite Fields to Compute Convolution," *IEEE Trans.*, 1975, Vol. IT-21, pp. 208–213.
7. Reed, I. S., and Truong, T. K., "Complex Integer Convolution Over a Direct Sum of Galois Fields," *IEEE Trans.*, 1975, Vol. IT-21, pp. 657–661.
8. Vegh, E., and Leibowitz, L. M., "Fast Complex Convolution in Finite Rings," *IEEE Trans.*, 1976 Vol. ASSP-24, pp. 343–344.
9. Golomb, S. W., Reed, I. S., and Truong, T. K., "Integer Convolutions Over the Finite Field GF($3 \cdot 2^n + 1$)," *SIAM J. on Applied Math.*, Vol. 32, No. 2, March 1977.
10. Pollard, J. M., "Implementation of Number-Theoretic Transforms," *Electro. Lett.*, 1976, Vol. 12, pp. 378–379.
11. Liu, K. Y., Reed, I. S., and Truong, T. K., "Fast Number-Theoretic Transforms for Digital Filtering," *Electron. Lett.*, 1976, Vol. 12, pp. 644–646.
12. Reed, I. S., Truong, T. K., and Liu, K. Y., "A New Fast Algorithm for Computing Complex Number-Theoretic Transforms," *Electron. Lett.*, 1977, pp. 278–280.
13. Reed, I. S., and Truong, T. K., "Fast Mersenne-Prime Transforms for Digital Filtering," to be published in *Proc. IEE.*
14. Winograd, S., "On Computing the Discrete Fourier Transform," *Proc. Nat. Acad. Sci. USA*, 1976, Vol 73, pp. 1005–1006.
15. Winograd, S., *On Computing the Discrete Fourier Transform*, Research Report, Mat. Science Dept., IBM Thomas J. Watson Research Center, Yorktown Heights, New York, 10592.
16. Agarwal, R. C., and Cooly, J. W., "New Algorithm for Digital Convolution," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-25, pp. 392–410, Oct. 1977.
17. Reed, I. S., Kwoh, Y. S., Truong, T. K., and Hall, E. L., "X-Ray Reconstruction by Finite Field Transforms," *IEEE Transactions on Nuclear Science*, Vol. NS-24, No. 1, February 1977.
18. Niven, I., and Zuckerman, H. S., *An Introduction to the Theory of Numbers*, John Wiley and Sons, Inc., New York, 1966.
19. Cooley, J. W., and Tukey, J. W., "An Algorithm for the Machine Calculation of Complex Fourier Series," *Math. Comput.*, Vol. 19, pp. 297–301, April 1965.

Table 1. Complexity of hybrid DFT for real and complex data

$d = q'$	$q' - 1$	No. of integer multiplications for real data	No. of integer multiplications for complex data
367	$2 \cdot 3 \cdot 61$	488	976
373	$2^2 \cdot 3 \cdot 31$	620	1240
733	$2^2 \cdot 3 \cdot 61$	1220	2440
1831	$2 \cdot 3 \cdot 5 \cdot 61$	4880	9760
1861	$2^2 \cdot 3 \cdot 5 \cdot 31$	6200	12400
2441	$2^3 \cdot 5 \cdot 61$	8540	17080

Table 2. Complexity of new hybrid algorithm for DFT

d	Factors	New Algorithm No. of integer multiplications for complex data	Radix-2 FFT No. of real multiplications $2d \log_2 d$
4096	2^{12}		98,304
4476	373 X 4 X 3	14,880	
8192	2^{13}		212,992
8796	733 X 4 X 3	29,280	
16384	2^{14}		458,752
20888	373 X 8 X 7	89,280	
32768	2^{15}		983,040
41048	733 X 8 X 7	175,680	
62664	373 X 8 X 7 X 3	267,840	
65536	2^{16}		2,097,152
123144	733 X 8 X 7 X 3	527,040	
131072	2^{17}		4,456,448
262144	2^{18}		9,437,184
268560	373 X 16 X 9 X 5	1,740,960	
524288	2^{19}		19,922,944
527760	733 X 16 X 9 X 5	3,425,760	

VLBI/Laser Intercomparison Project: Session 2

H. F. Fliegel
Tracking Systems and Applications Section

Recent experiments are described which directly intercompare Very Long Baseline Interferometry (VLBI) with laser ranging to artificial satellites by measuring the vector lengths of the same intercontinental baselines. These experiments are part of the VLBI/Laser Intercomparison Project, a continuing task to assess the suitability of VLBI to various geophysical applications.

I. Introduction

Several systems are now under development by NASA and other agencies which should be capable of measuring baselines of several thousand kilometers with an accuracy of 5 cm, and baselines of several hundred kilometers to 3 or 4 cm:

Very Long Baseline Interferometry (VLBI)

Laser ranging to artificial satellites

Lunar laser ranging

Radio tracking of artificial satellites, which has already been implemented by the National Geodetic Survey (NGS) and the Defense Mapping Agency (DMA) with repeatability of 50 cm or better.

The great potential accuracy of the new systems presents both opportunity and challenge: the opportunity of measuring quantities important to civil engineering and navigation with great precision — land subsidence, movement along faults, regional deformation, and the rotation of the earth; but the challenge of making certain that the new systems really are delivering their rated accuracy, which exceeds all previous standards of comparison.

II. Project History

A VLBI/Laser Intercomparison Project has been established in order to compare VLBI techniques with laser ranging systems by conducting a series of measurements between common sites. A goal of the task is to assess the suitability of VLBI for geophysical applications. "Snapshot" experiments are being used to illustrate the key developments of the VLBI technique as it matures. The effort includes the co-location and intercomparison of VLBI with laser ranging systems.

The Intercomparison Project was established to ensure that VLBI development and the intercomparison with laser systems takes place in a timely and forthright manner. The Project also selects certain key tests to illustrate the development of VLBI. The validation experiments rely primarily on the hardware and software capability and related support associated with the DSN Advanced Systems (310 Program) and the DSN operational equipment development programs (311 and 312 Programs). No funding in the Intercomparison Project has been allocated for equipment development, equipment level analysis, installation or maintenance. This approach is analogous to flight project relationships to the TDA/DSN, where the necessary equipment capability is developed in a mission-

independent mode. The Intercomparison Project supports data acquisition costs involved in validation experiments that are extraordinary to normal DSN operations. It also bears the direct costs incurred by non-DSN facilities in support of validation activity. The Intercomparison Project acquires necessary consumable items, and supports the data processing of validation experiments.

III. Project Experiments

The Project is divided into several Sessions, or groups of experiments, each of which concludes with a report assessing the performance of VLBI at a given stage in its development. Session I consisted of a series of measurements between pairs of antennas in the Goldstone complex, and was intended primarily to give an accurate picture of the state of VLBI development at the beginning of the Project, and to demonstrate organizational preparedness to perform more ambitious experiments. Session II, which is now being completed, is a direct comparison of VLBI with satellite laser ranging over intercontinental baselines.

This is the first real test of the application of space technology to geodesy, in which two independent systems of equally high precision have been used to measure the distance and orientation of the same two points separated by several thousand kilometers. If the results agree, they will encourage reasonable confidence that both systems are working to their expected level of accuracy. Further, more rigorous tests of accuracy will then be justified.

The Session II experiments were designed to intercompare not only VLBI with laser but results from different kinds of VLBI hardware, called Mark I and Mark II. Using either Mark I or Mark II, the observing strategy is to observe each radio source simultaneously from all stations for several minutes. Mark I switches among five channels at X-Band frequencies with an instantaneous bandwidth of 360 kHz, recording on ordinary computer tape, 3 min per tape. Mark II switches among 3 channels at X-band frequencies with an instantaneous bandwidth of 2 MHz, on video tape, each tape containing about one hour of data. The Mark I and Mark II data were recorded simultaneously at the observing stations. It is not practical to take VLBI and laser data strictly simultaneously; but the experiments were designed to be completed over a few months, with the several VLBI experiments interlaced in time with the laser measurements, so that the possibility that earth crustal movement might create an unverifiable difference between the VLBI and laser determinations of the baselines might safely be ruled out.

Accordingly, three mobile laser satellite (MOBLAS) tracking units were deployed at the Haystack Observatory, the Owens Valley Observatory, and at Goldstone (DSS 14). Four months were required to maximize the possibility of three station simultaneous observations. During December, January, and February, JPL, GSFC, and the group at Haystack cooperatively scheduled MKI/MKII experiments at Goldstone, Haystack, and Owens Valley. MOBLAS systems occupied sites at Goldstone, Haystack, and Owens Valley from November 1, 1977, to March 31, 1978. JPL coordinated pad requirements at Owens Valley and provided additional equipment (BWS — Band Width Synthesis Units) to Haystack and Owens Valley in order to operate the MKII system. JPL/GSFC/Haystack technical personnel cooperated to define a mutually agreeable observing strategy. There were three VLBI observing sessions: December 13 through 15, 1977, and January 13 through 16 and February 24 through 26, 1978. JPL was responsible for providing water vapor radiometer support at both Goldstone and Owens Valley. The Haystack Observatory provided water vapor radiometer support at Haystack. The Haystack 37-m antenna was equipped for simultaneous S/X operation with system temperatures of 150 K at S-band and 70 K at X-band, hydrogen masers, phase calibrators, Mark I and Mark II recorders, and a water vapor radiometer. DSS 14 was operated in the following configuration: S- and X-band receivers, 40 MHz synthesized bandwidth, 20-30 K system temperature at zenith, 4 Mb/s record rate, no phase/cable calibration, and hydrogen (H) masers for the last two experiments. A water vapor radiometer was available at Goldstone for the last two experiments. The Owens Valley 39-m antenna was equipped with S- and X-band receivers, phase and cable calibration, H-maser, water vapor radiometer, and Mark I and Mark II recorders.

Of the three mobile laser satellite tracking units, MOBLAS vans 1 and 3 were already equipped with high-powered green lasers delivering a pulse 6 ns (180 cm) in length, and measuring range with a precision of approximately 25 cm. The MOBLAS 2 unit, which was deployed to Owens Valley, was equipped with a similar laser, so that all stations could simultaneously range to LAGEOS, a satellite in a high earth orbit. Other satellites which are routinely tracked by MOBLAS units include Beacon Explorer 3, GEOS-1, GEOS-3, Starlette, Timation, and NTS-2. Since the Satellite Laser Ranging System was committed to support SEASAT from May 1978 at least through 1980, and will not be available again at sites where its performance can readily be assessed, OTDA and OSTDS jointly agreed that even a limited intercomparison with the VLBI technique would be worthwhile.

For MOBLAS, the measurement objective was ±10 cm precision for the shorter baseline (Goldstone/Owens Valley) and ±50 cm on the longer baseline (Haystack/Goldstone), based on the assumption that at least 30 passes could be measured simultaneously from all sites during the several

months' occupation. A combination of bad weather and equipment failure may have degraded the laser satellite data from the goal.

A third system which has been deployed at all three sites is the satellite doppler positioning system developed by the Defense Mapping Agency (DMA) and now regularly used by the National Geodetic Survey (NGS). The NGS standard field procedure is to observe 40 passes of suitable satellites of the Navy Navigation Satellite System, also known as the TRANSIT system. The number of passes is larger than that expected to be observed by the MOBLAS system in the Intercomparison Project experiments, even though the NGS site occupation time is shorter, primarily because radio data can be acquired even in cloudy weather. Doppler positions are acquired by the point positioning system, in which the orbit of the satellite is taken from an ephemeris and not adjusted. The NGS uses the most precise satellite ephemerides available from the DMA, which are produced using the NWL 10E gravity model and the NWL 9D coordinate set for the tracking stations providing data from which the ephemerides are generated. Typical precision of a satellite doppler position is 0.5 m in latitude and height, and about 1 m in longitude. The NGS assumed responsibility for the local surveys, which tied together the VLBI antenna, the MOBLAS unit, and the satellite doppler geodetic control point at each of the three sites.

The purpose of Session II has been to provide an accurate "snapshot" of VLBI and MOBLAS performance. There were no preset requirements of accuracy to be met. A system is to be judged a failure in the Haystack/OVRO/Goldstone experiments in any of three, but only three, catastrophic eventualities:

(1) Failure to acquire data over the 4-month interval.

(2) Generation of inconsistent data.

(3) Generation of data in significant disagreement with conventional geodesy, that is, 10 or 20 m in error between Haystack and Goldstone, or a meter or so in error between Goldstone and OVRO.

Nevertheless, based on previous performance, we expect the VLBI data on the intercontinental lines to be of 20 – 30 cm quality.

The VLBI data from Session II are now being cross-correlated and seem satisfactory. A project review in which the VLBI, laser, and doppler satellite results will be presented and compared is scheduled at NASA Headquarters in July 1978.

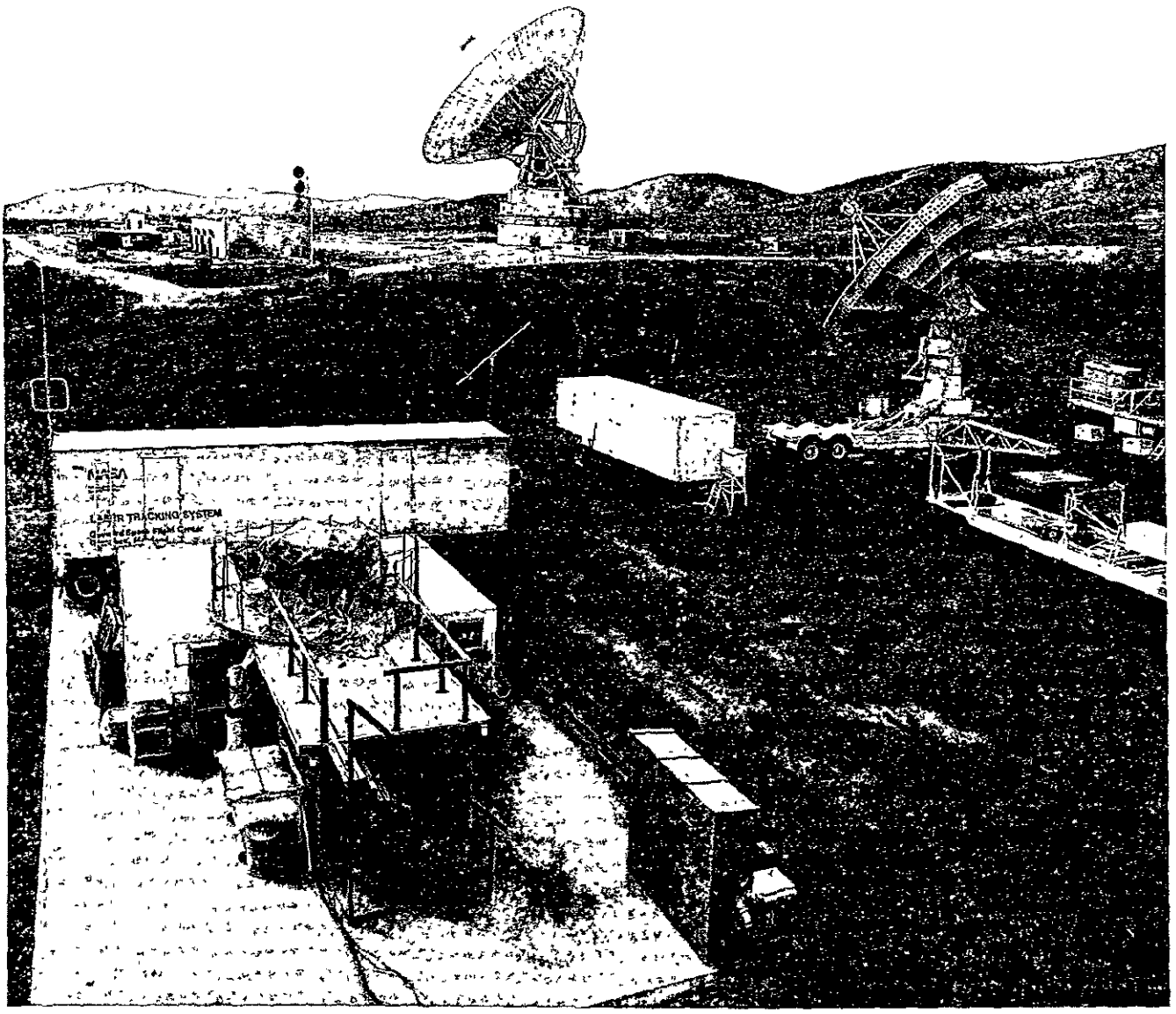

Fig. 1. A MOBLAS satellite laser ranging unit is brought to Goldstone as part of the VLBI/Laser Intercomparison Project. DSS 14 is in the background (top, center) and the ARIES 9-m transportable antenna on the right.

A New Algorithm for Computing Primitive Elements in *GF* (q^2)

I. S. Reed
University of Southern California

T. K. Truong and R. L. Miller
Tracking and Data Acquisition Engineering

A new method is developed to find primitive elements in the Galois field of q^2 elements $GF(q^2)$, where q is a Mersenne prime. Such primitive elements are needed to implement transforms over $GF(q^2)$.

I. Introduction

Several authors (Refs. 1 and 10) have proposed the use of the fast Fourier transform (FFT) over finite fields or rings. Such transforms can be used to compute circular convolutions of real sequences without round-off error. In Ref. 5, the authors extended the integer transforms of Rader by defining a complex number-theoretic transform over the Galois field $GF(q^2)$, where $q = 2^p - 1$ is a Mersenne prime for $p = 2, 3, 5, 7, 13, 17, 19, 31, 61, \cdots$. An algorithm was developed to compute elements of order 2^k for $1 \leqslant k \leqslant p + 1$. With this, an FFT algorithm of length 2^k was developed over $GF(q^2)$. Recently, the authors in Refs. 11 and 12 stated without proof the following result:

Let $GF(q^2)$ be a Galois field, where q is a Mersenne prime. If $d|(q^2 - 1)$, where $d = 2^k \cdot m$, m odd, $3 \leqslant k \leqslant p + 1$, and where $p|m$ for $p \leqslant m \leqslant 2^{p-1} - 1$, then there exists a generator α of the multiplicative subgroup G_d of order d, such that α satisfies

$$\alpha^{d/8p} \equiv (1 + \hat{i}) \bmod q$$

$$\alpha^{d/8} \equiv \text{one of forms } \pm 2^{(p-1)/2} (1 \pm \hat{i}) \bmod q$$

$$\alpha^{d/4} \equiv \hat{i} \text{ or } -\hat{i} \bmod q$$

and

$$\alpha^{d/2} \equiv -1 \bmod q \tag{1}$$

Using the above properties in $GF(q^2)$, a mixed high-radix transform of $2^k \cdot p$ points over $GF(q^2)$, where $3 \leqslant k \leqslant p+1$, can be developed. Such an algorithm for $GF(q^2)$ appears comparable in speed to that given by Winograd (Ref. 13).

In this article, a new method is presented for finding the primitive elements of $GF(q^2)$. Also a technique for finding an element of order d that satisfies Eq. (1) is given.

II. An Algorithm for Finding Primitive Elements in GF (q^2) where q is a Mersenne Prime

If q is a Mersenne prime, the order t of the multiplicative group with generator α of $GF(q^2)$ factors as follows:

$$t = (2^p - 1)^2 - 1 = 2^{p+1}\,(2^{p-1} - 1)$$

To find a primitive element of $GF(q^2)$, the following theorem is important.

Theorem 1: If q is a prime number, $t = q^n - 1$ and $t = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$, where the p_i are distinct primes, then γ is a primitive element of $GF(q^n)$, if and only if γ satisfies the congruences

$$\gamma^{t/p_1} \not\equiv 1 \bmod q \qquad (2)$$

$$\gamma^{t/p_2} \not\equiv 1 \bmod q$$

$$\vdots$$

$$\gamma^{t/p_k} \not\equiv 1 \bmod q$$

Proof: If γ is a primitive element of $GF(q^2)$, then $t = q^n - 1$ is the smallest positive number such that $\gamma^t \equiv 1 \bmod q$. Thus γ satisfies the congruences of Eq. (2).

Now assume γ satisfies congruences Eq. (2) and that $t = q^n - 1$ is not the smallest positive number for which this is true. Then there exists an integer ℓ for which $1 < \ell < t$ and $\ell \mid t$ such that $\gamma^\ell \equiv 1 \bmod q$. But when we have

$$\frac{t}{\ell} = p_i u$$

for some u and prime p_i in the factorization of t. Thus $t/p_i = \ell u$. Hence $\gamma^{t/p_i} \equiv 1 \bmod q$. This is contrary to assumption.

If q is a Mersenne prime, then $t = q^2 - 1 = 2^{p+1}(2^{p-1} - 1) = 2^{p+1} \cdot p_2^{e_2} \cdot p_3^{e_3} \cdots p_k^{e_k}$, where $p_i \neq p_j$ for $i \neq j$. Assume $\gamma = a + \hat{i}b$ is a primitive element of $GF(q^2)$. Then, by Theorem 1, γ satisfies

$$\gamma^{\frac{q^2-1}{p_1}} \equiv \gamma^{(2^{p+1}(2^{p-1}-1))/2} \equiv \gamma^{2^p(2^{p-1}-1)} \not\equiv 1 \bmod q$$

$$\gamma^{\frac{q^2-1}{p_2}} \equiv \gamma^{2^{p+1} \cdot p_2^{e_2-1} p_3^{e_3} \cdots p_k^{e_k}} \not\equiv 1 \bmod q$$

$$\vdots$$

$$\gamma^{\frac{q^2-1}{p_k}} \equiv \gamma^{2^{p+1} p_2^{e_2} \cdot p_3^{e_3} \cdots p_k^{e_k - 1}} \not\equiv 1 \bmod q \qquad (3)$$

Since $q = 2^p - 1$, it is easily seen that

$$\gamma^{2^p} \equiv (a + \hat{i}b)^{2^p} \equiv (a + \hat{i}b)(a + \hat{i}b)^q$$

$$\equiv (a + \hat{i}b) \sum_{k=0}^{q} \binom{q}{k} a^{q-k} (\hat{i}b)^k \equiv (a + \hat{i}b)(a^q + (\hat{i}b)^q) \bmod q$$

By Fermat's theorem,

$$(a + \hat{i}b)\,(a^q + (\hat{i}b)^q) \equiv (a + ib)\,(a + \hat{i}^q b) \bmod q$$

But since $q \equiv 3 \bmod 4$,

$$\gamma^{2^p} \equiv (a + \hat{i}b)\,(a - \hat{i}b) \equiv a^2 + b^2 \bmod q$$

Thus Eq. (3) becomes

$$\gamma^{\frac{q^2-1}{p_1}} \equiv (a^2 + b^2)^{2^{p-1}-1} \equiv (a^2 + b^2)^{\frac{q-1}{2}} \not\equiv 1 \bmod q$$

$$\gamma^{\frac{q^2-1}{p_2}} \equiv (a^2 + b^2)^{2 \cdot p_2^{e_2-1} \cdot p_3^{e_3} \cdots p_k^{e_k}} \equiv (a^2 + b^2)^{\frac{q-1}{p_2}}$$

$$\not\equiv 1 \bmod q$$

$$\vdots$$

$$\gamma^{\frac{q^2-1}{p_k}} \equiv (a^2 + b^2)^{2 \cdot p_2^{e_2} \cdot p_3^{e_3} \cdots p_k^{e_k-1}} \equiv (a^2 + b^2)^{\frac{q-1}{p_k}}$$

$$\not\equiv 1 \bmod q \tag{4}$$

By Theorem 1, the element $a^2 + b^2$ that satisfies Eq. (4) must be a primitive element of $GF(q)$. Thus if one can find a primitive element c in $GF(q)$ such that $c^{(q-1)/p_i} \equiv 1$ for $i = 1, 2, \cdots, k$, then the problem of finding a primitive element of $GF(q^2)$ is reduced to the problem of solving the congruence

$$a^2 + b^2 \equiv c \bmod q \tag{5}$$

for a and b.

To solve Eq. (5) let $X \equiv a^2 \bmod q$ and let $Y \equiv -b^2 \bmod q$. Then Eq. (5) becomes

$$X - c \equiv Y \bmod q \tag{6}$$

Proposition: Let $c \in GF(q)$, then there exist $a, b \in GF(q)$ such that $c = a^2 + b^2$.

Proof: If $q = 2^m$, then since $c^{2^m} = c$, it follows that $c = (c^{2^{m-1}})^2 + 0^2$. If q is odd, then

$$GF(q) = \{0, a_1, \cdots, a_{(q-1)/2}\}$$

$$\cup \{-a_1, -a_2, \cdots, -a_{(q-1)/2}\}$$

It is readily seen that there are exactly $(q + 1)/2$ elements of the form a^2 in $GF(q)$, and that there are $(q + 1)/2$ elements of the form $c - b^2$ for a fixed $c \in GF(q)$. Thus

$$\{a^2 \mid a \in GF(q)\} \cap \{c - b^2 \mid b \in GF(q)\} \neq \emptyset$$

Consequently we can express $c = a^2 + b^2$.

By the Proposition, one can choose the numbers X and Y from the set of integers $1, 2, \cdots, 2^p - 2$, such that X is a square and $Y = X - c$ is a nonsquare. Thus it is sufficient to let

$$a^2 \equiv X \bmod q$$

$$b^2 \equiv -X + c \equiv -Y \bmod q \tag{7}$$

By a procedure precisely similar to that used to find the solutions of congruence Eq. (14) in Ref. 5, the solutions of congruence (7) are given by

$$a \equiv \pm X^{2^{p-2}} \bmod 2^p - 1$$

$$b \equiv \pm(-Y)^{2^{p-2}} \bmod 2^p - 1 \tag{8}$$

The following theorem is often useful for finding solution of Eq. (5).

Theorem 2: If 3 is a primitive element of $GF(q)$, then the solution of the congruence

$$a^2 + b^2 \equiv 3 \bmod q$$

for a and b are given by

$$a = (2^{\frac{p-1}{2}} + 1)$$

and

$$b = (2^{\frac{p-1}{2}} - 1)$$

Proof: Note that $3 \equiv 2^p - 1 + 3 \bmod q$. Thus

$$3 \equiv 2^p + 2 \equiv (2^{\frac{p-1}{2}} + 1)^2 + (2^{\frac{p-1}{2}} - 1)^2$$

$$\equiv a^2 + b^2 \bmod q$$

and

$$a = 2^{\frac{p-1}{2}} + 1$$

$$b = 2^{\frac{p-1}{2}} - 1$$

are identically the solutions of the congruence.

To find a solution of $x^c \equiv 1 + \hat{i} \bmod q$ where $c = 2^k \cdot m/8p$, assume $\gamma = a + \hat{i}b$ is a primitive element of $GF(q^2)$. Then, using a computer program, one can find an integer j such that $\gamma^j \equiv 1 + \hat{i} \bmod q$. Hence

$$(\gamma^j)^{(q^2-1)/c} \equiv (1 + \hat{i})^{(q^2-1)/c}$$

$$\equiv ((1 + \hat{i})^{8p})^{(q^2-1)/2^k \cdot m} \bmod q \qquad (9)$$

where $2^k \cdot m \mid q^2 - 1$.

It was shown in Ref. 12 that the element $(1 + \hat{i})$ is an element of order $8p$ in $GF(q^2)$. Thus Eq. (9) becomes

$$(\gamma^j)^{(q^2-1)/c} \equiv 1 \bmod q$$

Since γ is primitive, this implies $c \mid j$. Thus

$$(\gamma^{j/c})^c \equiv (1 + \hat{i}) \bmod q$$

where $c = 2^k \cdot m/8p$ for $3 \leqslant k \leqslant p - 1$, where $p \nmid m$ for $p \leqslant m \leqslant 2^{p-1} - 1$. Hence, $\alpha = \gamma^{j/c}$ is an element of order $d = 2^k \cdot m$ in $GF(q^2)$ such that α satisfies (1).

Consider a simple example:

Example 1: Let $q = 2^p - 1 = 2^3 - 1$. Then $q^2 - 1 = 2^{p+1}(2^{p-1} - 1) = 2^4 \cdot 3$. Find an element $\alpha = a + \hat{i}b$ of $GF(7^2)$ such that satisfies

$$\alpha^{(q^2-1)/8p} \equiv \alpha^{48/24} \equiv \alpha^2 \equiv (1 + \hat{i}) \bmod q$$

$$\alpha^{48/8} \equiv \alpha^6 \equiv 2(-1 + \hat{i}) \bmod q$$

$$\alpha^{48/4} \equiv \alpha^{12} \equiv -\hat{i} \bmod q$$

$$\alpha^{48/2} \equiv \alpha^{24} \equiv -1 \bmod q$$

First we note that 3 satisfies the congruences:

$$3^{\frac{7-1}{2}} \equiv -1 \not\equiv \bmod 7 \qquad (10)$$

and

$$3^{\frac{7-1}{3}} \equiv 3^2 \equiv 2 \not\equiv 1 \bmod 7$$

Thus by Theorem 1, 3 is a primitive element of $GF(7)$. Using Eq. (10) in Eq. (4) and Eq. (5) the problem of finding a primitive element in $GF(q^2)$ is reduced to the solution of the congruence,

$$a^2 + b^2 \equiv 3 \bmod 7 \qquad (11)$$

By Eq. (6) the solution of Eq. (11) is equivalent to the solution of

$$X - 3 \equiv Y \bmod 7$$

where $X \equiv a^2 \bmod 7$ and $Y \equiv -b^2 \bmod 7$. It can be shown that $Y = 6$ is nonsquare and $X = -Y + 3 = 9$ is a square. Thus, by Eq. (7),

$$a^2 \equiv 9 \equiv 2 \bmod 7$$

$$b^2 \equiv -6 \equiv 1 \bmod 7 \qquad (12)$$

Hence by Eq. (8), the solution of congruence (12) are given by

$$a \equiv -2^{2^{p-2}} \equiv -2^2 \equiv 3 \bmod 7$$

$$b \equiv (1)^{2^{p-2}} \equiv (1)^3 \equiv 1 \bmod 7$$

Note that since $c = 3$ in Eq. (11), the solution of the congruence Eq. (11) can be also obtained by using Theorem 2. It is evident that Theorem 2 yields the same solution, as follows:

$$a \equiv (2^{\frac{p-1}{2}} + 1) = 2^{\frac{3-1}{2}} + 1 = 3$$

$$b \equiv 2^{\frac{p-1}{2}} - 1 = 2^{\frac{3-1}{2}} - 1 = 1$$

Hence $\gamma = -4 + \hat{i}$ is a primitive element of $GF(7^2)$. Since $(1 + \hat{i})^{48/2} \equiv (1 + \hat{i})^{24} \equiv 1 \bmod q$, then $x^2 \equiv 1 + \hat{i}$ has a solution. To find such a solution it is necessary to find an integer j such that $\gamma^j \equiv 1 + \hat{i} \bmod q$. In this case, $j = 14$, i.e., $\gamma^{14} \equiv 1 + \hat{i} \bmod q$. Thus

$$(\gamma^{14/2})^2 \equiv 1 + \hat{i} \bmod q$$

Hence $\gamma^{14} \equiv 1 + \hat{i} \bmod q$ and the desired element is $\alpha = \gamma^7$ of order 48 in $GF(7^2)$. Evidently $\alpha = \gamma^7$ satisfies also

$$\alpha^6 \equiv (1 + \hat{i})^3 \equiv 2(-1 + \hat{i}) \bmod q$$

$$\alpha^{12} \equiv -\hat{i} \bmod q$$

and

$$\alpha^{24} \equiv -1 \bmod q$$

With the same procedure used in Example 1 and then using a computer program, the primitive elements for a number of different Mersenne primes were found. These are shown in Table 1.

Acknowledgment

The authors wish to thank Dr. N. A. Renzetti, Manager of Tracking and Data Acquisition Engineering, and the members of the Advanced Engineering Group in that organization at the Jet Propulsion Laboratory for their early support, suggestions and encouragement of the research which led to this paper. We also thank Dr. C. A. Greenhall for his computer programming assistance.

References

1. Pollard, J. M., "The Fast Fourier Transform in a Finite Field," *Math. Comput.*, 1971, 25, pp. 365–374.
2. Schonhage, A., and Strassen, V., "Schnelle Multiplikation Grosser Zahlen," *Computing*, 1971, 7, pp. 281–292.
3. Rader, C. M., "Discrete Convolution via Mersenne Transforms," *IEEE Trans.*, 1972, C-21, pp. 1269–1273.
4. Agarwal, R. C., and Burrus, C. S., "Number Theoretic Transforms to Implement Fast Digital Convolution," *Proc. IEEE*, 1975, 63, pp. 550–560.
5. Reed, I. S., and Truong, T. K., "The Use of Finite Fields to Compute Convolutions," *IEEE Trans.*, 1975, IT-21, pp. 208–213.
6. Reed, I. S., and Truong, T. K., "Complex Integer Convolution Over a Direct Sum of Galois Fields," *ibid*, 1975, IT-21, pp. 657–661.
7. Vegh, E., and Leibowitz, L. M., "Fast Complex Convolution in Finite Rings," *ibid*, 1976, ASSP-24, pp. 343–344.
8. Golomb, S. W., Reed, I. S., and Truong, T. K., 'Integer Convolutions Over the Finite Field GF($3 \cdot 2^n + 1$)," *SIAM J. Appl. Math.*, Vol. 32, No. 2, March 1977.
9. Pollard, J. M., "Implementation of Number-Theoretic Transforms," *Electron. Lett.*, 1976, 12, pp. 378–379.
10. Liu, K. Y., Reed, I. S., and Truong, T. K., "Fast Number-Theoretic Transforms for Digital Filtering," *ibid*, 1976, 12, pp. 644–646.
11. Reed, I. S., Truong, T. K., and Liu, K. Y., "Fast Algorithm for Computing Complex Number-Theoretic Transforms," *Electronics Letters*, 12th May 1977, 13, No. 10, pp. 278-280.
12. Reed, I. S., and Truong, T. K., "Addendum to Fast Algorithm for Computing Complex Number-Theoretic Transforms," to be published in *Electronics Letters*.
13. Winograd, S., "On Computing the Discrete Fourier Transform," *Proc. Nat. Acad. Sci. USA*, 1976, 73, pp. 1005–1006.
14. Vinogrador, I. M., *Elements of Number Theory*, Dover Publications, New York, 1954.

Table 1. Primitive elements of $GF(q^2)$ for different q

	$q = 2^p - 1$	Primitive element of $GF(q)$	$a^2 + b^2 \bmod q$	Primitive element of $GF(q^2)$
3	7	3	$3^2 + 1^2$	$3 + \hat{i}$
5	31	3	$5^2 + 3^2$	$5 + \hat{i}3$
7	127	3	$9^2 + 7^2$	$9 + i\,7$
13	32767	17	$4^2 + 1^2$	$4 + \hat{i}$
17	131071	3	$257^2 + 255^2$	$257 + \hat{i}255$
19	524287	3	$513^2 + 511^2$	$513 + \hat{i}511$
31	2147483647	53	$7^2 + 2^2$	$7 + \hat{i}2$

Examination of the DSN X-Band Weather Specifications

C. A. Greenhall
TDA Engineering

Effects of weather on DSN system performance at X-band are examined by comparing a record of 64-meter system noise temperatures with weather observations taken at approximately the same times and places.

I. Introduction

As part of the specification of itself to the user, the Deep Space Network provides a probabilistic description of the weather-induced increases in signal attenuation and system noise temperature (SNT) at X-band. The description takes the form of a set of tables of X-band zenith cloud attenuation vs cumulative probability; the tables reside in Module TCI-40 (DSN Telecommunications Interfaces, Atmospheric and Environmental Effects) of Document 810-5; Rev. D (Ref. 1). In this article, they will be called the "810-5 tables."

The Voyager Project, whose planetary encounters could be affected seriously by weather, requested the DSN to reexamine the 810-5 tables in the light of two recent sources of information. The first source is a record of X-band zenith system noise temperatures measured from March 1976 to March 1977 at the three 64-meter antenna stations during precalibration and postcalibration of Viking tracks. We shall use the name "76-77 SNT data" or simply "SNT data" to refer to this data set. The second source is the result of a contract executed by Meteorology Research, Inc. (MRI), the same firm that provided the data for the 810-5 tables. MRI was given only the measurement times of the 76-77 SNT data, not the actual SNT values. They returned a data set containing their estimates of weather and X-band zenith cloud attenuation at those times. We shall use the name "76-77 weather data" or simply "weather data" to refer to this data set. Since cloud attenuation can be converted to added noise temperature, one can compare the 76-77 weather data to fluctuations in the 76-77 SNT data. The idea is to increase confidence in the 810-5 tables by checking the 76-77 SNT and weather data against each other. This assumes that the 810-5 tables arose from the same kind of weather data collection and processing as the 76-77 weather data. Unfortunately, this proviso could never be satisfied fully because the 810-5 tables are based largely on climatic summaries, not daily weather observations.

We shall present a comparison of the 810-5 tables, the 76-77 SNT data, and the 76-77 weather data. It will be seen that when the SNT and weather data are compared point by point, the correlation is low. The cumulative distributions agree better. On the other hand, the 810-5 tables look quite different from either the 76-77 SNT or the weather data.

II. Description of Data Sets

A. 810-5 Tables

The present edition (the second) of these tables is based on two reports by MRI. The first report, published in 1970 (Ref. 2), contains a set of cumulative distribution functions (cdf) of X-band (and S-band) zenith cloud attenuation. There is a distribution for each choice of DSN site (Goldstone, Canberra, Madrid), season (December-February, March-May,

June-August, September-November), and day quarter (centered at 0000, 0600, 1200, and 1800 local time), except that no 0000 distributions for Canberra were given.

The Goldstone distributions are based on a year of cloud observations at Daggett, California; a summary of the data was compared to long-term averages at Edwards Air Force Base to make sure that the single year of record at Daggett was not abnormal. For Canberra and Madrid, it was necessary to rely on long-term statistics of cloud cover by seasons because daily cloud data were not available.

The distributions account for nonprecipitating clouds only. Unfortunately, the random variable whose distributions MRI provided is not cloud attenuation, but maximum cloud attenuation times the fraction of the sky covered. This quantity can roughly be considered as the time average of attenuation over a period long enough for many clouds to pass by the antenna.

MRI estimated that the attenuation values (given in dB) "should be correct within a factor of 10–20 on a statistical basis." Needless to say, these huge error estimates destroyed the credibility of the results. In 1976, MRI published a second report (Ref. 3), which outlined a procedure for modifying the 1970 distributions and put somewhat narrower multiplicative error bounds on the results. The modification concerns the possibility that some of the clouds that previously were treated as ice might actually have been liquid water clouds, which have much greater attenuation. This modification tends to push the distribution toward higher attenuations.

The JPL Telecommunications Division carried out the procedure of the 1976 report; the resulting distributions have been used by the Voyager Project for planning their encounter sequences. Independently, the author executed his own interpretation of the procedure. The multiplicative error bounds were folded into the distributions in such a way that they do not appear in the final result, which is a set of distributions with no error bounds. These distributions, which constitute the 810-5 tables, are essentially a smoothed version of those produced by the Telecommunications Division.

We remark here that the 0000 Canberra distributions simply duplicate those of 0600; we shall not attempt to defend this arbitrary choice.

B. 76-77 SNT Data Set

This record of 64-meter zenith X-band system noise temperature measurements is shown in Fig. 1. There are 273 data for DSS 14 (Goldstone), 324 data for DSS 43 (Canberra), and 284 data for DSS 63 (Madrid). System noise temperatures were measured by the Y-factor method. Each measurement is accompanied by the precalibration or postcalibration time tag (GMT). The actual time of the measurement was estimated by

$$\text{measurement time} = \text{time tag} - 1 \text{ hour for precal}$$
$$= \text{time tag} + \tfrac{1}{2} \text{ hour for postcal}$$

These time estimates were converted to local standard time and given to MRI. Accompanying the data are occasional notes about equipment problems or bad weather. Those points are marked on Fig. 1. The *absence* of a note, however, does not imply the absence of problems with equipment or weather.

The portion of DSS 14 data between 76-162 (year-day of year) and 76-240 has a mean 4.8 Kelvin above the rest of the record. This must have been due to equipment problems, not weather. To rescue these data, they were lowered by 4.8 K during processing. (Two data then fell below 21 K; these were deleted.)

On 76-316, a new X-band traveling-wave maser was installed at DSS 43. (The note "maser orange" in Fig. 1(b) refers to the new maser, which was being burned in.) The SNT quieted down immediately. Accordingly, none of the data before 76-317 was used. This is unfortunate because there are many genuine bad-weather points in this period. The deletion left 107 data for processing. Nevertheless, all 324 measurement times were given to MRI.

C. 76-77 Weather Data Set

Items from this data set that were used for the present study are:

(1) Time tag; MRI gathered weather data for the nearest hour to the measurement time furnished them from the 76-77 SNT data set.

(2) Fraction of sky covered by clouds.

(3) Fraction of sky covered by optically opaque clouds.

(4) Physical temperature of low clouds, if any, else the temperature of midlevel clouds, if any.

(5) Estimated maximum zenith X-band cloud attenuation; the actual zenith attenuation would vary from zero to the maximum as clouds passed through the antenna beam. Clear-air (oxygen and water vapor) attenuation is not included; if there were no clouds or only high clouds, then zero attenuation was returned.

The author converted attenuation to added noise temperature by the formula

$$T = T_p \left(1 - 10^{-A/10}\right) \tag{1}$$

where T = added noise temperature (Kelvin), T_p = physical cloud temperature (Kelvin), and A = cloud attenuation (dB).

D. SNT_0 Data Set

This is a subset of the 76-77 SNT data set consisting of the SNT measurements at the times for which MRI reported zero cloud attenuation. The SNT_0 data show the weather-independent scatter in SNT measurements, provided that one may neglect variations in clear-air attenuation and rely on the indications of clear weather in the 76-77 weather data.

III. Comparison of the 76-77 SNT and Weather Data Sets

A. Point by Point

Figure 2 shows scatter plots of the 76-77 SNT data vs the corresponding added noise temperatures from the 76-77 weather data for the three 64-meter stations. We have plotted only the points for which MRI reported a positive cloud attenuation; on the right side of each plot we show the histogram of the remaining points, namely the SNT_0 data. (If those points had been plotted, they would be superimposed along the left edges of the plots.) Off-scale points are shown at the tops of the plots.

In an ideal situation, the points would cluster about a line of slope 1, that is, each increase in cloud noise would be matched by an increase in system noise temperature. For those points below the "regression cutoff" lines, a least-squares straight line was fitted and a correlation coefficient computed. The correlation is 0.37 for Goldstone, 0.35 for Canberra, and 0.17 for Madrid. Clearly, the regression lines cannot be taken seriously.

We see two causes for the weakness of the coupling between the two data sets. First, the SNT_0 data show that much of the scatter in the SNT data is caused by nonweather effects. Second, the SNT and weather data sets are separated, both in time and in space. The measurement times of corresponding data in the two sets could differ by as much as one hour, for the SNT time is only an estimate, and the weather time is given only to the nearest hour. We estimate the standard deviation of the time difference to be about 25 minutes. As for the spatial separation, the weather data for Goldstone are based on observations at Daggett, about 60 km from DSS-14. We do not know the situations at Canberra and Madrid.

If cloud attenuation varied slowly over time and space, these separations might not drastically decouple the data sets. Unfortunately, sky noise records from a dedicated X-band radiometer at Goldstone exhibit sharp, narrow peaks separated by quiet periods. An example can be seen in Fig. 1 of Ref. 4. Two of the peaks have widths less than two hours.

To sum up, an item from the 76-77 weather data is an estimate of the added noise when looking through the thickest part of a typical cloud found in the general vicinity of the station at about the same time as the SNT observation. One could not expect a high correlation between these data sets.

B. Distribution Function Comparison

Having observed that the 76-77 SNT and weather data sets are nearly orthogonal, we turn to the examination of their one-dimensional cumulative distribution functions (cdf). A first cut at a comparison is given in Fig. 3, which was displayed at a Voyager Telecommunication Review. The distributions of added noise from the 76-77 weather data start at zero Kelvin, whereas the SNT distributions start at above 20 Kelvin. Although a baseline system noise temperature of 25 ± 3 Kelvin is specified in Ref. 1, we do not want to assume an arbitrary baseline, preferring rather to let the data speak for themselves in this regard. Hence, we shifted the cdf's so that they coincide at their medians, the 50-percent points.

The 76-77 weather cdf curve is actually the cdf of added noise temperature times the fraction of the sky covered by clouds; this is the crude cloud cover adjustment used in Ref. 2.

Because the SNT data contain nonweather fluctuations, comparing them directly with the weather data is perhaps ill-advised. Nevertheless, we can make a few observations if we confine ourselves to the higher probability levels, above the bulk of the nonweather fluctuations of the SNT. For Goldstone, the agreement is good above the 90% level; for Canberra, the agreement is good through the 90% level, but at the 95% level one can see the tail of the SNT cdf leveling off. This reflects the thin scattering of high system noise temperatures, 40 K and above, visible in Figs. 1(b) and 2(b). Such temperatures (more than 15 K above baseline) are not seen in the weather data, but we remind the reader that the weather data do not include precipitation. For Madrid, the weather data are clearly pessimistic. Again, above 95%, the SNT data have a few extremely high system temperatures.

The second comparison method folds the SNT_0 distribution into the 76-77 weather distribution before comparison with the SNT distribution. Let X and Y be independent random variables having the distributions of the SNT_0 and weather data. Then X represents the zenith SNT measurements during clear weather, and Y the weather-induced increases in SNT. The distribution of $X + Y$ is the convolution of the SNT_0 and 76-77 weather distributions; it is reasonable to compare it with the distribution of the total 76-77 SNT data set.

A different method of accounting for cloud cover is used here. Suppose that the maximum cloud brightness temperature is 4.2 K and the opaque cloud cover fraction is 0.7. We shall crudely model the situation by supposing that 7 out of 10 narrow-beam zenith observations around this time would see a 4.2-K cloud temperature, and that 3 out of 10 would see only clear air or a thin cloud section. Thus, when binsorting this weather datum, we add 0.7 to the bin for 4.2 K, and 0.3 to the bin for zero Kelvin. The sum of the bin counts is still the total number of data, as with an ordinary binsort.

Figure 4 shows the cdf's for 76-77 SNT and SNT_0 + 76-77 weather. The third curve is the cdf of SNT_0 + 810-5; we shall treat this in Section IV. For Goldstone, the good agreement is to be expected because the nonweather variations dominate the weather-induced ones in this desert climate. At Canberra and Madrid, it can be seen that the SNT_0 + weather distributions are more pessimistic than the SNT distributions up to the 95% level for Canberra and the 97% level for Madrid. Beyond these levels, the tails of the SNT distributions are extended by the few system temperatures greater than 40 K.

IV. Comparison of 810-5 With the 76-77 SNT and Weather Data

Recall that the purpose of this study is to validate (or invalidate, as the case may be) the DSN cloud attenuation distributions in Document 810-5. These represent long-term average probability levels. To do this, we have compared the SNT and weather data sets of 1976-1977 to each other in order to make conclusions about MRI-derived statistics in general. On the other hand, although the 1976-1977 weather is not expected to resemble closely the overall long-term climate, it is inevitable that one should compare the 810-5 tables with the two recent data sets, especially since the 810-5 distributions for Canberra and Madrid are based largely on climatic summaries, while the recent 76-77 weather data come from daily observations.

Figure 5 compares the 810-5 tables with the cdf of the 76-77 weather data set. Since the SNT data play no role, the entire set of 324 weather data for Canberra was used. The second method of Section III-B was used to adjust the weather data for cloud cover. The measurement times of the weather data, being dependent on the vagaries of Viking passes, are far from uniformly distributed over the times of day and seasons of the year. As noted in Section II-A, for each DSN site there are 16 distributions in the 810-5 tables, one for each season and day quarter. To produce the 810-5 distribution of Fig. 5 for a given DSN site, the number of 76-77 SNT times that fall into each of the 16 categories was counted and an appropriate weighted average of the 16 distributions computed. To convert attenuation to added noise temperature, Eq. 1 with T_p = 260 K was used. (The 76-77 weather data set shows that typical cloud temperatures fall in the range 270–280 K, but 260 K was used because Ref. 1 specifies it.)

In order to show what total system temperature measurements would look like if weather-induced noise followed the 810-5 distributions, we show on each section of Fig. 4 a cdf obtained by convolving the SNT_0 distribution with a weighted average of 810-5 added noise distributions. (For Canberra, only the 107 times of the "good" SNT data were used to determine the weights.)

One can see in both Fig. 4 and Fig. 5 that the upper tails of the 810-5 distributions are much heavier than those of the recent MRI weather observations. They are also heavier than the tails of the SNT distributions, especially for Madrid, in spite of the inclusion of rain in the SNT data and its exclusion from the 810-5 distributions.

V. Conclusions

From the comparison between the distributions of 76-77 SNT and of SNT_0 + 76-77 weather in Fig. 4, we conclude that on the average the added noise temperatures given by MRI are slightly higher than the SNT data. For Goldstone, the difference is obscured by the scatter in the SNT and SNT_0 data. For Canberra and Madrid, the 76-77 weather cdf is 2.5 K and 3.5 K higher at the 90% level than the SNT cdf. Beyond 95%, though, the SNT distribution reflects the presence of a few system noise temperatures beyond 40 K, these cannot occur in the weather data because the latter do not account for precipitation. It appears that the method used by MRI to convert a weather and cloud observation to X-band attenuation is reasonably accurate.

If we had not looked directly at the 810-5 distributions, we would be tempted to conclude from the foregoing that they provide a slightly pessimistic estimate of the effect of weather on DSN communication. The striking differences between the 810-5 distributions and the recent MRI weather data raise doubts, however. How much of these differences is caused by possible atypical weather during March 1976 to March 1977 at the DSN sites, and how much is caused by poor estimates of cloud attenuation statistics from climatic summaries instead of from daily and hourly records? It is a general consensus that the upper tails of the 810-5 distributions are far too heavy, hence no matter how well the 76-77 SNT and weather data agreed, they could not validate 810-5. It is therefore strongly recommended that the DSN collect a long-term database of system noise temperatures together with simultaneous

measurements of weather parameters and indicators of microwave configuration. A system for doing this is being implemented; it includes a noise-adding radiometer for accurate measurement of system temperature, the Meteorological Monitor Assembly for collecting ground weather parameters, and, later, a water-vapor radiometer for estimating the depth of liquid water in the antenna beam. As such a database accumulates over the years, statistics gleaned from it will be built into a more accurate set of specifications for the effect of weather on DSN communication.

References

1. *Deep Space Network/Flight Project Interface Design Handbook,* Document 810-5; Rev. D, Jet Propulsion Laboratory, Pasadena, California.
2. Smith, T. B., and Chien, C. W., *Cloud Cover Effects on Signal Attenuation at DSN Sites,* Meteorology Research, Inc., Altadena, California, April 24, 1970.
3. Carbone, R. E., and Smith, T. B., *Attenuation Due to Clouds at X-Band,* Meteorology Research, Inc., Altadena, California, February 16, 1976.
4. Greenhall, C. A., "Simulation of Time Series by Distorted Gaussian Processes," *Deep Space Network Progress Report* 42-37, pp. 146–151, Jet Propulsion Laboratory, Pasadena, California, February 15, 1977.

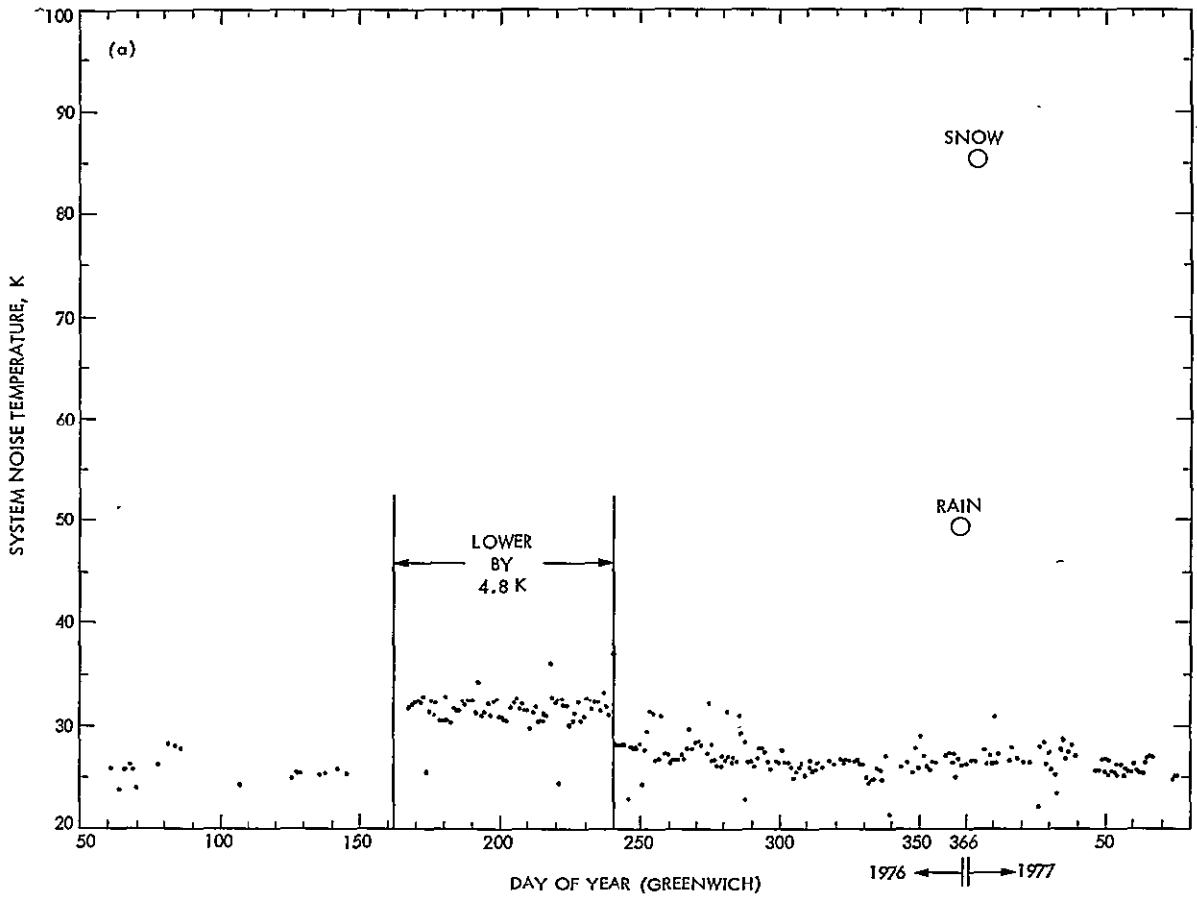

Fig. 1. Record of 76-77 SNT data: (a) Goldstone (DSS 14); (b) Canberra (DSS 43); (c) Madrid (DSS 63)

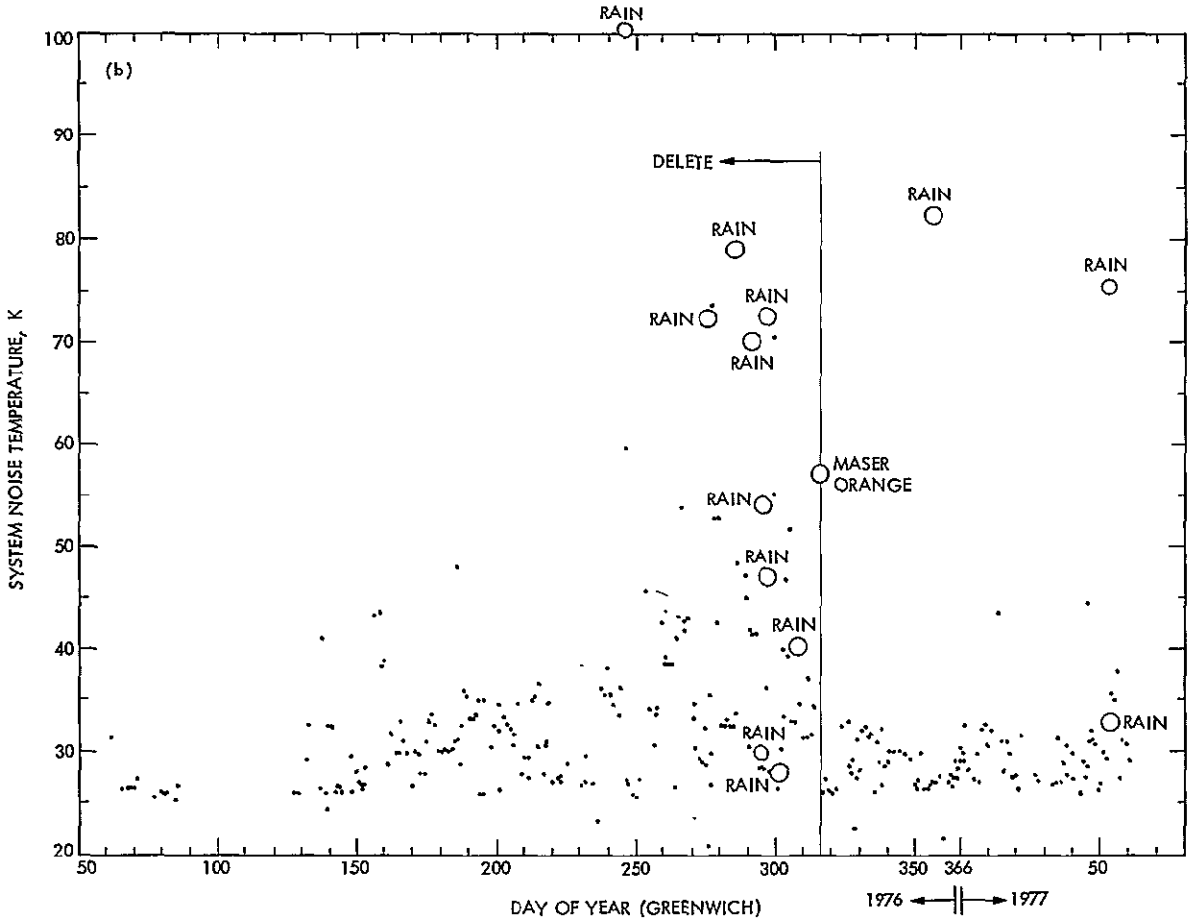

Fig. 1 (contd)

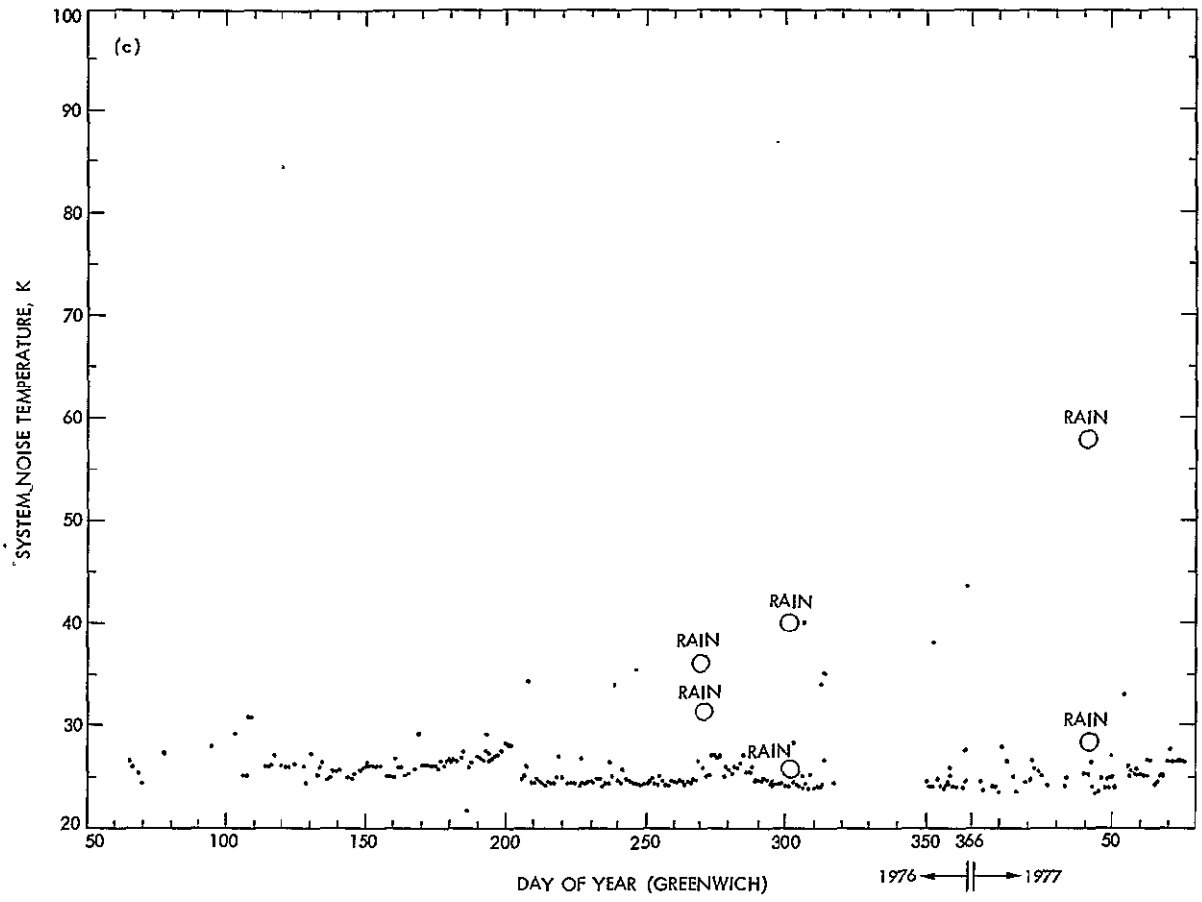
(c)
SYSTEM NOISE TEMPERATURE, K
100
90
80
70
60
50
40
30
20
RAIN
RAIN
RAIN
RAIN
RAIN
RAIN
50
100
150
200
250
300
350
366
50
DAY OF YEAR (GREENWICH)
1976
1977

Fig. 1 (contd)

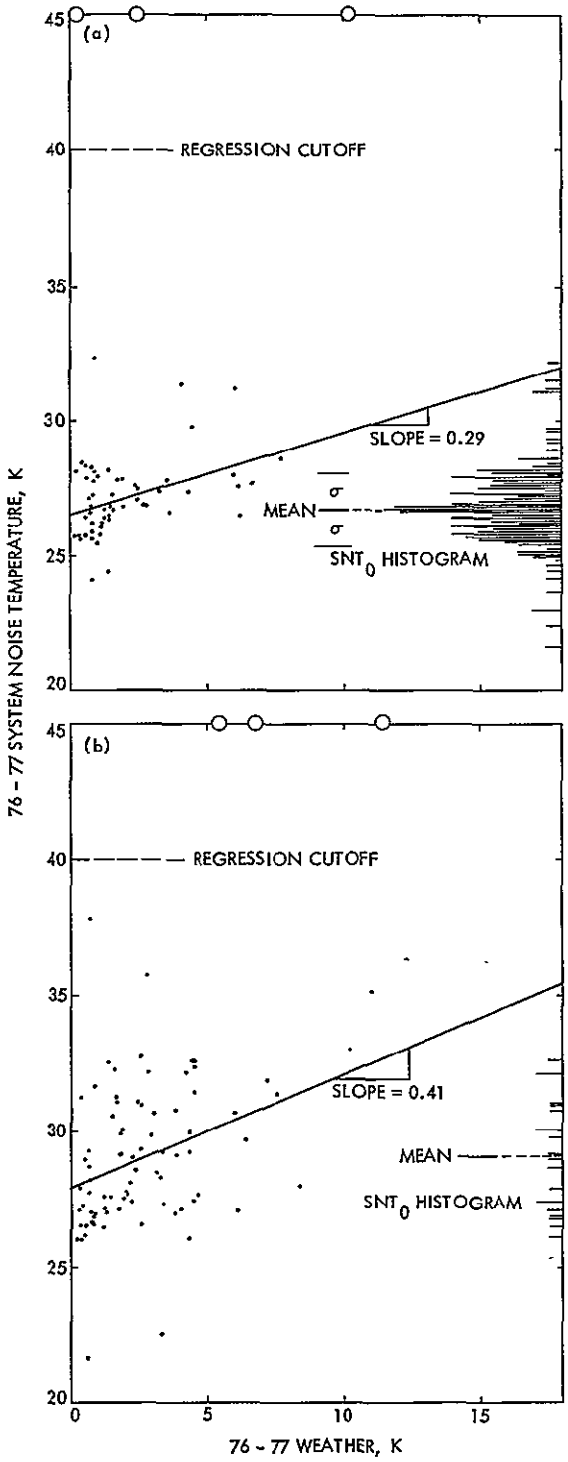

Fig. 2. Scatter plot of 76-77 SNT data vs 76-77 weather data, with histogram of SNT_0 data: (a) Goldstone; (b) Canberra; (c) Madrid

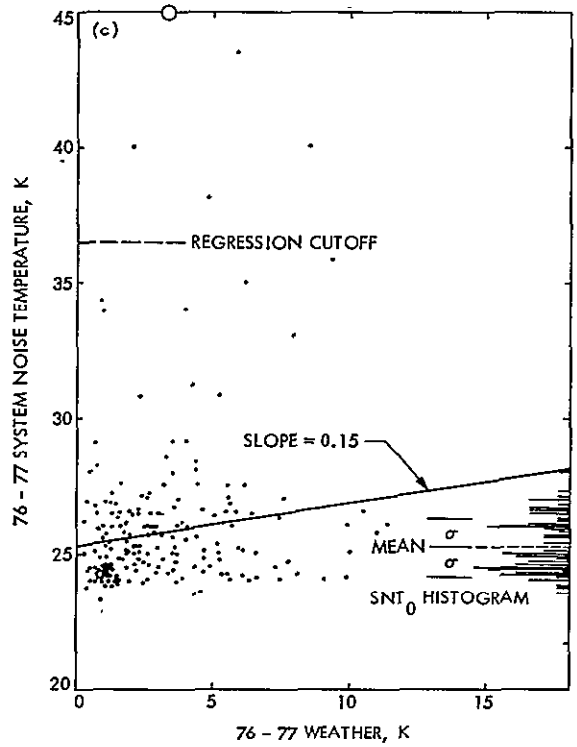
(c)
45
40
35
30
25
20
0
5
10
15
76 - 77 SYSTEM NOISE TEMPERATURE, K
76 - 77 WEATHER, K
REGRESSION CUTOFF
SLOPE = 0.15
MEAN
σ
σ
SNT0 HISTOGRAM

Fig. 2 (contd)

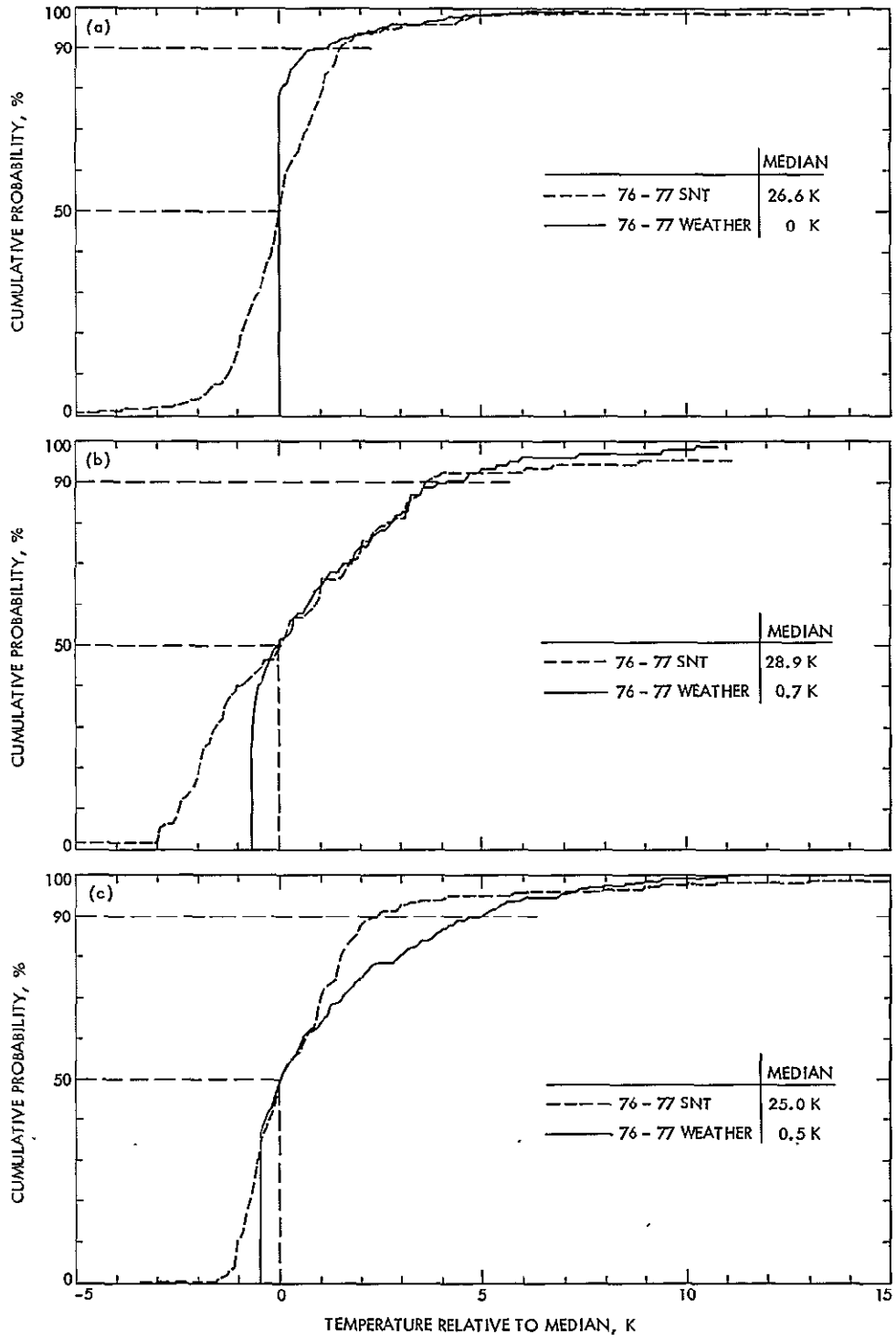

Fig. 3. Cumulative distribution functions of 76-77 SNT and weather data, tied at medians: (a) Goldstone; (b) Canberra; (c) Madrid

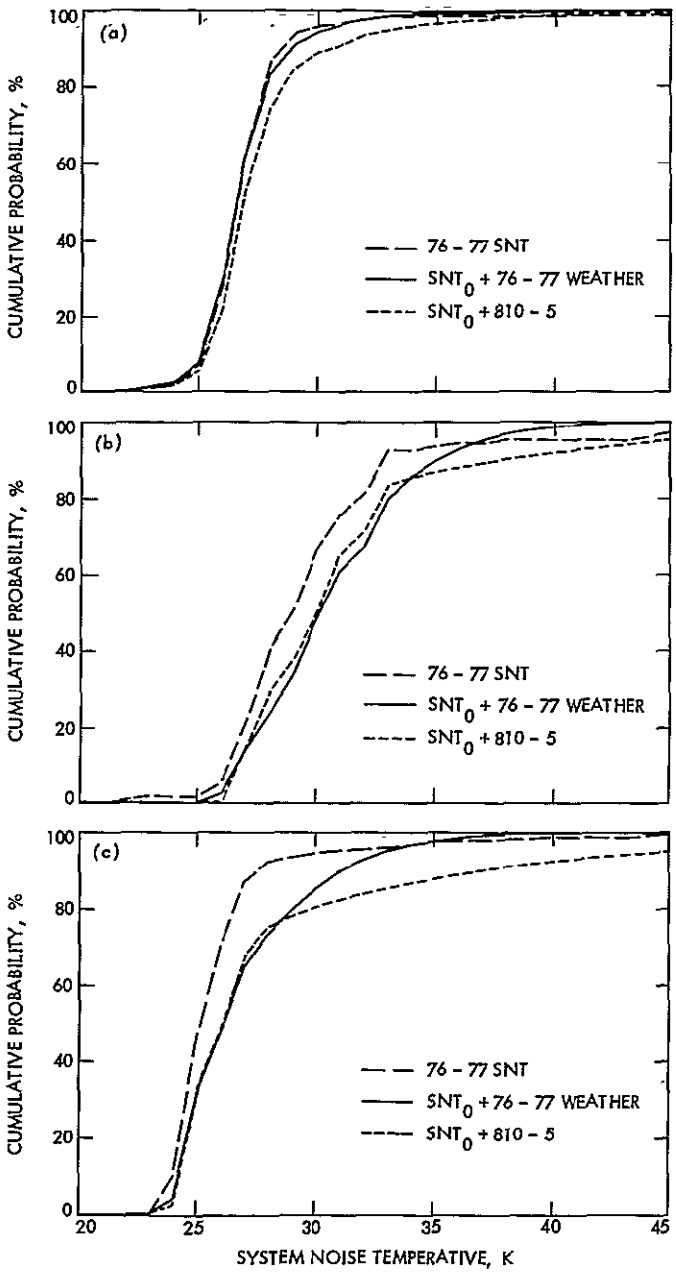

Fig. 4. Cumulative distribution functions of 76-77 SNT, SNT_0 + 76-77 weather, and SNT_0 + 810-5: (a) Goldstone; (b) Canberra; (c) Madrid

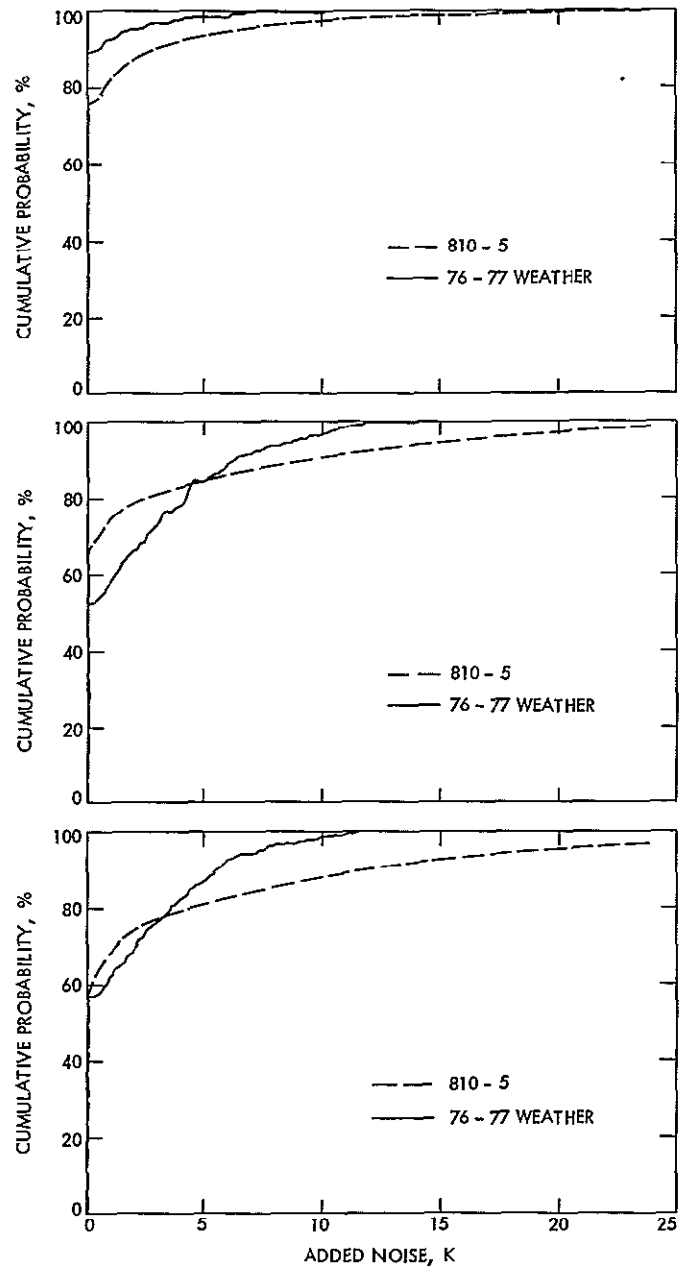

Fig. 5. Cumulative distribution functions of 76-77 weather and 810-5: (a) Goldstone; (b) Canberra; (c) Madrid

Radio-Frequency Carrier Arraying for High-Rate Telemetry Reception

M. H. Brockman
Telecommunications Science and Engineering Division

A method for increasing the sensitivity for radio-frequency reception is to array receiving systems or stations in such a manner as to provide signal-to-noise ratio improvement relative to a single receiving system or station. Radio-frequency carrier arraying for high-rate telemetry that provides signal-to-noise ratio improvement for RF carrier reception and demodulation represents one element of such an array.

I. Introduction

Increasing the sensitivity for radio-frequency reception can be achieved by arraying receiving systems or stations in such a manner as to provide signal-to-noise ratio improvement relative to a single receiving system or station. The following material relates to radio-frequency carrier arraying which provides signal-to-noise ratio improvement for coherent RF carrier reception and RF carrier demodulation. This article is directed toward high-rate telemetry reception with residual RF carrier. A later article will consider RF carrier arraying for low-rate telemetry reception with its attendant smaller RF carrier margin. In this article, the components of equivalent system noise temperature including galactic noise and atmospheric noise are assumed to be statistically independent among the various stations (antennas) of the array. A later article will consider the equivalent system noise temperature in more detail.

II. Receiver Configuration

Figure 1 illustrates the method for achieving RF carrier arraying. Two receiving systems are shown to illustrate the concept. However, additional receiving systems can be added to expand the signal-to-noise improvement capability for RF carrier arraying. In addition, each receiving system can be fed from its own antenna. The input signal to N receiving systems or stations is a radio-frequency signal ($\cos \omega_{RF} t$) phase modulated with telemetry and/or a ranging waveform. Consider, as an example, the case for telemetry where the RF carrier is phase modulated with a square-wave subcarrier ($\mathrm{Cos}\ \omega_{sc} t$) at a peak modulation index m_{pD} that is in turn, biphase-modulated with data $D(t)$. Refer to Fig. 1 during the following development. The signal at the input to the low-noise amplifier in Receiving System 1 is

$$2^{1/2} A_1(t) \cos \left[\omega_{RF} t + \theta_{i1} + D(t+\tau_{i1}) \cdot m_{pD} \cdot \mathrm{Cos}\left(\omega_{sc} t + \theta_{sc_{i1}}\right)\right] + n_{i1}(t) \tag{1a}$$

For a binary modulation waveform, the carrier component becomes

$$2^{1/2}A_1(t) \cdot \cos m_{pD} \cos(\omega_{RF}t + \theta_{i1})$$

The sideband component becomes:

$$+2^{1/2}A_1(t) \sin m_{pD} \cdot D(t + \tau_{i1}) \cdot \text{Cos}(\omega_{sc}t + \theta_{sc_{i1}}) \cdot \sin(\omega_{RF}t + \theta_{i1}) \tag{1b}$$

The noise component is $n_{i1}(t)$.

For Receiving System 2, the corresponding input signal is

$$2^{1/2}A_2(t) \cdot \cos m_{pD} \cos(\omega_{RF}t + \theta_{i2}) + 2^{1/2}A_2(t) \cdot \sin m_{pD} \cdot D(t + \tau_{i2})$$

$$\cdot \text{Cos}(\omega_{sc}t + \theta_{sc_{i2}}) \cdot \sin(\omega_{RF}t + \theta_{i2}) + n_{i2}(t) \tag{2}$$

The other $N-2$ receiving systems will have corresponding input signals. Since Antenna 1 and Antenna 2 are separate antennas, physically separated, the phase shift on the RF carrier (θ_{i2}) and square wave subcarrier ($\theta_{sc_{i2}}$) and the group delay on the data (τ_{i2}) in Receiving System 2 are different from the corresponding phase shifts and modulation group delay in Receiving System 1. In general, the differences in RF carrier phase shift ($\theta_{i2} - \theta_{i1}$), subcarrier phase shift ($\theta_{sc_{i2}} - \theta_{sc_{i1}}$) and modulation group delay ($\tau_{i2} - \tau_{i1}$) will vary with time during a station pass. The above applies in general to all N receiving systems. The terms $n_{i1}(t)$, $n_{i2}(t)$, etc., represent a combination of galactic noise, atmospheric noise, noise in the antenna sidelobes due to the Earth, noise due to losses in microwave reflectors, and noise due to losses in microwave components all lumped with noise due to input amplifier(s). This combined noise is measured relative to reference temperature load(s) connected to the amplifier input (during the measurement) and designated as the operating equivalent system temperature T_{op_1}, T_{op_2}, etc. The noise term $n_{i1}(t)$ has a double-sided noise spectral density $N_{o_1}/2 = (k \cdot T_{op_1} \cdot 1)/2$ watts/Hz where k is Boltzmann's constant, 1.38×10^{-23} joule/°Kelvin. The noise terms $n_{i2}(t)$, etc., have noise spectral densities related to T_{op_2}, etc., as above.

After passing through the low-noise amplifier in Receiving System 1, expression (1) can be written as:

$$k_{LN1}\Big[2^{1/2}A_1(t) \cos m_{pD} \cos(\omega_{RF}t + \theta_{LN10})$$

$$+ 2^{1/2}A_1(t) \sin m_{pD}\, D(t + \tau_{LN10}) \text{Cos}(\omega_{sc}t + \theta_{sc_{LN10}}) \sin(\omega_{RF}t + \theta_{LN10})\Big] + n_{LN10}(t) \tag{3}$$

where k_{LN1} represents the gain of the low-noise amplifier and $k_{LN1} \cdot A_1(t) \ll 1$. The corresponding expression at the output of the low-noise amplifier in Receiving System 2 is:

$$k_{LN2}\Big[2^{1/2}A_2(t) \cos m_{pD} \cdot \cos(\omega_{RF}t + \theta_{LN20}) + 2^{1/2}A_2(t) \sin m_{pD}$$

$$\cdot D(t + \tau_{LN20}) \text{Cos}(\omega_{sc}t + \theta_{sc_{LN20}}) \cdot \sin(\omega_R t + \theta_{LN20})\Big] + n_{LN20}(t) \tag{4}$$

where

$$k_{LN2} \cdot A_2(t) \ll 1.$$

The transfer characteristics of low-noise Amplifiers 1 and 2 modify the RF (i.e., $\theta_{LN10} = \theta_{i_1} + \theta_{LN1}$) and subcarrier (i.e., $\theta_{sc_{LN10}} = \theta_{sc_{L1}} + \theta_{sc_{LN1}}$) phase shift as well as data group delay (i.e., $\tau_{LN10} = \tau_{i1} + \tau_{LN1}$) and band limit and amplify by K_{LN} the receiver noise to $n_{LN10}(t)$ and $n_{LN20}(t)$. The other $N-2$ receiving systems will have corresponding signal expressions at this point in their respective receiving systems.

The first local oscillator signal, which is derived from a voltage-controlled oscillator (VCO_1) through a frequency multiplier (XM) is:

$$2^{1/2} \cos (\omega_{LO} t + \hat{\theta}_{RF}) \tag{5}$$

at the first mixer in Receiving System 1. The term $\hat{\theta}_{RF}$ represents a noisy estimate of the input signal phase to the first mixer. Note that this first local oscillator signal (expression (5)) is also applied to the first mixers of the other receiving systems. In particular, this local oscillator signal will experience an additional phase shift of $Y_2 \cdot 2\pi + \theta_2$ to the first mixer of Receiving System 2, where Y_2 is an integer (a large integer for separate antennas). As a consequence, the first local oscillator signal (Receiving System 2) becomes:

$$2^{1/2} \cos (\omega_{LO} t + \hat{\theta}_{RF} + \theta_2) \tag{6}$$

For Receiving System N, the additional first local oscillator phase shift is $Y_N \cdot 2\pi + \theta_N$ which provides a local oscillator signal $2^{1/2} \cos (\omega_{LO} t + \hat{\theta}_{RF} + \theta_N)$ to its first mixer.

The first IF amplifier in Receiving System 1 has a variable gain including first mixer loss of K_{IFV1} (gain controlled by AGC). The output signal of this first IF amplifier becomes:

$$2^{1/2} A_1 \cos m_{pD} \cos [(\omega_{RF} - \omega_{LO})t + (\theta_{IFV10} - \hat{\theta}_{RF})] + 2^{1/2} A_1 \sin m_{pD} \cdot D(t + \tau_{IFV10}) \cdot \cos (\omega_{sc} t + \theta_{sc_{IFV10}})$$

$$\cdot \sin [(\omega_{RF} - \omega_{LO})t + (\theta_{IFV10} - \hat{\theta}_{RF})] + n_{IFV10}(t) \tag{7}$$

where

$$A_1 = K_{LN1} \cdot K_{IFV1} \cdot A_1(t)$$

is held essentially constant due to AGC and $A_1 \ll 1$. Using the same convention for notation as above for the low-noise amplifier, the terms θ_{IFV10} and $\theta_{sc_{IFV10}}$ represent the sum of all the RF ($\theta_{i1} + \theta_{LN1} + \theta_{IFV1}$) and subcarrier ($\theta_{sc_{i1}} + \theta_{sc_{LN1}} + \theta_{sc_{IFV1}}$) phase shifts, respectively, and data group delay ($\tau_{IFV10} = \tau_{i1} + \tau_{LN1} + \tau_{IFV1}$) up to this point in the system including coaxial lines. The receiver noise is band limited by the combined bandwidth of the low-noise amplifier and first IF amplifier, and amplified by $k_{LN1} \cdot k_{IFV1}$ to $n_{IFV10}(t)$.

The second local oscillator signal in Receiving System 1, which is derived from a frequency standard common to all N receiving systems, is:

$$2^{1/2} \cos(\omega_{REF1} t). \tag{8}$$

The signal at the output of the second IF distribution amplifier (with gain K_{DIST1} which includes coaxial line and the second mixer loss) becomes:

$$K_{DIST1} \cdot 2^{1/2} A_1 \cos m_{pD} \cos [(\omega_{RF} - \omega_{LO} - \omega_{REF1})t + (\theta_{DIST10} - \hat{\theta}_{RF})] + K_{DIST1} \cdot 2^{1/2} A_1 \sin m_{pD}$$

$$\cdot D(t + \tau_{DIST10}) \cos (\omega_{sc} t + \theta_{sc_{DIST10}}) \cdot \sin [\omega_{RF} - \omega_{LO} - \omega_{REF1})t + (\theta_{DIST10} - \hat{\theta}_{RF})] + n_{DIST10}(t) \tag{9}$$

where $\theta_{DIST10} = \theta_{IFV10} + \theta_{DIST1}$, $\theta_{sc_{DIST10}} = \theta_{sc_{IFV10}} + \theta_{sc_{DIST1}}$ and $\tau_{DIST10} = \tau_{IFV10} + \tau_{DIST1}$ (the phase shift and group delay up to this point in the system including coaxial lines). In addition, $n_{DIST10}(t)$ is the band limited and amplified (by $k_{LN1} \cdot k_{IF1} \cdot k_{DIST1}$) receiver noise at the output of the second IF distribution amplifier. The term $(\omega_{RF} - \omega_{LO} - \omega_{REF1})$ becomes $(\omega_{REF1} - \omega_{RF} + \omega_{LO})$ for $\omega_{REF1} > \omega_{RF} - \omega_{LO}$. Note that $(\omega_{REF1} - \omega_{RF} + \omega_{LO})$ or $(\omega_{RF} - \omega_{LO} - \omega_{REF1}) = \omega_{REF2}$. Also frequency $(\omega_{REF2} t)$ is the reference frequency for the phase detector in the RF carrier phase tracking loop, the AGC detector, and the telemetry IF channel phase detector (Fig. 1). Reference frequency $(\omega_{REF2} t)$ is also derived from the frequency standard common to all N receiving systems.

The second IF filter F_{A1} is narrow band relative to the telemetry sidebands and consequently filters the sidebands out. After passing through filter F_{A1}, the signal becomes:

$$K_{A1} \cdot K_{DIST1} \cdot 2^{1/2} A_1 \cos m_{pD} \cdot \cos [\omega_{REF2} t + (\theta_{FA10} - \hat{\theta}_{RF})] + n_{FA10}(t) \tag{10}$$

where $\theta_{FA10} = \theta_{DIST10} + \theta_{FA1}$ and $n_{FA10}(t)$ represents the amplified receiver noise related to T_{OP1} in the noise bandwidth of F_{A1}. Note that since the gain from Expression (10) in Fig. 1 to the AGC detector is constant at $K_{2nd\ IF_1}$ (minus the summing junction loss), the attenuator (A_1) at the input to the second IF amplifier in Receiving System 1 provides a means for setting the signal level at (9) to provide the required output level for the telemetry subcarrier spectrum in Receiving System 1 when RF carrier arraying with Receiving System 2, etc.

Proceeding in Receiving System 2 as above (for Receiving System 1), the output signal of the first IF amplifier in Receiving System 2 (from expressions (4) and (6)) becomes:

$$2^{1/2} A_2 \cos m_{pD} \cos [(\omega_{RF} - \omega_{LO})t + (\theta_{IFV20} - \hat{\theta}_{RF} - \theta_2)] + 2^{1/2} A_2 \sin m_{pD} \cdot D(t + \tau_{IF20}) \cdot \cos (\omega_{sc} t + \theta_{sc_{IFV20}})$$

$$\cdot \sin [(\omega_{RF} - \omega_{LO})t + (\theta_{IFV20} - \hat{\theta}_{RF} - \theta_2)] + n_{IFV20}(t) \tag{11}$$

where

$$A_2 = (K_{LN2} K_{IFV2} \cdot A_2(t))$$

and is held essentially constant due to AGC and $A_2 \ll 1$. Note that the receiving systems considered here are similar so that the signal-to-noise ratio at this point (11) in Receiving System 2 is nearly equal to that at the corresponding point in Receiving System 1. (The concept developed herein however, can be applied to arraying receiving systems where the signal-to-noise ratios are not equal.) The second local oscillator signal in Receiving System 2 is derived from a voltage-controlled crystal oscillator (VCO_2) through a frequency multiplier (XQ). From inspection of Fig. 1, the frequency *VCO2 XQ* is equal to ($\omega_{REF1}t$) although its phase is modified by differences in carrier phase as pointed out in the discussion relative to expressions (2), (4), and (6). The second local oscillator signal can be expressed as:

$$2^{1/2}\cos(\omega_{REF1}t - \hat{\theta}_{LO2}) \tag{12}$$

where $\hat{\theta}_{LO2}$ represents the estimate of the RF carrier input phase to the second mixer (Receiving System 2). The phase estimate $\hat{\theta}_{LO2}$ is derived from a phase-locked loop whose closed-loop noise bandwidth is a small fraction of the closed-loop noise bandwidth of the RF carrier phase tracking loop discussed below (expression (16)). Hence the noise on phase estimate $\hat{\theta}_{LO2}$ is much less than that on $\hat{\theta}_{RF}$ (first local oscillator signal). Note that if the carrier signal-to-noise ratio in Receiving System 2 were significantly less than that in Receiving System 1, this effect could be offset by making the closed-loop noise bandwidth in Receiving System 2 a still smaller fraction of that in Receiving System 1.

The signal at the output of the second IF distribution amplifier (with gain K_{DIST2} which includes coaxial line and second mixer loss) is.

$$K_{DIST2} \cdot 2^{1/2}A_2 \cos m_{pD} \cos[(\omega_{RF} - \omega_{LO} - \omega_{REF1})t + (\theta_{DIST20} - (\hat{\theta}_{RF} - \hat{\theta}_{LO2}))]$$

$$+ K_{DIST2} \cdot 2^{1/2}A_2 \sin m_{pD} \cdot D(t+\tau_{DIST20}) \cos(\omega_{sc}t + \theta_{sc_{DIST20}})$$

$$\cdot \sin[(\omega_{RF} - \omega_{LO} - \omega_{REF1})t + (\theta_{DIST20} - (\hat{\theta}_{RF} - \hat{\theta}_{LO2}))] + n_{DIST20}(t) \tag{13}$$

where $\theta_{DIST20}(= \theta_{IFV20} - \theta_2 + \theta_{DIST2})$, $\theta_{sc_{DIST20}}$ and τ_{DIST20} represent the phase shift and group delay up to this point in System 2 including coaxial lines. Note from Fig. 1 that $(\omega_{RF} - \omega_{LO2} - \omega_{REF1}) = \omega_{REF2}$. The second IF distribution amplifier provides its output signal to the IF filter F_{A2} and to the telemetry IF channel in Receiving System 2. The second IF filter F_{A2} in Receiving System 2 has the same noise bandwidth as IF filter F_{A1} in Receiving System 1 (by design). The signal at the output of IF filter F_{A2} (with gain K_{A2}) can be expressed as:

$$K_{A2} \cdot K_{DIST2} \cdot 2^{1/2}A_2 \cos m_{pD} \cdot \cos[\omega_{REF2}t + (\theta_{FA20} - (\hat{\theta}_{RF} - \hat{\theta}_{LO2}))] + n_{FA20}(t) \tag{14}$$

where $\theta_{FA20} = \theta_{DIST20} + \theta_{FA2}$ and $n_{FA20}(t)$ represents the amplified receiver noise in the noise bandwidth of F_{A2} at the operating equivalent system noise temperature T_{op2}. This signal is provided as an input to the summing junction (Fig. 1) as well as to the RF carrier loop and AGC in Receiving System 2. Note that the phase shift from the output of IF filter F_{A2} (expression 14) in Fig. 1 to the phase detector which provides the error signal to the tracking filter (Receiving System 2) is a constant by design. Consequently the phase shifter marked *A*, which is in series with the reference 2 input ($\omega_{REF2}t$) in Receiving System 2 provides a means for setting the RF phase of expression (14) equal to the RF phase of expression (10) in Receiving System 1 at the summing junction. The second IF amplifier gain $K_{2nd\ IF_2}$ is designed so that the signal level ($K_{A2} \cdot K_{DIST2} \cdot 2^{1/2}A_2$) can be set as required relative to the signal level ($K_{A1} \cdot K_{DIST1} \cdot 2^{1/2}A_1$) in Receiving System 1 at the input to the summing junction.

III. Predetection Signal-to-Noise Ratio and Resultant Phase Noise

A. Receiving System 1

Assume for the moment that the second receiving system shown in Fig. 1 is not connected to the summing junction. The predetection carrier signal-to-noise power ratio in Receiving System 1 represented by expression (10) is:

$$P_{c1}/P_{n1} = \frac{A_1^2 \cos^2 m_{pD}}{NBW_{F_{A1}} \cdot N_{o1}} \quad \text{or} \quad \frac{P_{c1}}{NBW_{F_{A1}} \cdot N_{o1}} \tag{15}$$

where $NBW_{F_{A1}}$ represents the noise bandwidth of F_{A1} and N_{o1} is the one-sided noise spectral density for Receiving System 1 which was defined in the discussion relative to expression (1). The receiving system contains a second-order RF carrier phase tracking loop which utilizes a band pass limiter and a sinusoidal phase detector. Utilizing the information in Refs. 1 and 2 and limiting $\sigma_n \leqslant 1$ radian, the rms phase noise σ_{ϕ_n} at the output of the RF carrier tracking loop (i.e., on the first local oscillator signal) can be expressed as:

$$\sigma_{\phi_{n1}} = \left(\frac{\frac{N_{o1}}{2} \cdot 2B_{L1}}{P_{c1}}\right)^{1/2} \cdot \left[\frac{1 + \frac{P_{c1}}{NBW_{F_{A1}} N_{o1}}}{0.862 + \frac{P_{c1}}{NBW_{F_{A1}} N_{o1}}}\right]^{1/2} \cdot \left[\frac{N_{o1}B_{L_1}}{P_{c_1}} \cdot \frac{e\left(\frac{N_{o1}B_{L1}}{P_{c1}}\right)}{\sinh\left(\frac{N_{o1}B_{L1}}{P_{c1}}\right)}\right]^{1/2} \text{ radians, rms}$$

or collecting terms.

$$\sigma_{\phi_{n1}} = \frac{\frac{N_{o1}}{2} \cdot 2B_{L1}}{P_{c1}} \left[\frac{1 + \frac{P_{c1}}{NBW_{F_{A1}} \cdot N_{o1}}}{0.862 + \frac{P_{c1}}{NBW_{F_{A1}} \cdot N_{o1}}} \cdot \frac{e\left(\frac{N_{o1}B_{L1}}{P_{c1}}\right)}{\sinh\left(\frac{N_{o1}B_{L1}}{P_{c1}}\right)}\right]^{1/2} \text{ radians, rms} \tag{16}$$

The two-sided closed-loop noise bandwidth $2B_{L1}$ can be expressed as:

$$2B_{L1} = \frac{2B_{LO1}}{r_o + 1}\left(1 + r_o \frac{\alpha}{\alpha_{o1}}\right)$$

where $r_o = 2$ by design (0.707 damping) and $2B_{LO1}$ is the design point (threshold) two-sided closed-loop noise bandwidth in Receiving System 1. The term α_1 is the limiter suppression factor resulting from the noise-power-to-carrier-power ratio in NBW_{FA1}, α_1 has a value of α_{o1} at design point (threshold). At threshold, the predetection carrier signal-to-noise ratio in a noise bandwidth equal to the threshold closed-loop noise bandwidth ($2B_{LO}$) is unity (i.e., $P_c/2B_{LO} \cdot N_o = 1$).

B. Receiving System 1 Arrayed with Receiving System(s) 2 Through N

Consider connection of the second Receiving System to the summing junction. Designate the voltage coupling of Receiving System 2 into the summing junction as β_2 (relative to Receiving System 1). In addition, designate the carrier power-to-noise spectral density ratio of Receiving System 2 as γ_2^2 of that in Receiving System 1 where $\gamma_2 \leqslant 1$. Note that the carrier phase

noise that is fed from Receiving System 2 into the summing junction is small (because of much narrower closed loop noise bandwidths as discussed above) relative to the phase noise error on the first local oscillator. Consequently, the Receiver 1 and 2 carrier signals fed into the summing junction are coherent with a small differential phase jitter. However, their receiver noise voltages fed into the summing junction are statistically independent. As a result, the summed predetection carrier signal-to-noise power ratio at the output of the summing junction becomes:

$$\frac{P_{c_{1\Sigma 1,2}}}{P_{n_{1\Sigma 1,2}}} = \frac{(A_1 \cos m_{pD} + \beta_2 A_2 \cos m_{pD})^2}{NBW_{F_{A1}} \cdot N_{o1} + \beta_2^2 NBW_{F_{A2}} \cdot N_{o2}}$$

which can be expressed as

$$\frac{P_{c_{1\Sigma 1,2}}}{P_{n_{1\Sigma 1,2}}} = \frac{A_1^2 \cos^2 m_{pD}}{NBW_{F_{A1}} \cdot N_{o1}} \cdot \frac{(1 + \beta_2\gamma_2)^2}{1 + \beta_2^2}$$

or

$$\frac{P_{c_{1\Sigma 1,2}}}{P_{n_{1\Sigma 1,2}}} = \frac{P_{c1}}{NBW_{F_{A1}} \cdot N_{o1}} \cdot \frac{(1 + \beta_2\gamma_2)^2}{1 + \beta_2^2} \tag{17}$$

The rms phase noise $\sigma_{\phi_{n\Sigma 1,2}}$ at the output of the RF carrier tracking loop (i.e., on the first local oscillator signal) in Receiving System 1 due to predetection carrier signal-to-noise ratio can now be determined for two receiving systems arrayed by incorporating the information in expression (17) above into the expression for σ_{ϕ_n} (see expression (16)). The resulting expression becomes:

$$\sigma_{\phi_{n1\Sigma 1,2}} = \frac{\frac{N_{o1}}{2} \cdot 2B_{L1}}{P_{c1}} \cdot \frac{1}{\eta_2} \left[\frac{1 + \frac{P_{c1} \cdot \eta_2}{NBW_{F_{A1}} \cdot N_{o1}}}{0.862 + \frac{P_{c1} \cdot \eta_2}{NBW_{F_{A1}} \cdot N_{o1}}} \cdot \frac{\exp\left(\frac{N_{o1} B_{L1}}{P_{c1} \cdot \eta_2}\right)}{\sinh\left(\frac{N_{o1} \cdot B_{L1}}{P_{c1} \cdot \eta_2}\right)} \right]^{1/2} \text{ radians, rms} \tag{18}$$

where

$$\eta_2 = \frac{(1 + \beta_2\gamma_2)^2}{1 + \beta_2^2}$$

for two receiving systems arrayed. Note that for N similar receiving systems arrayed, the predetection carrier signal-to-noise power ratio in Receiving System 1 becomes

$$\frac{P_{c1\Sigma 1,\cdots N}}{P_{n1\Sigma 1,\cdots N}} = \frac{P_{c1}}{NBW_{F_{A1}} \cdot N_{o1}} \cdot \frac{(1 + \beta_2\gamma_2 + \beta_3\gamma_3 + \cdots + \beta_N\gamma_N)^2}{1 + \beta_2^2 + \beta_3^2 + \cdots + \beta_N^2} \tag{19}$$

The corresponding $\sigma_{\phi_{1n\Sigma 1 \cdots N}}$ can now be determined and becomes

$$\sigma_{\phi_{n1\Sigma 1,\cdots,N}} = \frac{\frac{N_{o1}}{2} \cdot 2B_{L1}}{P_{c1}} \cdot \frac{1}{\eta_N} \left[\frac{1 + \frac{P_{c1} \cdot \eta_N}{NBW_{F_{A1}} N_{o1}}}{0.862 + \frac{P_{c1} \cdot \eta_N}{NBW_{F_{A1}} N_{o1}}} \cdot \frac{\exp\left(\frac{N_{o1} \cdot B_{L1}}{P_{c1} \cdot \eta_N}\right)}{\sinh\left(\frac{N_{o1} \cdot B_{L1}}{P_{c1} \cdot \eta_N}\right)} \right]^{1/2} \text{ radians, rms} \qquad (20)$$

where

$$\eta_N = \frac{(1 + \beta_2\gamma_2 + \cdots + \beta_N\gamma_N)^2}{1 + \beta_2^2 + \cdots + \beta_N^2}$$

for N receiving systems arrayed.

The preceding paragraph has addressed determination of the rms phase noise $\sigma_{\phi n}$ at the output of the RF carrier tracking loop (first local oscillator signal) in Receiving System 1 due to the predetection carrier signal-to-noise ratio for two (up to N) receiving systems arrayed. Inspection of Fig. 1 and expression (14) above indicates that an additional phase noise term due to $\hat{\theta}_{LO2}$ (noise at the output of the carrier tracking loop in Receiving System 2) coupled into the summing junction must also be considered for two receiving systems arrayed. Note that since the closed-loop noise bandwidth of Receiving System 1 carrier tracking loop is much wider than that in Receiving System 2, the phase noise on $\hat{\theta}_{LO2}$ is tracked by this wider phase tracking loop and consequently appears as an (extremely small) error signal at the output of the summing junction. As a result of the above (and from inspection of Fig. 1), the term $\hat{\theta}_{RF}$ for the first oscillator signal (see expression (5) and (6)) now becomes

$$\hat{\theta}_{RF} + \frac{\beta_2\hat{\theta}_{LO2}}{1 + \beta_2}$$

Consequently expressions (7), (9), and (10) in Receiving System 1 and expressions (11), (13) and (14) in Receiving System 2 now have the term $\hat{\theta}_{RF}$ replaced by

$$\hat{\theta}_{RF} + \frac{\beta_2\hat{\theta}_{LO2}}{1 + \beta_2}$$

As a result, the total rms phase noise at the output of the principal (Receiving System 1) carrier tracking loop (i.e., on the first local oscillator signal) becomes:

$$\left(\sigma^2_{\phi_{n1\Sigma 1,2}} + \left(\frac{\beta_2\sigma_{\phi_{n2}}}{1 + \beta_2}\right)^2\right)^{1/2} \qquad (21)$$

for two receiving systems arrayed. The first term, $\sigma_{\phi_{n1\Sigma 1,2}}$ is the rms phase noise due to predetection carrier signal-to-noise ratio at the output of the summing junction (i.e., expression (18)). The rms phase noise $\sigma_{\phi_{n2}}$ contained in the second term above will be developed in the following material. Note that for N receiving systems arrayed the total rms phase noise on the first local oscillator signal becomes:

$$\left(\sigma^2_{\phi_{n1\Sigma 1,\cdots,N}} + \left(\frac{\beta_2 \sigma_{\phi_{n2}}}{1+\beta_2}\right)^2 + \cdots + \left(\frac{\beta_N \sigma_{\phi_{nN}}}{1+\beta_N}\right)^2\right)^{1/2} \tag{22}$$

where $\sigma_{\phi_{n1\Sigma 1,\cdots,N}}$ is expression (20) above. The expressions for $\sigma_{\phi_{n2}}$ (and $\sigma_{\phi_{nN}}$) will be developed in the following material.

The total rms phase noise shown in expressions (21) and (22) can be considered as due to an equivalent predetection carrier signal-to-noise ratio for two up to N receiving systems arrayed. Comparison of this equivalent predetection carrier signal-to-noise ratio with the initial predetection carrier signal-to-noise ratio in a single Receiving System (i.e., system 1) provides the improvement due to carrier arraying for high-rate telemetry.

C. Receiving System 2 Through N

Consider next the carrier phase tracking loop in Receiving System 2. Since the closed-loop noise bandwidth of the carrier tracking loop in Receiving System 2 is much narrower than that in Receiving System 1, the phase noise in Receiving System 1 carrier tracking loop (which has an rms value $\sigma_{\phi_{n\Sigma 1,2}}$ for two systems arrayed) produces a predetection carrier signal voltage reduction which has an expected value of

$$1 - \frac{\sigma^2_{\phi_{n\Sigma 1,2}}}{2}$$

The predetection carrier signal-to-noise power ratio in Receiving System 2 represented by expression (14) is:

$$\frac{P_{c_{2\Sigma 1,2}}}{P_{n2}} = \frac{A_2 \cos^2 m_{pD} \left(1 - \frac{\sigma^2_{\phi_{n1\Sigma 1,2}}}{2}\right)^2}{NBW_{F_{A2}} \cdot N_{o2}} \quad \text{or} \quad \frac{P_{c_2} \left(1 - \frac{\sigma^2_{\phi_{n1\Sigma 1,2}}}{2}\right)^2}{NBW_{F_{A2}} \cdot N_{o2}} \tag{23}$$

The carrier tracking loop in Receiving System 2 is a second-order loop ($r_o = 2$) which also utilizes a band pass limiter and a sinusoidal phase detector. The rms phase noise $\sigma_{\phi_{n2}}$ in Receiving System 2 carrier tracking loop becomes

$$\sigma_{\phi_{n2\Sigma 1,2}} = \frac{\frac{N_{o2}}{2} \cdot 2B_{L2}}{P_{c_{2\Sigma 1,2}}} \cdot \left[\frac{1 + \frac{P_{c_{2\Sigma 1,2}}}{NBW_{F_{A2}} N_{o2}}}{0.862 + \frac{P_{c_{2\Sigma 1,2}}}{NBW_{F_{A2}} \cdot N_{o2}}} \cdot \frac{\exp\left(\frac{N_{o2} \cdot B_{L2}}{P_{c_{2\Sigma 1,2}}}\right)}{\sinh\left(\frac{N_{o2} \cdot B_{L2}}{P_{c_{2\Sigma 1,2}}}\right)}\right]^{1/2} \text{ radians, rms} \tag{24}$$

where

$$P_{c_{2\Sigma 1,2}} = P_{c2}\left(1 - \frac{\sigma^2_{\phi_{n1\Sigma 1,2}}}{2}\right)^2$$

for two receiving systems arrayed.

For N receiving systems arrayed, the predetection carrier signal-to-noise power ratio in Receiving System 2 becomes:

$$\frac{P_{c_{2\Sigma 1,\cdots N}}}{P_{n2}} = \frac{P_{c2}\left(1 - \frac{\sigma^2_{\phi_{n1\Sigma 1,\cdots N}}}{2}\right)^2}{NBW_{F_{A2}} \cdot N_{o2}} \tag{25}$$

and the corresponding $\sigma_{\phi_{n2\Sigma 1,\cdots N}}$ can be determined by substitution of $P_{c_{2\Sigma 1,\cdots N}}$ in place of $P_{c_{2\Sigma 1,2}}$ in expression (24). Note that the term $\sigma_{\phi_{n1\Sigma 1,\cdots N}}$ in expression (25) above should be replaced by an rms value similar to that shown in expression (22) with the term

$$\frac{\beta_2 \sigma_{\phi_{n2}}}{1+\beta_2}$$

deleted. This iteration provides a very small effect on carrier signal-to-noise ratio improvement for the design parameters and number of receiving systems considered in this report. The predetection carrier signal-to-noise power ratio in Receiving System N becomes:

$$\frac{P_{c_{N\Sigma 1,\cdots N}}}{P_{n_N}} = \frac{P_{c_N}\left(1 - \frac{\sigma^2_{\phi_{n1\Sigma 1,\cdots,N}}}{2}\right)^2}{NBW_{F_{AN}} \cdot N_{o_N}} \tag{26}$$

The resultant rms phase noise $\sigma_{\phi_{nN}}$ in Receiving System N carrier tracking loop is:

$$\sigma_{\phi_{nN\Sigma 1,\cdots,N}} = \frac{\frac{N_{o_N}}{2} \cdot 2B_{LN}}{P_{c_{N\Sigma 1,\cdots,N}}} \cdot \left[\frac{1 + \frac{P_{c_{N\Sigma 1,\cdots,N}}}{NBW_{F_{AN}} \cdot N_{o_N}}}{0.862 + \frac{P_{c_{N\Sigma 1,\cdots,N}}}{NBW_{F_{AN}} \cdot N_{o_N}}} \cdot \frac{\exp\left(\frac{N_{o_N} \cdot B_{L_N}}{P_{c_{N\Sigma 1,\cdots,N}}}\right)}{\sinh\left(\frac{N_{o_N} \cdot B_{L_N}}{P_{c_{N\Sigma 1,\cdots,N}}}\right)}\right]^{1/2} \text{ radians, rms} \tag{27}$$

IV. Acquisition

The following discussion relates to RF carrier acquisition for the arrayed receiving systems shown in Fig. 1. Prior to RF carrier acquisition, receiving systems 2 through N are disconnected (switched out) from the summing junction. The first local oscillator ($MxVCO_1$) in Receiving System 1 is swept or set in frequency to the expected (or predicted) reception frequency. Radio frequency carrier acquisition is obtained for carrier levels from threshold (defined in discussion following expression (16)) to strong signal levels. RF carrier acquisition in Receiving System 1 provides the first local oscillator signal for receiving systems 2 through N.

The carrier tracking loop in Receiving System 2 is acquired by setting the oscillator frequency (VCO_2xQ) equal (zero beat frequency) to reference frequency 1 (Fig. 1). Receiving System 2 is now connected into the summing junction to accomplish RF carrier arraying of two systems. This procedure is repeated in rapid succession until all remaining receiving systems are acquired and RF carrier arrayed for reception.

V. Performance

The improvement in carrier signal-to-noise ratio that can be achieved by arraying receiving systems for high-rate telemetry can now be determined for a representative set of design parameters using the preceding development. The following parameters are used for Receiving System 1:

Threshold two-sided noise bandwidth ($2B_{LO1}$) = 12 Hz

Predetection IF filter noise bandwidth (NBW_{FA1}) = 2200 Hz

The parameters used for Receiving System 2 through N are:

Threshold two-sided noise bandwidth ($2B_{LO2,\cdots,N}$) = 0.1 Hz

Predetection IF filter noise bandwidth ($NBW_{FA2,\cdots,N}$) = 2200 Hz

Figure 2 shows the carrier signal-to-noise ratio improvement resulting from arraying up to twelve similar receiving systems for high-rate telemetry reception. Performance is shown for γ_N values of 1.0, 0.95, 0.9 and 0.84 where β_N is equal to γ_N. As developed earlier in this report, γ_N^2 is the carrier signal-to-noise spectral density ratio of Receiving System N relative to Receiving System 1. Also β_N is the voltage coupling of Receiving System N into the summing junction relative to Receiving System 1.

Figure 3 shows the effect on carrier signal-to-noise ratio improvement for γ_N values of 1.0, 0.95, 0.9 and 0.84 when β_N is varied. Performance is shown for two, three, four and ten receiving systems arrayed for high-rate telemetry. The effect of the carrier phase noise from Receiving Systems 2 through N (see expression (22)) coupled into the principal (Receiving System 1) RF carrier tracking loop can be seen in Fig. 3. This effect becomes more apparent as the number of systems arrayed is increased. Note that, for ten receiving systems arrayed for high-rate telemetry reception, reducing the coupling β from unity results in an increase in carrier signal-to-noise ratio improvement.

Some initial measurements of radio-frequency carrier signal-to-noise ratio improvement have been made in the laboratory by arraying two, three, and four receivers. For convenience, these initial measurements were made with the following parameters; Receiver 1:

Threshold two-sided noise bandwidth ($2B_{LO1}$) = 152 Hz

Predetection IF filter noise bandwidth (NBW_{FA1}) = 2200 Hz

The parameters for Receivers 2, 3 and 4 were:

Threshold two-sided noise bandwidth ($2B_{LO2,3,4}$) = 1 Hz

Predetection IF filter noise bandwidth ($NBW_{FA2,3,4}$) = 2200 Hz

These parameters provide essentially the same performance as shown in Figs. 2 and 3 for two, three, and four receiving systems arrayed.

Measurement of two receivers (1 and 2) arrayed, with γ_2 and β_2 equal to one, provided a RF carrier signal-to-noise ratio improvement of 2.7 dB. Calculated improvement is 2.8 dB. Note that a measurement with $\beta_2 = 1$ and $\gamma_2 = 0.88$ provided a carrier signal-to-noise ratio improvement of 2.4 dB, which agrees with predicted performance. Measurement of three receivers (1, 2, and 3) arrayed provided a 4.1-dB improvement in RF carrier signal-to-noise ratio. For this measurement γ_2 and $\gamma_3 = 1$, $\beta_2 = 0.4$ and $\beta_3 = 0.49$. These parameters provide a predicted improvement of 4.1 dB. With four receivers (1, 2, 3, and 4) arrayed, measured improvement in RF carrier signal-to-noise ratio was 4.6 dB. For this measurement γ_2 and $\gamma_3 = 1$, $\gamma_4 = 0.86$, $\beta_2 = 0.3$, $\beta_3 = 0.38$ and $\beta_4 = 0.36$. These parameters provide a predicted improvement of 4.6 dB. For the above measurements, the effective system noise temperatures of receivers 1, 2, 3 and 4 were 3540, 3360, 3360 and 4640 degrees kelvin, respectively. It is planned to make further measurements with parameters optimized (i.e., γ_N and $\beta_N = 1$).

References

1. Tausworthe, R. C., *Theory and Practical Design of Phase-Locked Receivers, Vol. I.*, TR 32-819, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1966.

2. Tausworthe, R. C., "Limiters in Phase-Locked Loops: A Correction to Previous Theory," *Space Programs Summary No. 37-54*, Vol. III, pp. 201-203, Jet Propulsion Laboratory, Pasadena, Calif., 1968.

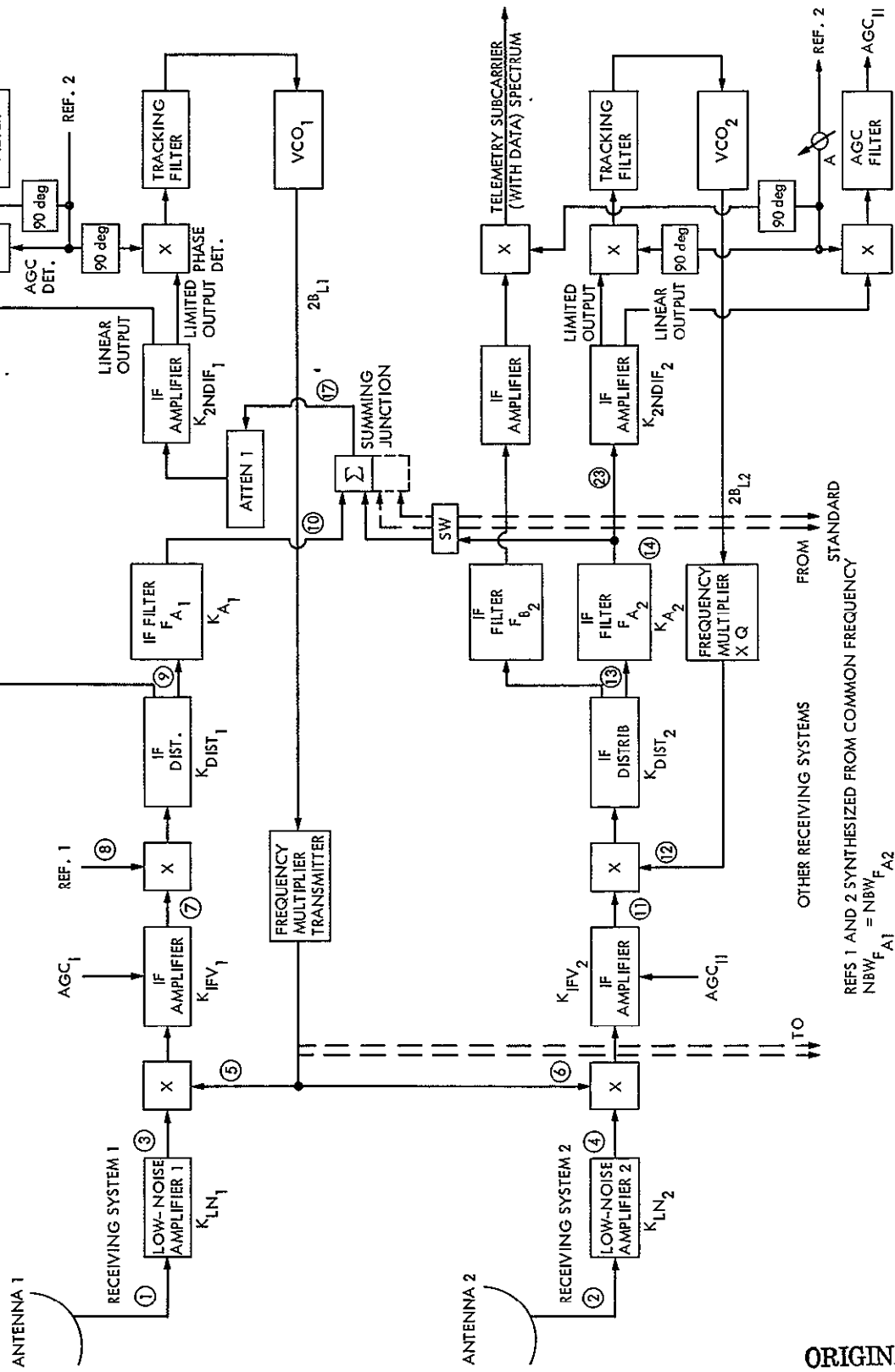

Fig. 1. Radio frequency carrier arraying for high-rate telemetry reception

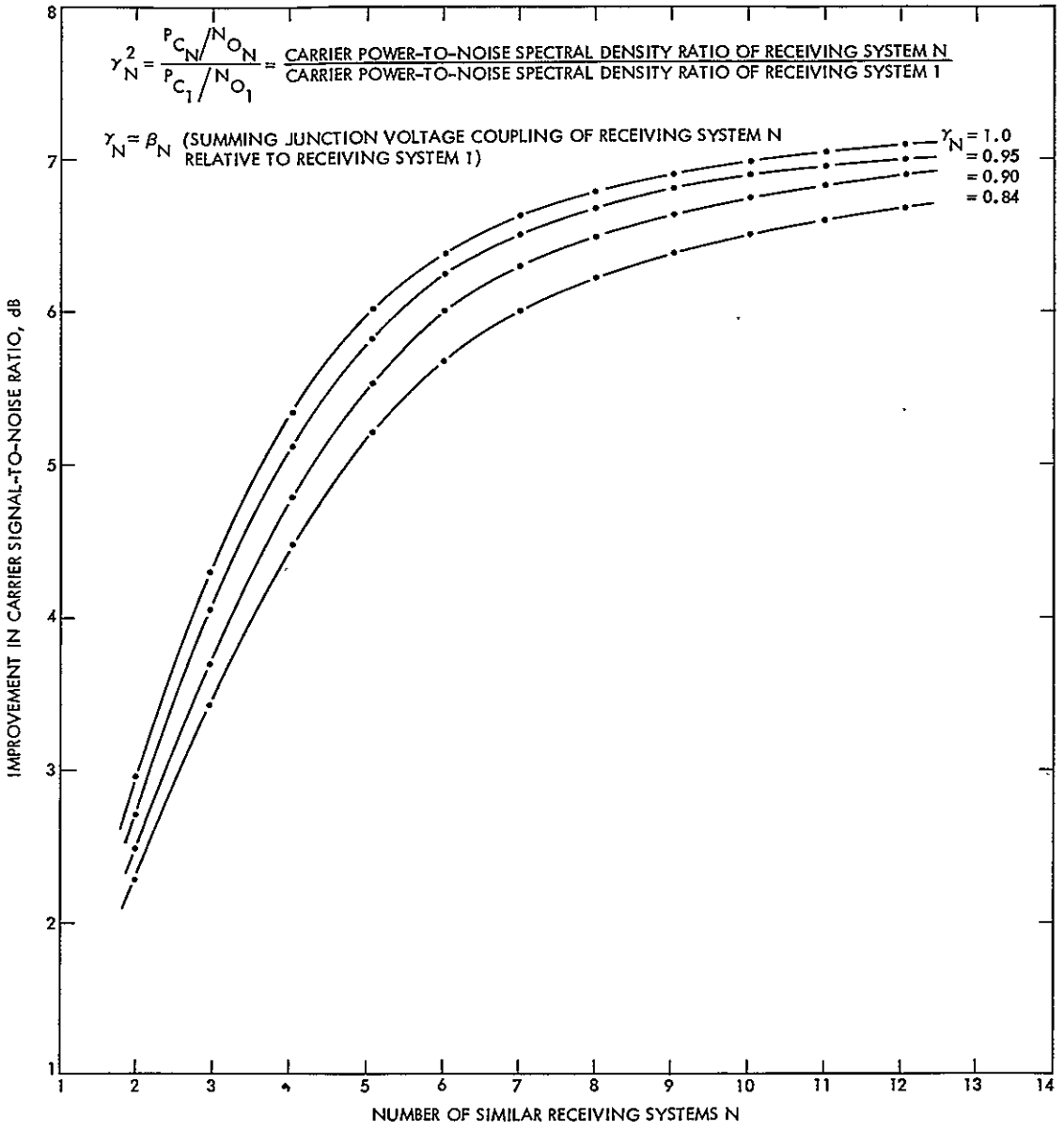

Fig. 2. Effect of summing junction voltage coupling on carrier signal-to-noise ratio improvement (receiver arraying for high-rate telemetry reception)

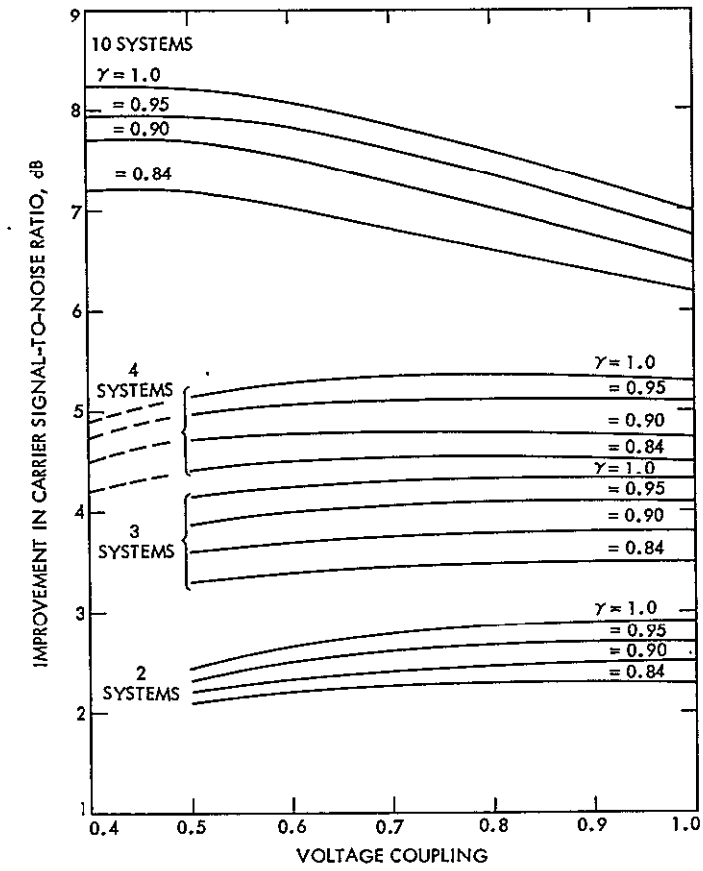

Fig. 3. Signal-to-noise ratio improvement for radio frequency carrier arraying for high-rate telemetry reception

CCIR Paper on the Radiocommunications Requirements for Systems to Search for Extraterrestrial Life

D. Nightingale
Telecommunications Systems Section

Three separate JPL papers and one Japanese paper were originally submitted to Study Group 2 of the International Radio Consultative Committee (CCIR) on the subject of the search for extraterrestrial intelligence. During the Final Meeting of Study Group 2 in Geneva in September – October 1977, a working party headed by Mr. Sam Brunstein of JPL combined these four papers into a single report. This article presents this report in its final CCIR format. The report considers propagation factors, preferred frequency bands, system characteristics and requirements, and interference.

In the three previous issues of the *DSN Progress Report* N. F. de Groot has presented a series of papers on telecommunications for deep space research as adopted by Study Group 2 of the International Radio Consultative Committee (CCIR). The background information contained in the first of those articles *(DSN Progress Report 42-42)* applies equally to this present article.

The Spectrum Engineering Group of the Jet Propulsion Laboratory supports the NASA Ames Research Center in assisting the U.S. Department of State to prepare for the 1979 General World Administrative Radio Conference. This support consists of conducting studies and preparing CCIR documents in connection with the allocation and protection of radio frequency bands to be used in the search for extraterrestrial intelligence (SETI).

Three separate reports on SETI were prepared within the Spectrum Engineering Group and were presented by the United States to the CCIR Study Group 2 Final Meeting held in Geneva during September 1977. A working group, headed by Mr. Sam Brunstein of JPL, was assigned the task of combining the three papers together with a fourth (presented by the Japanese delegation) into a single report. The composite report was recommended to the full international Study Group 2 in Geneva, was approved, and will be presented for adoption by the CCIR Plenary Assembly in mid-1978. Upon approval by that body, the report will be published as part of CCIR Volume 2, *Space Research and Radio Astronomy.*

As with the reports of deep space research telecommunications previously presented in the *Deep Space Network Progress Report,* this SETI paper is also reproduced in its original form to illustrate both the style and format of CCIR documents.

Documents
CCIR Study Groups
Period 1974 - 1978

Doc. 2/277-E
23 September 1977

Original: English

WORKING GROUP 2-D

DRAFT NEW REPORT

RADIOCOMMUNICATION REQUIREMENTS FOR SYSTEMS TO SEARCH FOR EXTRATERRESTRIAL LIFE

(Question 17/2)

1. Introduction

Question 17/2 on Radiocommunication Requirements for Systems to Search for Extraterrestrial Life has been adopted by the CCIR. The present Report discusses the general background and technical matters related to this topic.

1.1 Background

Many scientists believe that life is common in our galaxy and that it may develop to a civilization. Civilizations with similar technical achievements to ours could communicate with each other by radio waves up to distances of 100 light years.

Cocconi and Morrison [1959] first pointed out the possibility of communication from an extraterrestrial intelligence (ETI) and proposed to search for a signal. Independently, Drake et al. attempted to receive signals from possible civilizations on nearby stars. Similar attempts have been made at other observatories since then, and these works have been reviewed by Sagan and Drake [1975]. The first "aimed" signal was sent to space from the Arecibo Observatory in November 1974 [NAIO, 1975].

Using state-of-the-art technology it is feasible to detect radio signals arriving at the Earth from other civilizations in the galaxy. Such a programme is called SETI (search for extraterrestrial intelligence).

There are presently several SETI programmes in progress [Sagan and Drake, 1975]. These include the following:

1) Bridle and Feldman, at Algonquin Radio Observatory in Canada, are searching nearby stars at 22.2 GHz, near the H_2O line.

2) Dixon and Cole, at the Ohio State University Radio Observatory, are making an all sky survey near the 1.4 GHz hydrogen line [Dixon, Cole and Kraus, 1977]. This survey has been continuously in progress for three years.

3) Drake and Sagan, using the Arecibo Observatory in Puerto Rico, are observing several nearby galaxies at 1 420, 1 653, and 2 380 MHz.

4) Kardashev, using the Eurasian Network, in the USSR, is searching for pulsed signals, with hemispherical coverage.

5) Troitsky, using the Eurasian Network, is searching for pulsed signals in an all sky survey at 1.9, 1.0 and 0.6 GHz.

6) Zuckerman and Palmer, using the NRAO Observatory in Greenbank, are searching nearby F, G, and K type stars near 1 420 MHz.

7) The United States National Aeronautics and Space Agency is currently conducting a search near 1.5 GHz.

1.2 Average distance between civilizations

Average distance between the civilizations is inversely proportional to the cube root of the space density of the civilizations, which is proportional to their average life.

For the civilized life within 100 light years to have a high probability, one must assume an average life of at least 10^7 years.

1.3 Other civilizations

Some experimenters may assume that the other civilization is superior to ours, based on the following argument. We have been able to communicate with an equivalent civilization by radio waves only during the last 30 years. Consequently, if they can communicate, but are inferior to us, the state of development of the other civilization must be within the same interval of 30 years. As 30 years is an extremely short time compared with the time scale of evolution of life, the probability that this would occur is very low. Similar arguments show that they are unlikely to be only slightly superior to us. Arguments in the previous section also require an average life of communicating civilizations of the order of 10^7 years. We conclude therefore that the other civilizations are probably superior to ours.

It is also possible that such civilizations have formed a community through radio communications and that they have been continuously sending signals to suggest that we join the community.

1.4 Consequences of success

Interstellar communication is merely hypothetical before the first contact is made. However, as soon as a contact is established, practical implications to us may be significant. The large capacity communication following the first contact may contain information far superior to our knowledge.

1.5 Types of stars to be searched

Stars which are similar to the sun may have planets suitable for life similar to that on the Earth. Such stars have surface temperatures of 4500 to 6500 K and luminosity of 0.3 to 3 of the sun, and are known as main sequence stars with spectral types of F, G and K [Sagan, 1973].

2. Characteristics of the signal

2.1 Transmitted signal

Nothing can be known with certainty about radio signals transmitted by an extraterrestrial, intelligent society (ETI). However, it can at least be assumed that the signals may have any of the characteristics currently known to human science. Thus, the transmitted signals may have any carrier frequency, modulation, e.i.r.p., or polarization. In addition, since the location of the source is unknown, the nature and magnitude of the source Doppler frequency drift is also unknown.

Using practicable terrestrial facilities (100 m diameter antennae, 1 MW transmitter and a 20 K receiver) communication could be maintained at a rate of 10 bps within a distance of 100 light years.

The United States routinely generates an e.i.r.p. of 130 dBW near 2 GHz during planetary radar experiments at Arecibo, Puerto Rico. An e.i.r.p. this high or higher is possible for a signal from an ETI.

Some experts believe that the signal would be from a point source, polarized, variable with time, and have a narrow bandwidth (of the order of 10 Hz). A very simple form of modulation, coding etc., might be used to make processing at this end simple. This assumes that the signal is intended for reception by a civilization other than the one that is transmitting.

In this case, initial contact would be through a beacon signal to establish a large capacity communication. This signal would possibly contain minimum information to let us know their existence, factors related to the communications following, etc. Subjects such as mathematics, geometry, physics etc., might be also contained in order to help us to understand systems of modulation, coding, grammar, etc.

2.2 Propagation considerations

2.2.1 Interstellar medium

The interstellar medium is a magneto-ionic plasma with non-uniform characteristics that vary with time, distance and direction. The medium is inhomogeneous, anisotropic, and dispersive. The result of these properties is a change in the characteristics of the transmitted signal as it passes through the medium.

Changes in the polarization and spectrum of the transmitted signal are especially important.

Faraday rotation

One effect of the medium will be Faraday rotation of the polarization vector. This may be in either direction depending on the characteristics of the path. Rotations as great as 10 radians or more have been observed at a frequency of 1 420 MHz [Whiteoak, 1974].

Since the medium is anisotropic, there may also be some conversion between linear and circular polarization.

Spectral changes

The interstellar medium causes changes in the spectrum of a signal transversing it. If the signal is unmodulated, the medium broadens the signal spectrum; if the signal is modulated, the medium broadens and distorts the modulation spectrum. The effect is to limit both the maximum and minimum bandwidths for observation of a coherent signal.

The best available data concerning limitation of the maximum bandwidth comes from the measurement of pulsars [Lee and Jokipii, 1976]. These measurements indicate that the maximum coherent bandwidth at 1 GHz is approximately 2 000 Hz over a path length of about 10^3 light years. The bandwidth appears to increase with frequency and decrease with distance.

Pulsar observations have also shown the minimum coherent bandwidth to be on the order of 10^{-2} to 10^{-3} Hz.

2.2.2 Atmosphere of the Earth

Attenuation

For clear weather conditions, atmospheric attenuation is a significant factor above about 20 HGz (Report 233-3, Rev. 76, p. 69). When rain is taken into account, attenuation becomes more significant at frequencies greater than 3 GHz. Atmospheric attenuation is shown in Fig. 1, based on data from Report 564 [Geneva, 1974].

Polarization and spectrum

The ionosphere will affect the polarization and spectrum of the signal in transit. The effect is expected to be small compared to the effect of the interstellar medium.

2.3 Arriving signal

2.3.1 Noise

Signals from an ETI will arrive at the Earth mixed with background radiation.

Extraterrestrial

Outside the atmosphere of the Earth, background noise consists primarily of three components. These are: the 2.7 K black body isotropic radiation that fills the universe, galactic noise and quantum noise (caused by fluctuations in the rate of arrival of RF quanta at the receiver) [Oliver, 1973a]. These noise contributors are shown in Fig. 2.

Galactic noise temperature decreases steeply with increasing frequency. It is directional in nature and is maximum at the galactic equator, diminishing rapidly at greater galactic latitudes. At a latitude of $\pm 5^{\circ}$ it is about twice the amount near the poles.

The black body background noise temperature is constant in the microwave region at frequencies less than about 60 GHz. At higher frequencies it diminishes. This black body background radiation is isotropic.

Quantum noise temperature increases with increasing frequency, but is significantly less than the other contributors below 60 GHz. It is also independent of direction.

The spectra of emissions of molecules and free radicals contained in interstellar space modifies the noise at certain frequencies. Hydrogen, hydroxyl, and formaldehyde are examples.

Near 1 420 MHz the emissions of neutral hydrogen cause the noise to be above the background.

The OH lines, through maser action, are very narrow and are often variable with time [Weaver et al. 1965].

The H_2CO line, through the anti-maser effect, exhibits a negative excitation temperature and, consequently, the background radiation temperature in the direction of dark clouds is below the 3K background temperature. On the basis of current knowledge this is a unique situation [Palmer, et al. 1969].

The resultant total sky noise temperature outside the atmosphere of the Earth is shown in Fig. 3. As can be seen, there is a broad minimum between approximately 1 and 100 GHz, which is called the free-space microwave window.

From the atmosphere of the Earth

The contributions of the atmosphere to sky noise are discussed in detail in Report 234-3 [Geneva, 1974], Vol. V, page 93. The noise contribution of rain is also discussed in Report 564 [Geneva, 1974], Vol. II, page 253. The effect of the atmosphere is to reduce the frequency range of the sky noise temperature minimum.

Total sky noise

When noise effects of the atmosphere are combined with extraterrestrial noise, the frequency of the noise minimum is reduced to the range between 1 and 10 GHz for clear weather, and 1 to 3 GHz for precipitation in rain climate 4 (as an example), considering rainfall rates not exceeded for more than 0.01% for an average year. See Report 234-3. The resultant total sky noise temperature is shown in Fig. 4.

2.3.2 Received flux-density

Signals will have travelled over interstellar distances, and therefore could be at extremely low power flux-density levels upon arrival at the Earth. A graph of power flux-density at the Earth versus e.i.r.p. for sources at various distances is shown in Fig. 5. The distance to the nearest star is about 4 light years, the diameter of the galaxy is about 100 thousand light years, and the distance to nearby galaxies is around 10^7 light years.

2.3.3 Frequency

The frequency and the time rate of change of frequency of the signals arriving at the Earth are unknown.

Frequency drift considerations

If the transmitting source is on a planet of some other star, then relative motion between the source and our search system can occur because of (a) radial velocity of the other star with respect to the sun, (b) orbital velocity of the Earth and of the other planet, and (c) the rotation of the Earth and the other planet.

The first of these leads to a Doppler shift which is for all practical purposes constant over the interval of time of the search.

Planetary orbital motion and rotation produce Doppler frequency shifts that vary nearly sinusoidally with time (Doppler drift).

There could be, for example, four sinusoidal components to the received signal. These are due to the orbital and diurnal motions of the Earth and the transmitter. The magnitude of the Doppler drift is directly proportional to the transmitted frequency.

The effects of Earth motions can be calculated and allowance can be made for its motion.

Frequency drift can also be caused by instabilities in the receiver and in the transmitter.

2.3.4 Modulation

The arriving signal may be unmodulated, or have any of the known modulation types.

If the source modulation bandwidth exceeds the coherent bandwidth of the medium, the modulation will be distorted or destroyed.

Spectral lines in the transmitted signal will arrive with some broadening caused by the propagation medium, but essentially intact.

2.3.5 Polarization

The arriving signal may have any known linear, circular, or elliptical polarization with any orientation.

2.3.6 Direction

Because the source location is unknown, the direction of the arrival is also unknown.

2.3.7 Time of arrival

Signals may be intermittent.

3. Preferred frequency bands

3.1 Receiving sensitivity considerations

Noise

It may be assumed by some experimenters that there is no prior knowledge of the characteristics of the signal sought, or of the signal source distance. This leads to the conclusion that the signal power flux-density from an extraterrestrial source may be very small, and therefore the search must be made in a portion of the electromagnetic spectrum where physical factors allow maximum sensitivity.

Considerations of sensitivity make background radio noise a strong factor in the determination of a preferred frequency band.

Accommodation of the need for maximum available search time means that atmospheric precipitation effects on system sensitivity should also be taken into account for a terrestrial system.

Considering the discussions of §§ 2.2.1 and 2.3.1, the smallest background noise is obtained between 1 and 100 GHz for a spaceborne system. For a terrestrial system, noise is least between 1 and 10 GHz when only clear weather is considered. If rain effects are also taken into account, the noise minimum is reduced to between 1 and 3 GHz.

Frequency drift

Maximum receiver sensitivity is obtained by using very narrowband receiving channels; but, in order for the receiver output to develop fully, the signal must remain in the receiver bandwidth for a time somewhat longer than the reciprocal of the receiver bandwidth. For an arbitrary Doppler frequency drift rate and assuming matched filter detection, it can be shown that [Oliver, 1973b]

$$B = Af_T^{1/2} \tag{1}$$

where

f_T is the transmitted frequency;

B is the minimum channel bandwidth for full response to the drifting signal;

A is a constant of proportionality.

Thus, in order to allow minimum channel bandwidth and thehighest sensitivity, it is necessary to confine doppler drift to a single channel during the integration time. For the unknown components this can only be done by receiving at the lowest practical frequency.

Maximum sensitivity

The combined effects of noise and doppler drift on total system sensitivity can be expressed as a figure-of-merit (F_m):

$$F_m = T_s f_S^{1/2} \tag{2}$$

where:

T_s is the total equivalent sky noise temperature;

f_S is the frequency being searched.

The best sensitivity is indicated by the smallest figure-of-merit.

This figure-of-merit is shown in Fig. 6 both with and without the atmospheric noise contribution. The maximum sensitivity is obtained in either case in the 1-3 GHz region. For an earth-based system the maximum sensitivity is near 1.5 GHz.

3.2 Spectral line considerations

Experimenters may also assume that an ETI may place importance on some of the same molecular emission frequencies that are considered important by human science. In this case it may be desirable to search near specific spectral lines.

Cocconi and Morrison [1959] proposed using a frequency of 1 420 MHz (the neutral hydrogen spectral line) which was the only known spectral line at that time. Subsequently, a number of molecular spectral lines have been detected in the microwave and mm wave region, and the selection of the frequency band must be reviewed.

In selecting the line, the following characteristics should be considered:

(a) The frequency must be known as unique or particular to both parties.

(b) It must have an intrinsic advantage that is known to both parties.

(c) There must be a high probability of receiving a signal by chance.

The hydroxyl lines

The lambda-type doubling lines of OH radicals (1 612, 1 665, 1 667 and 1 720 MHz) and K-type doubling line of formaldehyde (H_2CO, 4 830 MHz) have been suggested as being more suitable than the other molecular lines [Morimoto et al. 1977].

A frequency 1 666 MHz would be a good choice as it is in the middle of the main components (1 665 and 1 667 MHz), clear from the OH lines, but close enough to be monitored for the lines. Furthermore the lines are regularly monitored with a high frequency resolution in many objects from many observatories, and there is a high possibility that a narrowband artificial signal would be received by chance. This fact must be known to the superior extraterrestrial civilizations.

The formaldehyde line

The formaldehyde molecule at 4 830 MHz exhibits anti-maser action as mentioned in § 2.3.1. A civilization on a star seen by us in front of a formaldehyde cloud could realize this situation and tend to send us signals at this frequency.

The hydrogen line

Observation directly at the hydrogen frequency would suffer from a high background noise. Observations near this line may be profitable.

The "water hole"

The region between the hydrogen line near 1 420 MHz and the set of hydroxyl lines near 1 650 MHz has been called the "water hole." Water based life forms may see this region as significant. These lines of the dissociation products of water are landmarks in that region of the microwave window from 1 to 2 GHz which physical considerations have indicated as most sensitive for the search.

Others

There are, of course, other molecular interstellar lines which might have importance, such as the water line at 22.23 GHz, or the formaldehyde line at 14.5 GHz. However, sky noise and doppler drift considerations make a search near these other molecular transition frequencies less attractive.

3.3 Summary

Considerations of earth-based SETI system sensitivity lead to a preferred frequency near 1.5 GHz. If importance is attached to the hydrogen and hydroxyl lines this frequency region becomes even more attractive.

However, if observers assume that an ETI may attach significance to the noise reducing properties of formaldehyde clouds, then 4 830 MHz is also a preferred frequency.

The bands between the OH lines near 1 650 MHz are considered of secondary preference.

The above considerations are not complete, and a SETI may be carried on at any frequency. Although SETI is a type of space reserach, it is often performed in the same frequency bands as radio astronomy because of spectral line considerations and the relative lack of radio interference.

4. Search system characteristics and requirements

The factors of §§ 1 and 2 lead to the following requirements on the search system:

1) Maximum practicable sensitivity

2) Capable of receiving from any direction

3) Capable of receiving any polarization

4) Capable of searching a large frequency band with very narrowband frequency resolution

5) Nearly continuous operation and minimum practical total search time.

4.1 Sensitivity

Under certain conditions the minimum detectable power of a signal is given by [Oliver, 1971]:

$$P = kT\frac{1 + \sqrt{1 + Bt}}{t} \qquad (3)$$

where k is the Boltzmann constant, B is the bandwidth of the receiver channel in Hertz, T is the equivalent system noise temperature in Kelvins, and t is the integration time in seconds.

Maximum sensitivity to a coherent signal will be achieved by a combination of low noise temperature, narrowband frequency resolution, long integration times, and antennae with large collecting area.

Because of the unknowns in the possible modulation, best sensitivity and maximum likelihood of detection can be achieved with a system designed to search for and detect single spectral lines (carriers).

4.2 Bandwidth

For matched filter detection, the flux-density detectable by a search system is proportional to the search bandwidth, maximum sensitivity being achieved when the search bandwidth matches the signal bandwidth (which may be as narrow as the interstellar medium permits). At the same time it is necessary to search over a very wide frequency range. Many spectrum analyzers sweep a narrowband receiver across the band of interest. This procedure is not useful for SETI. What is necessary is to simultaneously search many adjacent narrow channels, perhaps as many as 10^9 channels having bandwidths as small as 0.01 Hz. This simultaneous search of many adjacent channels allows the system to reach maximum sensitivity, while at the same time being able to detect a signal which is drifting in frequency due to doppler effect, and also reducing the time necessary to search a very wide frequency band.

4.3 Operation

The SETI search may take decades or longer. The number of directions to be searched is very large; the possibility of intermittent transmissions will require long observation times in each direction; and integration of the receiver output to increase system sensitivity will make each observation relatively lengthy.

Steps must be taken to minimize the total search time, e.g., the system should be automated and designed for nearly continuous all-weather operation.

When a signal is detected, it must be immediately examined to determine if it is an ETI signal. The search system must be sufficiently reliable to do this at any time.

As this work may take a long period without practical results, a special arrangement for continuing the search may be necessary.

4.4 Location

For ground based operations, requirements are similar to those for radio-astronomy services, i.e., a place surrounded by mountains, far from large cities, etc.

In space, the far side of the moon offers freedom from man-made interference and atmospheric emission. The Lagrangian colinear equilibrium point (Fig. 7) in the earth-moon system (60,000 km behind the moon, where the gravitational force is balanced with the centrifugal force so that the observing spacecraft remains eclipsed from the Earth and well removed from the thermal emission from the lunar surface.

4.5 The search programme

SETI programmes to date have examined only a few targets and directions; they have searched over a relatively narrow band, and their sensitivity has been poor compared to the state-of-the-art.

Systems now performing searches are using existing radio telescopes.

Future programmes will have radio antennae dedicated to SETI, will examine many more targets and directions, will search much wider bandwidths, and will be considerably more sensitive than 1977 systems.

To gain increased sensitivity they may: use more sophisticated data analysis techniques; reduce the system noise temperature; reduce the single channel bandwidth; and use larger, more efficient antennae.

Future systems may search the entire sky, or they may look at individual stars. If observers desire to consider the background noise reduction caused by formaldehyde there are numerous stars which can be seen against the background of many dark clouds that cover a substantial fraction of the sky. Table I lists a sample of such stars, which are single stars, brighter than 8th magnitude, of spectral types of F, G, and K and are in front of dark clouds. The column 1 gives names of stars in the catalogue number in the AGK 3 catalogue, column 2, brightness in magnitude, column 3, spectral types, and columns 4 and 5 give equatorial coordinates of the stars.

Optimum search method

The above considerations are based on many debatable assumptions and there are many possible alternative systems; consequently, no single method can be specified as the optimum and it may be best to try several.

4.6 Search system design

Several earth-based SETI systems are currently being planned in the USA with capabilities and system parameters as outlined in Table II.

4.7 Summary

While the exact source direction, frequency, signal polarization, flux-density, modulation and doppler shift are unknown, the considerations discussed in this Report lead to the following conclusions:

1) The search system must be very sensitive and have the lowest possible noise temperature;

2) the system must be capable of searching large frequency bands continuously with narrowband frequency resolution;

3) the system must be capable of looking in any direction;

4) the system must sense all possible polarizations.

5. Interference considerations

A system for a search for extraterrestrial intelligence (SETI) will be a receive only system and will not cause interference. The SETI system will be characterized by low system noise temperature, antennae with a large capture area, wide instantaneous bandwidth with narrowband resolution, with signal integration and quasi-matched filter detection techniques. The receivers will have the highest practical sensitivity in order to detect the anticipated low power flux-density signals from an ETI. Hence, a SETI system will be very vulnerable to interference.

The system discussed in this Section is considered to be operating near 1.5 GHz. Other frequencies will also be used, especially ones used by radio astronomy.

5.1 SETI station factors pertinent to sharing

The SETI parameters considered in this Report are taken from Table II. Other values may be examined by using the information from that and this Report.

The minimum detectable signal for this system is -226 dBW in 300 Hz, or -272 dBW/m^2 in 300 Hz using an antenna of 300 metre diameter (such as the radio telescope at Arecibo, Puerto Rico).

5.1.1 Maser saturation

Systems under consideration in the USA envision the use of a maser to achieve the required low noise temperature. Typical masers exhibit nonlinear saturation effects when the total receiver input power in the maser passband exceeds about -120 dBW. Systems used for SETI are considering passbands up to 300 MHz.

5.1.2 Receiver saturation

The signal processor will have a finite linear dynamic range. Thus a sufficiently strong receiver input signal in a single channel will cause saturation. This will in turn cause the generation of spurious signals that will also affect nearby channels.

5.1.3 Minimum detectable signal degradation

A wideband, manmade signal at a level about 10 dB below the minimum detectable signal in each channel will appear to the signal processor as an increase in noise over a number of channels. This will cause a degradation of the receiver performance of about 0.4 dB in those channels.

5.1.4 False signal detection

A manmade coherent signal at or slightly below the minimum detectable level in a channel will be observed and cause a false alarm. This will in turn require observations sufficient to establish the detection as a false alarm. It will also mask a weaker ETI signal that may be present.

5.2 Interference protection

Based on § 5.1, reasonable protection will be afforded if the power in a single channel is about 10 dB below the minimum detectable signal. This requires that the power spectral density of wideband interference or the total power of CW interference in any single band and all sets of bands 1 Hz wide does not exceed -260 dBW/Hz referenced to the input terminals of the receiver.

Because of the possible fleeting nature of ETI signals, interference should not exceed an aggregate of 5 minutes per day. For the reasons discussed in Report 219-2 (New Delhi, 1970) this should be taken as 0.001% of the time for protection from terrestrial transmitters.

Near 1.5 GHz this protection will be afforded for a power spectral flux-density of -306 dBW/m^2 Hz on the boresight of a 300 meter antenna, or -235 dBW/m^2 Hz for an isotropic receiving antenna.

5.3 Sharing considerations

5.3.1 Line-of-sight paths

Consider the conservative case of a transmitting station with 1 dBW of power, a 0 dBi antenna, and the modulation uniformly spread over 1 MHz. The station would exceed the interfering power spectral flux-density given in § 5.2 for an isotropic receiving antenna at a range of approximately 160,000 km.

This leads to the conclusion that sharing with spaceborne stations is not practicable. Sharing with airborne stations above the horizon of a SETI station is also not practicable.

5.3.2 Over the horizon paths

Sharing with most ground-based transmitters appears feasible if an appropriate coordination procedure is adopted.

Sharing with high powered earth-based systems such as troposcatter and radiolocation may be difficult.

This sharing case requires further study.

5.3.3 Reflection from spaceborne objects

The reflection of terrestrial signals from spaceborne objects near earth may also cause a problem. This case needs further study.

5.4 Summary

Frequency sharing between a receiving SETI system and spaceborne or line-of-sight airborne transmitting systems is not feasible.

Frequency sharing between a receiving SETI system and earth-based transmitting systems is probably feasible in most cases with appropriate coordination.

6. Other cases

Although the above considerations are based on a number of reasonable assumptions. there are also several reasonable objections. Three important ones are outlined.

(a) There is an argument that transmitting a radio signal may be more efficient than just searching for signal. It was also pointed out that the first artificial radio signal occurred 60 years ago and this signal has already travelled 60 light years, and may be detected by a civilization within this distance. However, such unaimed signal at a distance of 60 light years is very weak and would be difficult to detect by other civilizations.

If the communicating civilization is regarded as far superior to ours, transmission of an aimed signal must not be attempted without due consideration of unexpected problems.

(b) Bracewell [1960] considered the possibility of a first contact through a space probe from the other civilization. Generally a space probe could be considered as uneconomical and therefore unlikely. However, several variations are possible and a more careful study is warranted.

(c) Morimoto [1967] has suggested a gaseous creature, which causes a population inversion in certain molecules in its body and sends and receives radio signals through maser action.

These possibilities are ignored in the present Report.

7. Conclusions

7.1 Preferred frequency bands

If experimenters assume that there is no a priori knowledge of the ETI signal, then considerations of maximum sensitivity and maximum search bandwidth lead to the conclusion that the region near 1.5 GHz having a bandwidth of several hundred MHz is a preferred region for SETI examination. Because of the possible significance that other water based life may attach to the spectral lines of hydrogen and hydroxyl radical, it is desirable that the band include these spectral lines. It should be defined so as to allow a reasonable search bandwidth on either side of these lines.

Consideration of these factors establishes a preferred frequency band several hundred MHz wide, near 1.5 GHz.

However, if observers assume that the ETI has knowledge of the location characteristics of formaldehyde clouds, and is considering their shielding in transmitting to the Earth, then a narrowband of frequencies centered at the 4 830 MHz formaldehyde line would be a preferred frequency.

Other frequencies will also be used, especially those currently used by radio astronomy. As the determination of optimum frequencies continues, other preferred bands will emerge.

7.2 Interference considerations

Some currently planned SETI systems will need protection at the receiving level of -235 dBW/m^{2}Hz near 1.5 GHz, if the antenna is assumed to have isotropic gain away from the main beam.

Some other SETI systems will use frequencies also used by radio astronomy. The protection currently considered appropriate in Report 224 will be adequate for their protection.

As the SETI systems and frequencies become better defined, additional protection criteria will emerge.

REFERENCES

AGK3, [1975], Star catalogue of positions and proper motions north of -2.5 Declination. Hamburger Sternwarte.

BRACEWELL, R.N. [1960], Communication from superior galactic communities. Nature, 186, 670.

COCCONI, G. and MORRISON, P. [1959], Searching for interstellar communications. Nature. 184, 884.

MORIMOTO, M. [1967], Dark clouds and signal from OZ (in Japanese). Shizen, 22, 17.

MORIMOTO, M., HIRABAYASHI, H. and JUGAKU, J., Preferred frequency for interstellar communications (in prep). Nature.

PALMER, P., ZUCKERMAN,B,B., BUHL, D. and SNYDER, L.E. [1969], Formaldehyde absorption in dark nebulae. Astrophys. Journal, 156, L47.

SAGAN, C. and DRAKE, F. [1975], The search for extraterrestrial intelligence. Scientific American, 232, 80.

SAGAN, C. ed. [1973], Communication with extraterrestrial intelligence. The MIT Press.

The Staff of the N.A.I.O. [1975], Arecibo Message of November 24. ICARUS, 26, 80.

WEAVER, H., WILLIAMS, D. R. W., DIETER, N. H. and LUM, W. T. [1965], Observations of a strong unidentified microwave line and emission from OH molecule. Nature, 208, 29.

COCHRANE, W. T., et al. [1967], What is the Fast Fourier Transform? Proc. of the IEEE, 5, 1964.

DIXON, R. W., COLE, D. M., and KRAUS, J. D., [1977], A modest all-sky search for narrowband radio radiation near the 21 cm hydrogen line. ICARUS, preprint.

LEE, L. C., and JOKIPII, J. R., [1976], The irregularity spectrum in instellar space. ASTROPHYS. J., 206, 735.

OLIVER, B. M., BILLINGHAM, J. et al. [1971], Project cyclops, NASA/Ames Research Centre, Code LT, Moffett Field, California 94035, USA.

SAGAN, C. and DRAKE, F. [May, 1975], The search for extraterrestrial intelligence, Scientific American 232, 83.

WHITEOAK, J. B. [1974], The observed characteristics of the local magnetic field, in Galactic Radio Astronomy, IAU Symposium No. 60, p. 137.

OLIVER, J. B., and BILLINGHAM, J. [1973a], Project Cyclops, NASA/Ames Research Centre, 40-41*.

OLIVER, J. B., and BILLINGHAM, J. [1973b], Project Cyclops, NASA/Ames Research Centre, 58*.

BIBLIOGRAPHY

SHKLOVSKI, I. S. and SAGAN, C. [1966], Intelligent Life in the Universe, Holden Day, Inc.

USSR [1966], Vnezemnye tsivilizatssi (Extraterrestrial civilizations) Proceedings of a conference, Byurakan, Arm. SSR.

Key words:

Extraterrestrial intelligence

Extraterrestrial life

Galactic civilizations

Interstellar communication

*Available for consultation in the CCIR Secretariat.

TABLE I.

A partial list of candidate stars for search at 4 830 MHz

AGK 3	No.	Magnitude	Type	Right Ascension	Declination (1950)
$+58^{o}$	161	2.8	F5	$0^{h}\ 6^{m}.5$	$+58^{o}\ 52'$
54	395	6.7	F5	4 5.4	54 42
38	491	6.2	F5	4 38.4	38 11
22	518	6.1	K0	5 16.3	22 3
+ 1	618	7.0	G5	5 44.0	+ 1 9
- 1	639	8.3	K2	5 46.2	- 1 48
+ 2	647	6.9	G0	5 47.9	+ 2 1
1	630	6.3	K0	5 49.8	1 51
6	643	7.7	G5	5 49.8	6 12
6	646	7.9	G0	5 50.8	6 15
3	717	7.9	K0	5 51.6	3 13
0	594	7.6	K0	5 52.5	0 58
2	684	7.3	K0	6 7.2	2 53
15	1946	7.9	G0	18 53.8	15 18
12	2040	8.4	G0	18 59.0	12 33
18	1834	8.0	K5	19 13.3	18 26
9	2447	6.7	F8	19 20.4	9 49
21	1990	8.6	G5	19 26.5	21 53
22	1969	8.7	K2	19 27.7	22 36
21	2002	7.6	G5	19 31.3	21 44
8	2614	7.6	K0	19 42.9	8 36
33	1851	7.9	K0	19 52.3	33 39
34	1945	8.0	K0	19 53.1	34 46
34	1947	5.0	K0	19 54.4	34 57
45	1663	8.1	K2	20 15.0	45 11
44	1848	7.8	G5	20 52.1	44 12
43	1905	7.9	K5	20 53.6	43 11
42	1952	7.5	K0	20 54.5	42 35
43	1910	7.9	K0	20 54.6	43 42
42	1956	8.4	K0	20 55.6	42 15
43	1923F	5.1	K5	21 3.1	43 44
57	1436	7.3	F8	21 23.6	57 52
54	1439	7.1	K0	21 39.1	54 39
49	1851	7.1	K0	21 40.8	49 22
60	1419	8.3	K0	21 50.9	60 6
56	1555	7.0	K2	22 14.6	56 58
59	1619	8.3	K0	23 25.9	59 47
58	1565	7.6	K0	23 28.9	58 53
59	1625	8.5	K5	23 29.4	59 11
58	1567	8.2	K2	23 29.4	58 49
60	1574	7.9	K0	23 31.0	60 8
60	1575	8.2	K0	23 31.3	60 12
+61	1471	7.6	K2	23 40.1	+61 24

TABLE II.

Earth-based SETI system characteristics

System noise temperature	10 K
Frequency resolution bandwidth	0.01 to 300 Hz
Integration time	10 to 10^5 sec
Number of adjacent frequency channels searched simultaneously	10^6 to 10^9
Bandwidth searched simultaneously	1 to 300 MHz
Tuneable frequency band capability	1 to 25 GHz
Polarizations received simultaneously	2 to 6
Minimum detectable signal power*	-226 dBW
Minimum detectable signal flux-density for a 300 m diameter antenna (e.g., Arecibo, Puerto Rico).*	-272 dBW/m^2

*These values are based on a 300 Hz resolution bandwidth, 10^4 second integration time, and 60% antenna aperture efficiency. Values for other system parameters may be calculated using equation (3) and the antenna capture area.

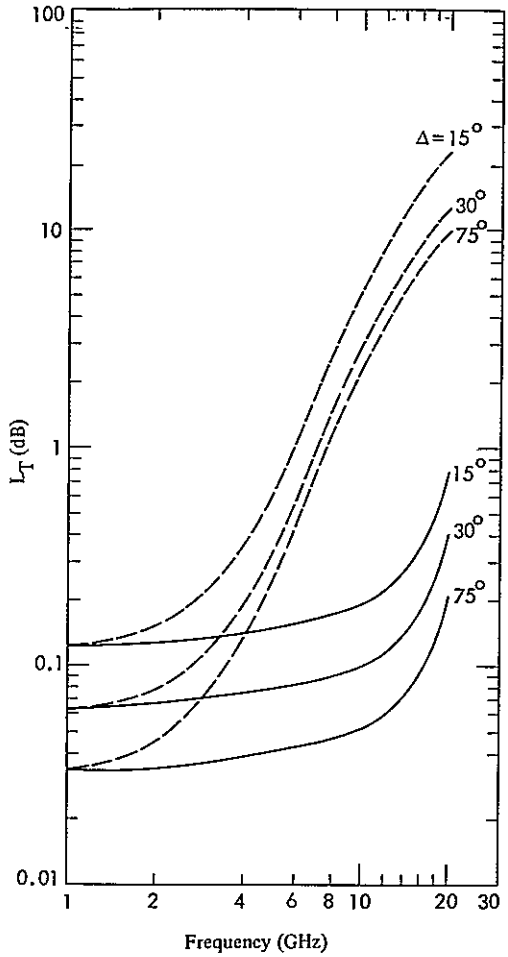

FIGURE 1

Space-to-earth attenuation (L_T)

— — — rain and atmosphere

——— clear weather, atmosphere only

Δ : elevation of earth station antenna

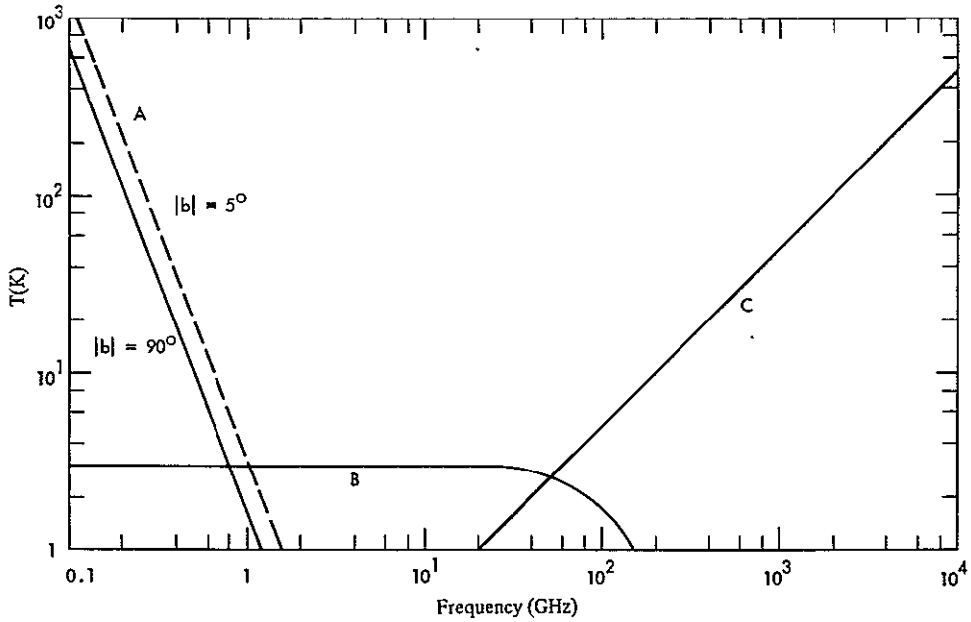

FIGURE 2

Background noise temperature (T) contributors outside the atmosphere of the earth

A: Galactic
B. Black body
C· Quantum effect
|b|: Absolute value of galactic latitude

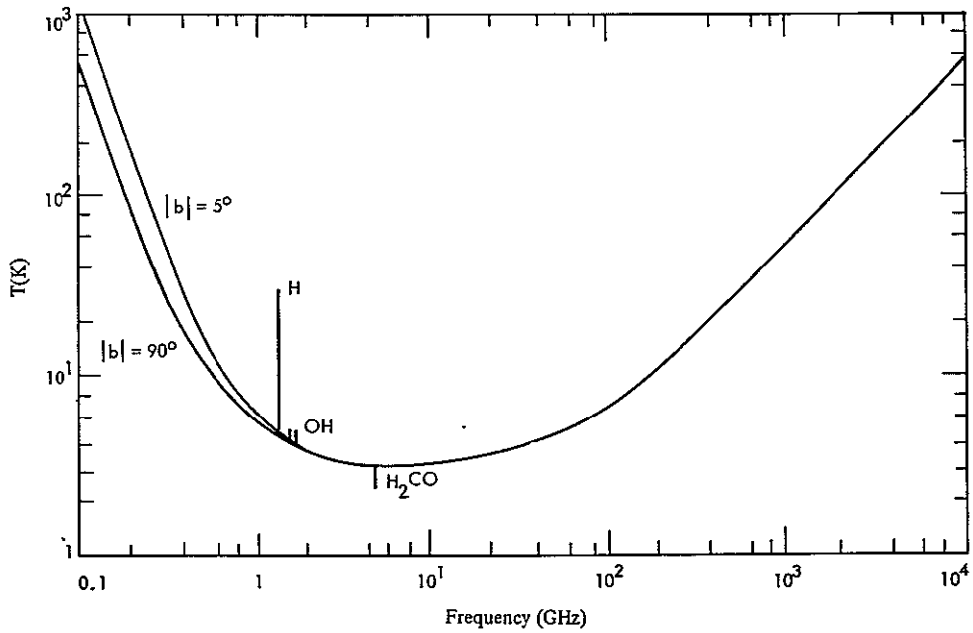

FIGURE 3

Total sky noise temperature (T) outside the atmosphere of the earth

| b|: the absolute value of the galactic latitude

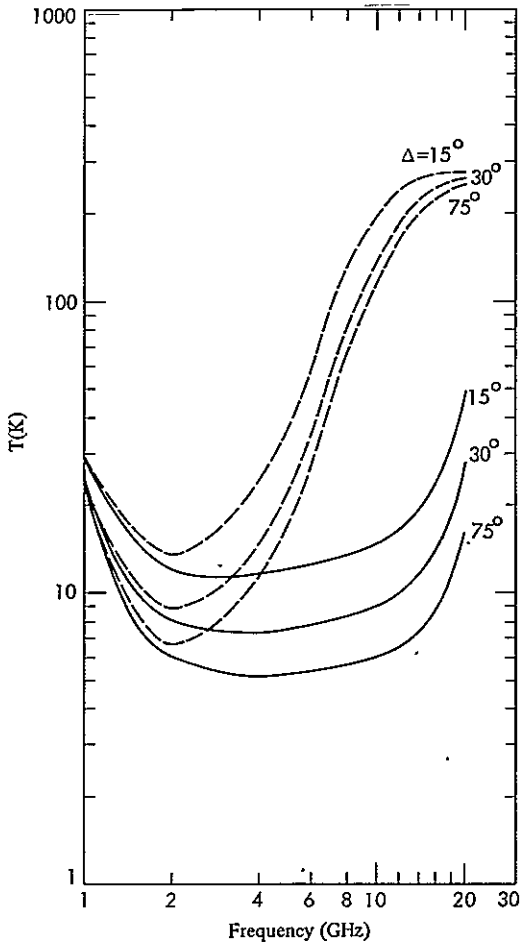

FIGURE 4

Total sky noise temperature (T) seen from the surface of the earth

— — — rain and atmosphere
——— clear weather, atmosphere only
Δ : elevation of earth station antenna

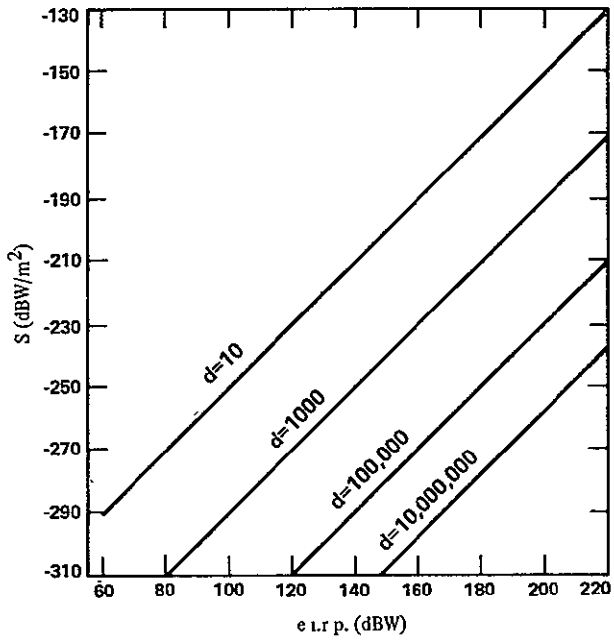

FIGURE 5

Received flux density vs. effective isotropically radiated power for unattenuated sources at various distances

d is distance in light years

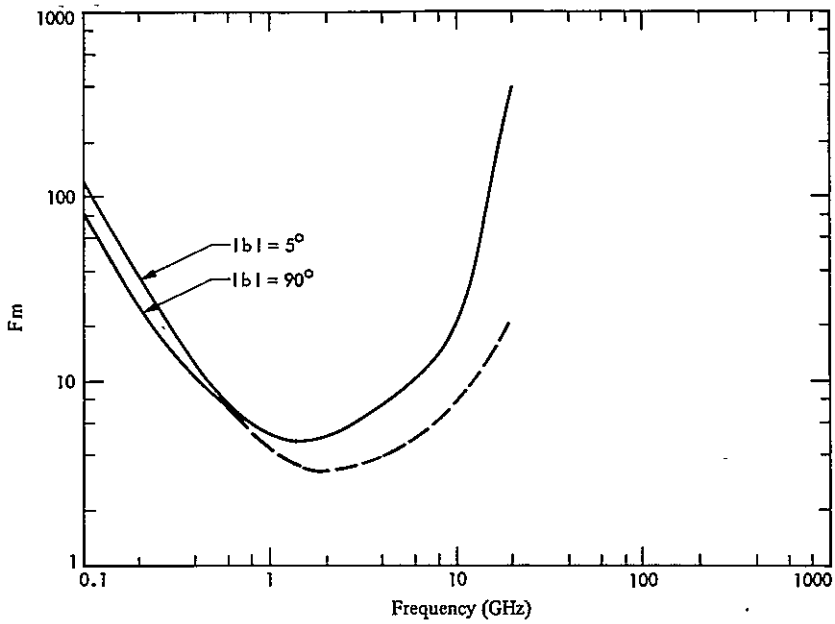

FIGURE 6

Combined noise and doppler drift system sensitivity figure of merit (F_m)

|b|: The absolute value of the galactic latitude

———Including zenith atmosphere noise
— — —Extraterrestrial noise only

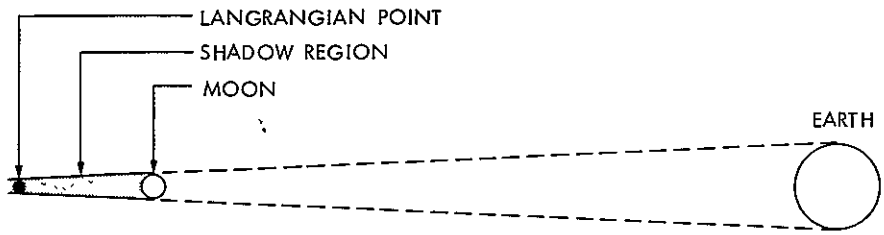

FIGURE 7

Lagrangian collinear equilibrium point

Analysis of Hydrogen Maser Frequency Drift Due to Possible Drifts in Load VSWR and Phase Angle of Reflection Coefficient

R. W. Beatty[1] and T. Y. Otoshi
Radio Frequency and Microwave Subsystems Section

Theoretical equations are derived for calculating the effects of load VSWR and reflection coefficient phase-angle drifts on hydrogen maser frequency stability. Sample calculations made for a typical JPL maser show that under special load conditions, a VSWR drift of 7.5×10^{-5}*/h or phase angle drift of 0.015 deg/h can produce a frequency drift of* $(10^{-14} f_0)$ *Hz/h where* f_0 *is the maser frequency of approximately* 1.42×10^9 *Hz.*

I. Introduction

It is well known that the frequency of a hydrogen maser oscillator can be pulled by the loading on its cavity. In applications requiring especially good frequency stability of a microwave source, such as in VLBI (very long baseline interferometry), very small load changes can cause significant frequency variations or drifts. In the following, this situation is analyzed in order to determine maximum drift sensitivity factors in a particular case.

II. Analysis

The case under consideration is shown in the simplified diagram of Fig. 1. The virtual resonant circuit located at the reference plane is shown.

The following familiar equations apply (for example, Ref. 1).

$$\left.\begin{aligned} Q_u &= \frac{\omega C}{G}, \quad Q_L = \frac{\omega C}{G + G_L} \\ Q_C &= \frac{\omega C}{G_L}, \quad Q_E = \frac{\omega C}{Y_0} \\ \beta &= \frac{Q_u}{Q_E} = \frac{Y_0}{G} \end{aligned}\right\} \tag{1}$$

where Q_U, Q_L are respectively the unloaded and loaded Q of the cavity; Q_C, Q_E are respectively the coupled and external Q; β is the coupling factor; and as shown in Fig. 1, G and G_L are the cavity and load conductances, respectively, and Y_0 is the characteristic admittance of the transmission line. It is

[1] R. W. Beatty is a consulting electronics engineer at 2110 Fourth Street, Boulder, CO 80302.

assumed in the above that Y_0 is positive real (i.e., $Y_0 = G_0$). Under loaded conditions, the resonant frequency of the cavity is given by

$$f_L = \frac{1}{2\pi\sqrt{L(C + C_L)}} \tag{2}$$

where the L and C are, respectively, the inductance and capacitance elements of the cavity shown in Fig. 1, and C_L is the equivalent load capacitance such that $B_L = \omega C_L$.

When the load is non-reactive ($B_L = 0$, $C_L = 0$), the cavity resonant frequency is

$$f = \frac{1}{2\pi\sqrt{LC}} \tag{3}$$

Let

$$\frac{\Delta f}{f} = \frac{f_L - f}{f} = \frac{f_L}{f} - 1 = \sqrt{\frac{C}{C + C_L}} - 1$$

$$\approx -\frac{C_L}{2C} \quad \text{for } C_L \ll C \tag{4}$$

Then

$$\frac{\Delta f}{f} \approx -\frac{\omega C_L}{2\omega C} = \frac{-B_L}{2\omega C} = -\frac{B_L}{Y_0}\left(\frac{1}{2Q_E}\right) \tag{5}$$

The hydrogen maser output frequency is pulled from the hydrogen line resonance by an amount that is proportional to the cavity resonant frequency shift according to the relationship (Ref. 2):

$$\Delta f_{osc} = \left(\frac{Q_L}{Q_H}\right)\Delta f \tag{6}$$

where Q_H is the Q of the hydrogen line. If we define f_0 to be the hydrogen line resonant frequency and $\delta = f - f_0$, then

$$\frac{\Delta f_{osc}}{f_0} = \left(\frac{Q_L}{Q_H}\right)\frac{\Delta f}{f\left(1 - \frac{\delta}{f}\right)} \approx \frac{Q_L}{Q_H}\left(\frac{\Delta f}{f}\right) \tag{7}$$

The above approximation is a good assumption because in practice, δ is normally made close to zero by adjustments made with a mechanical tuning slug in the cavity. From (1)

$$Q_L = \frac{\omega C}{G + G_L} = \frac{\frac{\omega C}{Y_0}}{\frac{G}{Y_0} + \frac{G_L}{Y_0}}$$

$$= \frac{Q_E}{\frac{1}{\beta} + \frac{G_L}{Y_0}} = \frac{\beta Q_E}{1 + \beta\left(\frac{G_L}{Y_0}\right)} \tag{8}$$

Then substitution into (7) gives

$$\frac{\Delta f_{osc}}{f_0} = \frac{\beta Q_E}{Q_H} \cdot \frac{1}{1 + \beta\left(\frac{G_L}{Y_0}\right)} \cdot \frac{\Delta f}{f} \tag{9}$$

Substitution of (5) gives

$$\frac{\Delta f_{osc}}{f_0} = -\frac{\beta}{2Q_H} \cdot \frac{\frac{B_L}{Y_0}}{1 + \beta\left(\frac{G_L}{Y_0}\right)} \tag{10}$$

We are interested in the rate of change or drift of $\Delta f_{osc}/f_0$ with time, or

$$\frac{d}{dt}\left(\frac{\Delta f_{osc}}{f_0}\right) = -\frac{\beta}{2Q_H} \cdot \frac{d}{dt}\left[\frac{\frac{B_L}{Y_0}}{1 + \beta\left(\frac{G_L}{Y_0}\right)}\right] \tag{11}$$

into which we may substitute

$$\frac{G_L}{Y_0} = \mathrm{Re}\left(\frac{Y_L}{Y_0}\right) = \mathrm{Re}\,\frac{1 - \Gamma_L}{1 + \Gamma_L} = \frac{1 - |\Gamma_L|^2}{|1 + \Gamma_L|^2} \tag{12}$$

$$\frac{B_L}{Y_0} = \text{Im}\left(\frac{Y_L}{Y_0}\right) = \text{Im}\,\frac{1-\Gamma_L}{1+\Gamma_L} = \frac{-2|\Gamma_L|\sin\psi_L}{|1+\Gamma_L|^2} \tag{13}$$

where $|\Gamma_L|$ and ψ_L are, respectively, the magnitude and phase angle of the load voltage reflection coefficient Γ_L at the reference plane shown in Fig. 1. For a general function F

$$\frac{d}{dt}F(|\Gamma_L|,\psi_L) = \left(\frac{\partial F}{\partial|\Gamma_L|}\right)\frac{d|\Gamma_L|}{dt} + \left(\frac{\partial F}{\partial\psi_L}\right)\frac{d\psi_L}{dt} \tag{14}$$

Hence

$$\frac{d}{dt}\left(\frac{\Delta f_{osc}}{f_0}\right) = \frac{\beta}{Q_H}\left\{\frac{[(1+\beta)-(1-\beta)|\Gamma_L|^2]\sin\psi_L}{[(1+\beta)+(1-\beta)|\Gamma_L|^2+2|\Gamma_L|\cos\psi_L]^2}\frac{d|\Gamma_L|}{dt}\right.$$

$$\left. + |\Gamma_L|\frac{[(1+\beta)+(1-\beta)|\Gamma_L|^2]\cos\psi_L + 2|\Gamma_L|}{[(1+\beta)+(1-\beta)|\Gamma_L|^2+2|\Gamma_L|\cos\psi_L]^2}\frac{d\psi_L}{dt}\right\} \tag{15}$$

This equation shows clearly the strong dependence of drift in oscillator frequency on the coupling coefficient β. If the cause of drift was in the load circuit or transmission line, a reduction in coupling would reduce it. Three other possible sources of oscillator frequency drift, not included in Eq. (15), are internal to the maser cavity and have been analyzed by Vessot, et al. (Ref. 2). Examination of their analysis reveals that if the cause of drift is in the maser cavity, a reduction in coupling would not lower the oscillator drift. Therefore, reduction of coupling factor would be one way to experimentally localize the source of frequency drift.

Although the differentiation was made with respect to time in Eq. (15), it could have been made with respect to any other variable. For example, if t in Eq. (15) were replaced by varactor voltage, one could arrive at the differential equation of the varactor tuning curve. Thus the above derivations have more general application than one might at first expect.

The load VSWR

$$\sigma_L = \frac{1+|\Gamma_L|}{1-|\Gamma_L|} \tag{16}$$

and

$$\frac{d\sigma_L}{dt} = \frac{2}{(1-|\Gamma_L|)^2}\frac{d|\Gamma_L|}{dt} \tag{17}$$

If $d\psi_L = 0$, the VSWR drift sensitivity is derived from Eq. (15) as

$$\frac{d\sigma_L}{dt} = \frac{2Q_H}{\beta}\left[\frac{d}{dt}\left(\frac{\Delta f_{osc}}{f_0}\right)\right] \times \frac{[(1+\beta)+(1-\beta)|\Gamma_L|^2+2|\Gamma_L|\cos\psi_L]^2}{(1-|\Gamma_L|)^2[(1+\beta)-(1-\beta)|\Gamma_L|^2]\sin\psi_L} \tag{18}$$

If $d\sigma_L = 0$, the load phase angle drift sensitivity is derived from Eq. (15) as

$$\frac{d\psi_L}{dt} = \frac{Q_H}{\beta|\Gamma_L|}\left[\frac{d}{dt}\left(\frac{\Delta f_{osc}}{f_0}\right)\right] \times \frac{[(1+\beta)+(1-\beta)|\Gamma_L|^2+2|\Gamma_L|\cos\psi_L]^2}{[(1+\beta)+(1-\beta)|\Gamma_L|^2]\cos\psi_L + 2|\Gamma_L|} \tag{19}$$

For a given $|\Gamma_L|$ or σ_L, the drift sensitivities depend upon ψ_L, and can vary over wide ranges. The maximum values with respect to ψ_L are of interest, and are as follows.

Case I. If $d\psi_L = 0$, maximum sensitivity to VSWR drift occurs according to Eq. (15) when $\psi_L = \pm\pi/2$, then

$$\frac{d\sigma_L}{dt} = \pm\frac{2Q_H}{\beta}\left[\frac{d}{dt}\left(\frac{\Delta f_{osc}}{f_0}\right)\right] \times \frac{[(1+\beta)+(1-\beta)|\Gamma_L|^2]^2}{(1-|\Gamma_L|)^2[(1+\beta)-(1-\beta)|\Gamma_L|^2]} \tag{20}$$

If we assume

$$Q_H = 10^9, \beta = 0.5, |\Gamma_L| = 0.1 \text{ and } \left[\frac{d}{dt}\left(\frac{\Delta f_{osc}}{f_0}\right)\right]$$

$$= 10^{-14}/\text{h},$$

then $d\sigma_L/dt = \pm 7.5 \times 10^{-5}$ per hour. Stated in another way, if $\psi_L = \pm \pi/2$, the VSWR drift must be kept to less than 7.5×10^{-5} per hour to keep the fractional frequency drift of the maser to less than 10^{-14}/h.

Case II. If $d|\Gamma_L| = 0$, maximum sensitivity to load phase angle drift occurs according to Eq. (15) when $\psi_L = \pi, 3\pi$, etc. Then

$$\frac{d\psi_L}{dt} = \frac{-Q_H}{\beta|\Gamma_L|}\left[\frac{d}{dt}\left(\frac{\Delta f_{osc}}{f_0}\right)\right][(1+\beta) + (1-\beta)|\Gamma_L|^2 - 2|\Gamma_L|] \tag{21}$$

If we assume the same parameters as in Case I, then,

$$\left|\frac{d\psi_L}{dt}\right| = 2.61 \times 10^{-4} \text{ radians per hour,}$$

or $\cong 0.015$ deg per hour

III. Application to Specific Circuit

As shown in Fig. 2, a maser is electronically tuned by varying the bias on a varactor connected to reflect energy back to the maser through a directional coupler. Assuming that the varactor is lossless, it acts like a variable capacitance. The locus of its reflection coefficient is a segment of the unit circle as shown in Fig. 3.

Both the signal incident upon the varactor and its reflected signal are attenuated or decoupled from the main line by the directional coupler. If the coupling is 10 dB for example, the varactor will produce $|\Gamma_L| = 0.1$ in the main line, and the locus of the varactor reflection coefficient variation will appear as in Fig. 4. Note that the small reflection coefficient circle has a radius of 0.1, and that it has rotated due to the delay along the path from the reference plane in the main line to the varactor. As we visualize the effect of different path lengths, we can see that for some lengths, the locus line cuts across lines of constant susceptance and causes maximum frequency change, but for other lengths the locus line produces no change in susceptance or in frequency. Normally, a path length is selected experimentally for which nearly maximum *change* in load susceptance is produced by the varactor. Any drift in varactor bias, for example, would produce a drift in phase ψ_L and the sensitivity would correspond to that calculated from Eq. (21).

IV. Effect of Fixed Reflections

In the absence of reflections other than from the varactor, the small reflection coefficient circle shown in Fig. 4 is centered on the origin. In this case, changes in varactor bias cause a shift in phase of the reflection coefficient in the main line, but no change in VSWR. Actually some small fixed reflections may be present from the receiver circuit, the directional coupler, and from connector discontinuities. These fixed reflections will displace the small reflection coefficient circle away from the origin as shown for example in Fig. 5 and result in the varactor changing the VSWR in the main line. It can be deduced from the diagram that a relatively small offset of this small reflection coefficient circle will not significantly affect the maximum drift sensitivities calculated from Eqs. (18) and (19). Nor will it greatly affect the values of ψ_L at which maximum effects occur. However, an offset comparable to radius of the reflection coefficient circle can have a significant effect. It can also be deduced from Fig. 5 that if a drift of magnitude or phase of the offset occurred, it would produce a drift of $|\Gamma_L|$ or ψ_L even though the varactor tuning was fixed.

Acknowledgement

Technical discussions with and contributions from Dr. W. Higa and R. Clauss are gratefully acknowledged.

References

1. Moreno, T., *Microwave Transmission Design Data,* Dover Publications, Inc., New York, 1948.

2. Vessot, R. F. C., et al., "Recent Developments Affecting the Hydrogen Maser as a Frequency Standard," in Final Report prepared for NASA entitled *Research and Development For a Ground-Based Hydrogen-Maser System,* Smithsonian Institute, Astrophysical Observatory, Cambridge, Mass., April 1972.

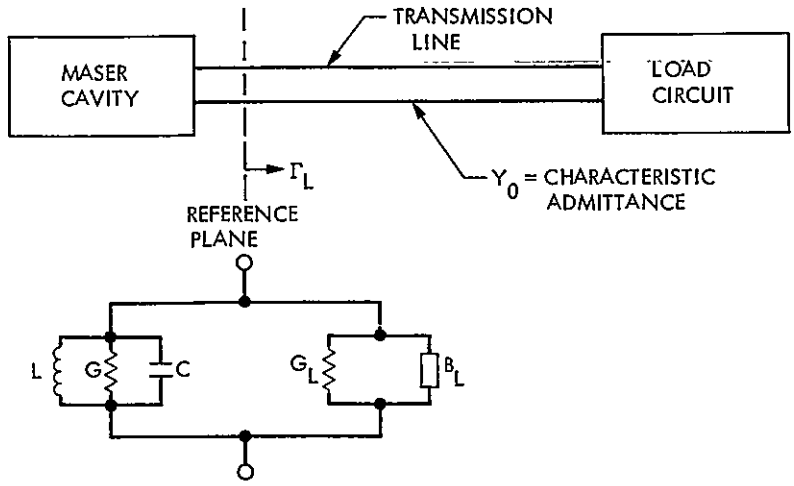

Fig. 1. Maser oscillator and load and equivalent lumped circuit

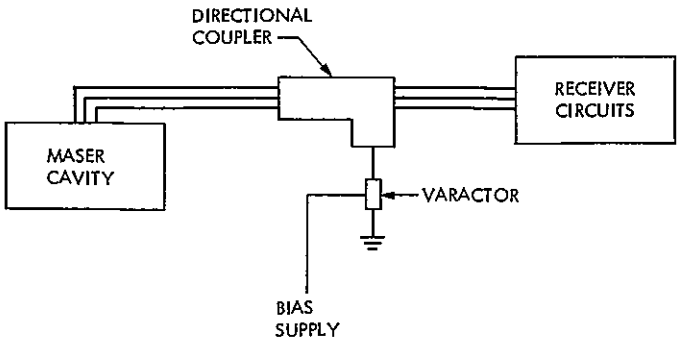

Fig. 2. Output circuit of maser oscillator

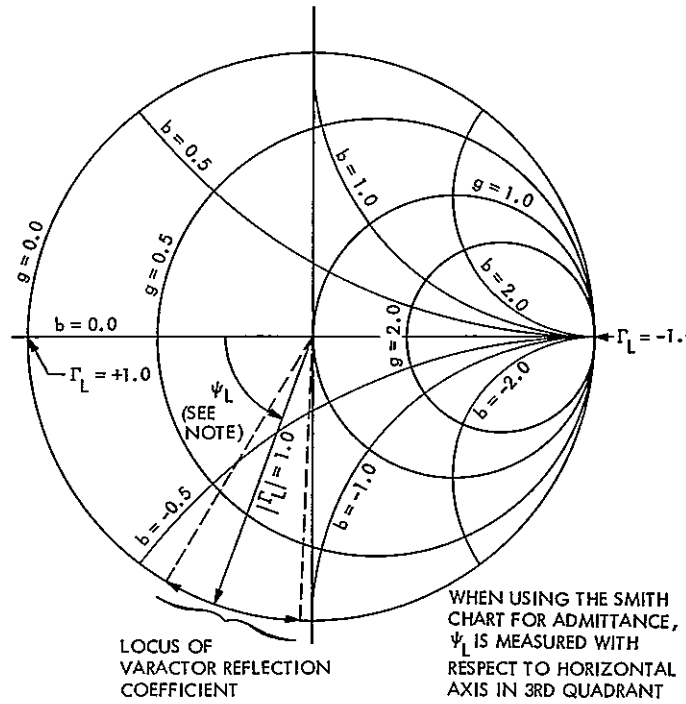

Fig. 3. Admittance and voltage reflection coefficient diagram for varactor

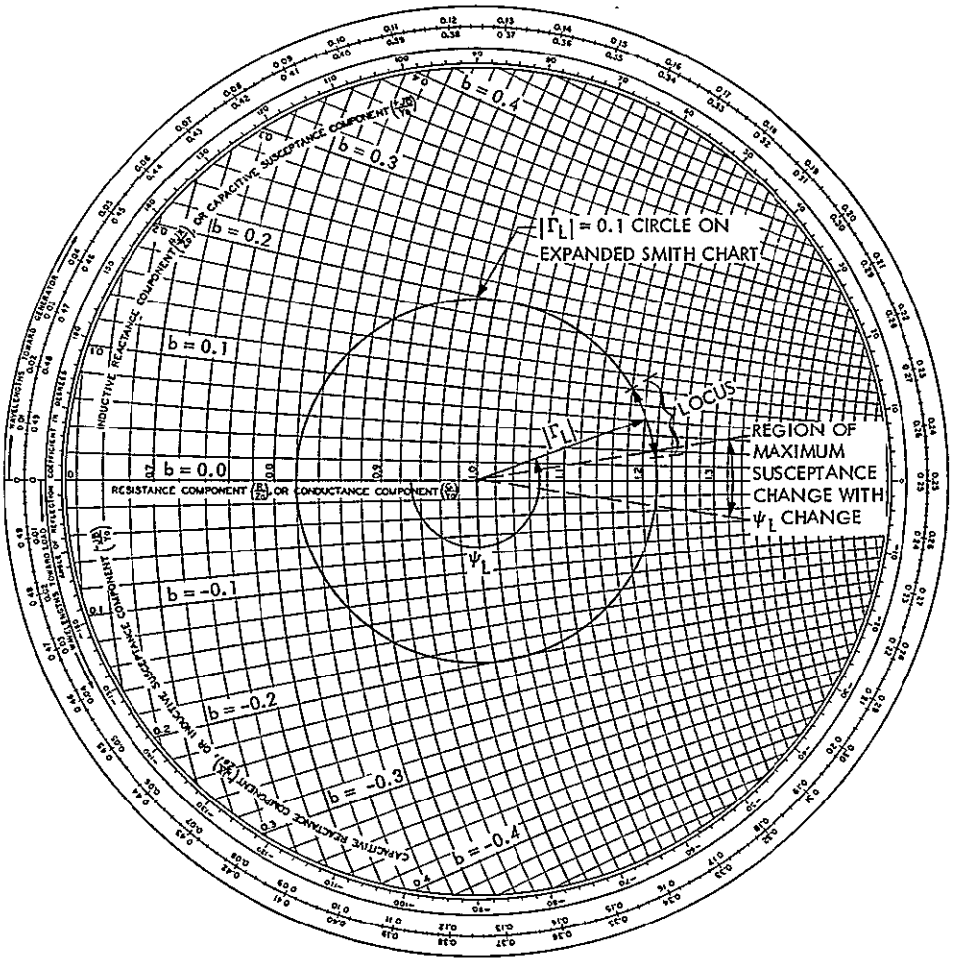

Fig. 4. Locus of varactor reflection coefficient variation in main line of directional coupler

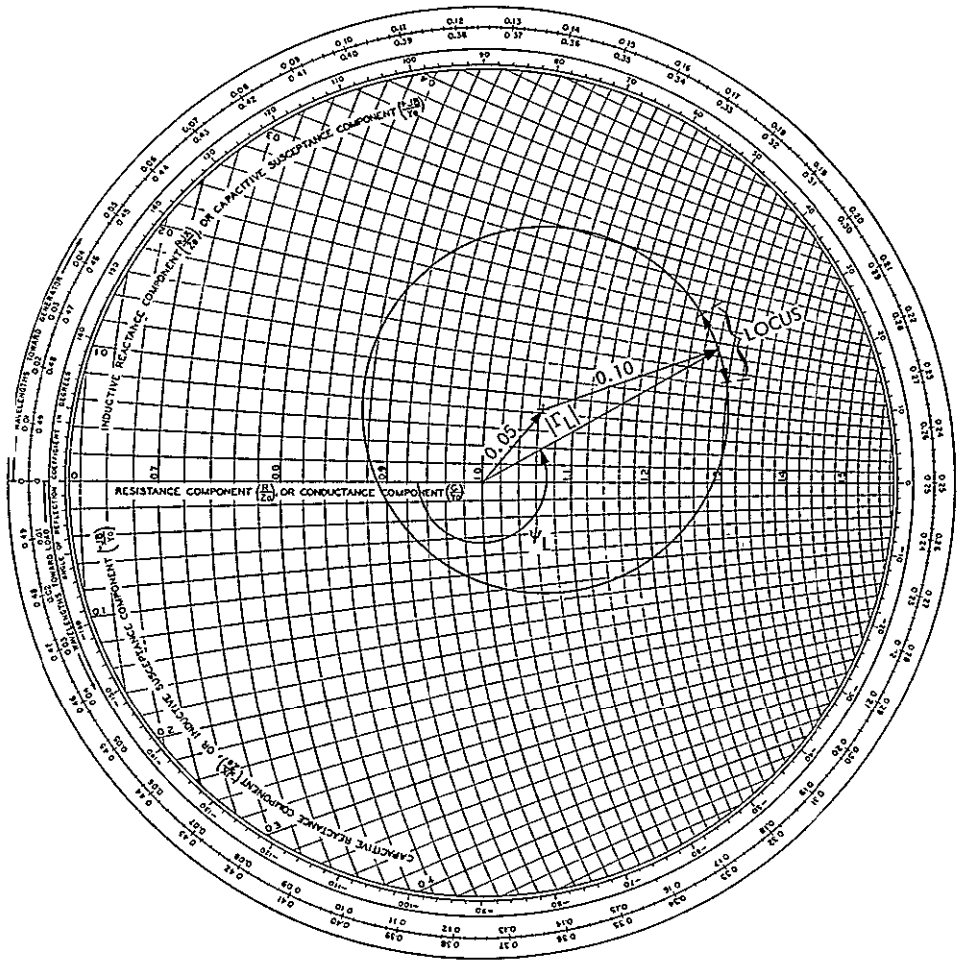

Fig. 5. Offset reflection coefficient locus of varactor as seen in main line

Lateral Forces Induced by a Misaligned Roller

H. McGinness
DSN Engineering Section

The magnitude of the lateral force induced by a roller, misaligned in its direction of travel, can be large and is not linearly proportional to the misalignment.

I. Introduction

If a metal roller which bears against a track is forced to move in a direction not perpendicular to the roller axis, a lateral force is induced. Figure 1 illustrates that the total displacement can be considered as composed of two orthogonal components, namely, one of pure rolling about the roller axis and the other of sliding parallel to the roller axis. The sliding force is $W\mu$, where W is the normal force between the roller and track and μ is the coefficient of sliding friction. The components of the sliding force parallel and perpendicular to the direction of constrained motion are respectively:

$$F_x = W\mu \sin\theta \tag{1}$$

$$F_y = -W\mu \cos\theta \tag{2}$$

For small misalignment angles, where $\sin\theta \approx \theta$ and $\cos\theta \approx 1$, Eqs. (1) and (2) indicate that the drag force, F_x, is proportional to the misalignment, whereas the lateral force has the approximately constant value of $W\mu$. According to the concept from which Eqs. (1) and (2) are derived, the lateral force is approximately $W\mu$, because it is assumed that there has been a finite misalignment angle so that sliding has existed. Once the full value of the sliding frictional force has been established, a static frictional force can be locked in after the sliding has ceased. One can conceive of a slight misalignment being gradually reduced while the wheel is rolling. At zero misalignment the full lateral force of $W\mu$ could exist, provided the zero position had not been overshot. This concept implies that the roller and track are rigid bodies.

In reality the roller and track are elastic bodies and are deformed slightly from the induced lateral force. The effect of this deformation is to prevent the full value of the lateral force from being locked in. Within some small value of the misalignment angle, the induced side force is likely to be proportional to the misalignment angle and not proportional to the cosine of the angle as Eq. (2) indicates. A limited amount of empirical data from Ref. 1 shows the lateral force/misalignment relationship within the small angle regime.

II. Empirical Results

A few tests made at JPL using the apparatus shown in Fig. 2 gave essentially the same results as those of Ref. 1. The experiments of Ref. 1 were all made with a steel roller

0.102 m in diameter and 0.025 m wide. The track was 0.051 m wide by 0.025 m thick by 2.74 m long. Both the roller and track were 4340 steel with a yield strength of 89.6×10^7 N/m^2 (130,000 psi). The track originally had a 0.38-micron rms finish in the direction of rolling and a 1.00-micron rms finish laterally. The surface of the roller originally had 0.38-micron rms finish in both directions. Data were taken after both the roller and track had been cleaned with acetone and again after being lubricated with SAE 10W motor oil. Additional data were taken after both the track and roller had been sandblasted to a surface finish of 4.50 microns rms. Subsequently the track and roller were rusted to simulate outdoor conditions. The force between the roller and track ranged from 1800 to 15100 N (400 to 3400 lb). This corresponds to a nominal Hertz stress of 22×10^7 to 64×10^7 N/m^2 (31,900 to 93,000 psi).

Some of the results of the Ref. 1 experiments are shown in Figs. 3, 4 and 5. In these figures the lateral force coefficient is the ratio between the lateral and normal forces. Figures 4 and 5 indicate that there are no significant differences between force coefficients produced by various surface conditions. Apparently the viscosity of the SAE 10W oil was insufficient to provide a film at these values of Hertz stress. Figures 3, 4 and 5 each show an almost linear relationship between lateral force coefficient and misalignment angles up to about 10 arc min, and a fairly constant value of lateral force coefficient for misalignments greater than 30 arc min.

III. Consequences of the Induced Lateral Forces

A method of obtaining perfect alignment of a roller in its direction of motion is through the use of caster, which means that misalignments are reduced to zero as the wheel advances. However, the reversal of motion of a wheel with caster causes a misalignment to increase. For those cases where the wheel must move in both directions, the wheel or roller design should provide for: (1) achieving and maintaining an alignment within certain small amounts in order to limit the lateral force to certain small values, or (2) accepting and coping with the large lateral force which will exist if the misalignment exceeds a few arc min of angle. Figures 3, 4 and 5 give some insight to the problem. It is emphasized that these figures are based upon a limited amount of empirical data and may not quantitatively apply to a specific application. It would appear, however, that the maximum lateral force coefficient is approximately the same as the sliding coefficient of friction when the wheel is sliding parallel to its axis of rotation. If this sliding friction coefficient could be determined for the particular application, it is believed that it would be approximately the same as the maximum induced lateral force coefficient.

Usually it is desirable that the load intensity between the wheel and track be as nearly uniform as possible. For some wheel suspension configurations the induced lateral force requires an interface moment for static equilibrium; thus the load-intensity across the-width-of-the wheel is nonuniform, the maximum intensity being much above the average. This same suspension configuration may well accommodate for a warped track. Therefore, it is most important to consider carefully the suspension design of a wheel or roller when induced side forces exist and caster cannot be employed.

IV. Examples of Wheel Suspensior Designs

Figure 6(a) shows a wheel without caster mounted so as to allow the wheel frame to pivot about point A in order to permit the wheel to remain flat against the track even if there is a small angular displacement between the track and base structure. Figure 6(b) shows a free body diagram of the wheel and frame. The forces V and S are, respectively, the total normal force between wheel and track and the induced lateral force. For instances of small angular displacement between the track and base structure, where the configuration remains essentially constant, a summation of moments about the wheel-track origin yields the following expression for the required equilibrating interface moment, M_i:

$$M_i = aS \mp VR\mu_B \tag{3}$$

where a and R are shown in Fig. 6b, μ_B is the coefficient of friction of the pivot. Setting $S = \mu V$ where μ is the coefficient of sliding friction between the wheel and track, Eq. (3) becomes:

$$\frac{M_i}{V} = a\mu \mp R\mu_B \tag{4}$$

The $\pm$ sign in Eqns. (3) and (4) accounts for the direction of motion of the pivot.

For the case where there is no relative angular displacement between the track and base structure, it is possible that the two terms on the right side of Eq. (4) always have opposite signs. In this case the value of M_i could be very low. If instead of an intentional pivot at A the wheel frame is solidly attached to the base frame, the effective value of $R\mu_B$ could be large enough to cancel the $a\mu$ term. For such a case, however, the real value of M_i will depend upon the elastic stiffnesses of the various parts of the structure. A certain 30-m-diam antenna is

known to have a very precise azimuth track. Its wheels are mounted so that there is no intentional pivot at point A. The wheel and track wear after four years of operation has been both small and uniform.

For the case where there is an intentional pivot at point A the two terms of the right side of Eq. (4) will at times have the same signs and the maximum absolute value of M_i/V will be:

$$\left|\frac{M_i}{V}\right|_{max} = \left|a\mu + R\mu_B\right| \tag{5}$$

A certain 100-m antenna had wheels suspended such that $R\mu_B$ was effectively very small but the ratio a/L was 6.58. After four years of operation there was enough nonuniform wear to necessitate a remachining of the wheels and track. These wheels and track are nominally identical to those mentioned in the previous example and the nominal load per wheel is the same.

Since the wheel and track coefficient of sliding friction μ is fixed, only a, R and μ_B are available for reduction in magnitude. If the center of R is placed at the wheel track origin, "a" becomes zero, but R becomes large if configured as shown by the dashed line of Fig. 6a.

A suspension system which effectively places point A at the wheel-track origin and effectively make R small is shown in Fig. 7. It has the advantage of allowing the wheel or roller to accommodate to a slightly warped track and does not require a larger equilibrating interface moment in the presence of an induced lateral force. A complete analysis of this design will appear in a separate report.

Reference

1. Cross, H. A., and Talbert, S. G., *Experimental Determination of Lateral Forces Developed by a Misaligned Steel Roller on a Steel Rail*, Machine Design Division of ASME, Oct. 2, 1963.

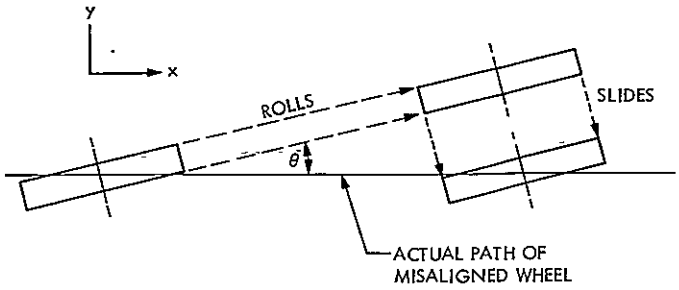

Fig. 1. Model of induced lateral force

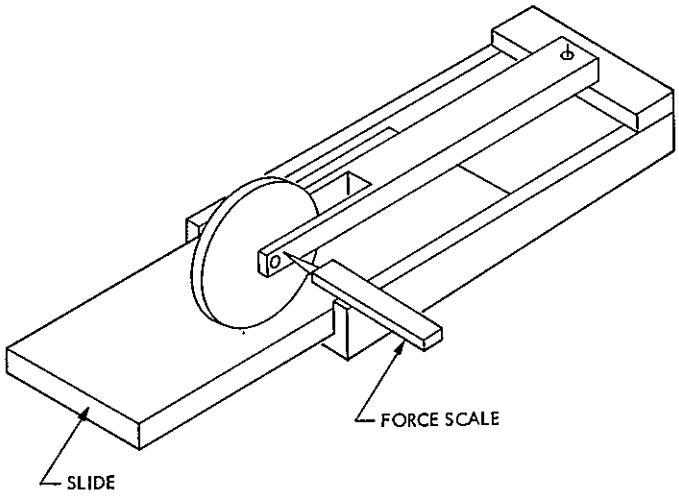

Fig. 2. Test apparatus used at JPL

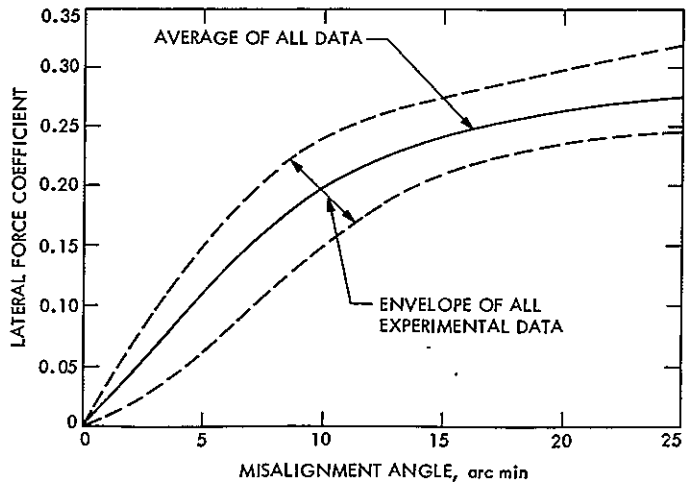

Fig. 3. Lateral force coefficient vs small misalignment

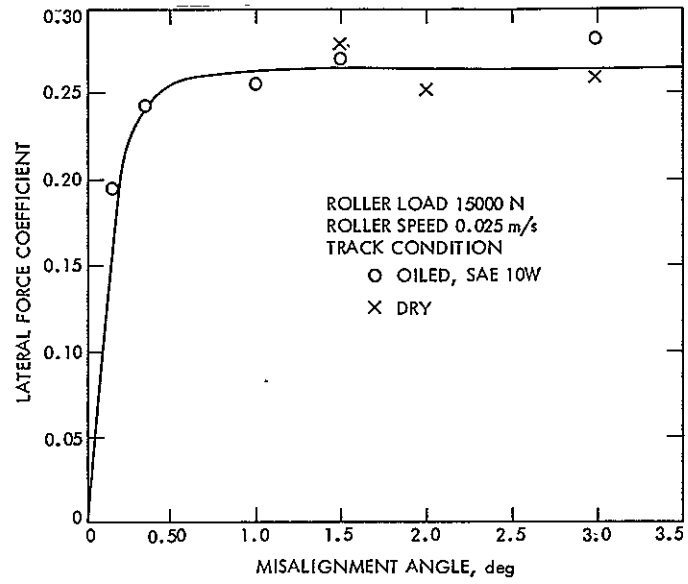

Fig. 4. Lateral force coefficient vs large misalignment

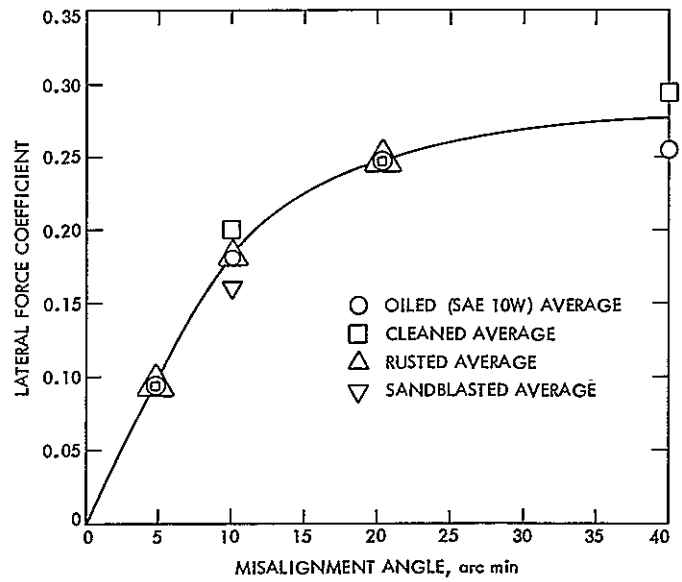

Fig. 5. Effect of surface condition on force coefficient

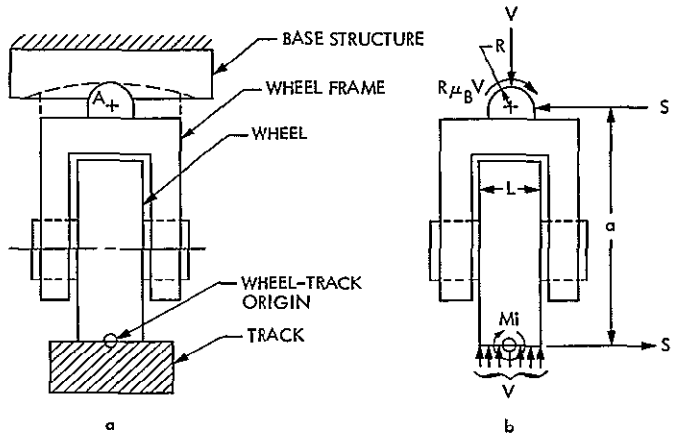

Fig. 6. Wheel suspension

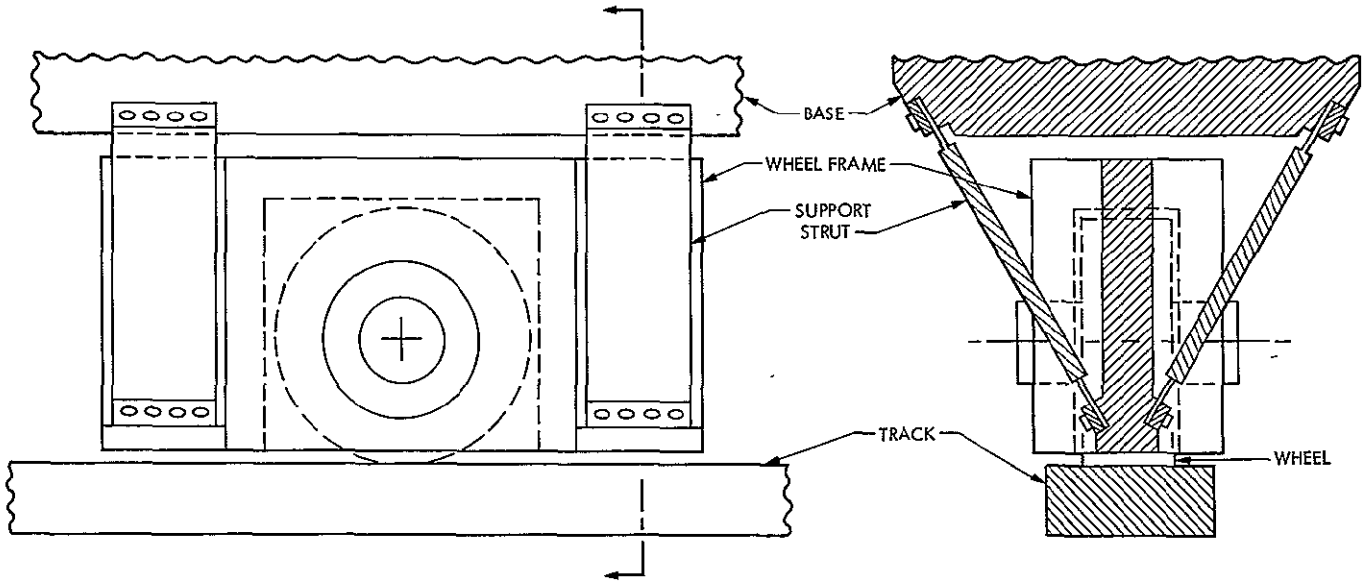

Fig. 7. Wheel suspension with small interface moment

Structural Design of a 64-Meter Low-Cost Antenna

M. S. Katow
DSN Engineering Section

The computer model of a 64-m ground antenna was almost completely generated by 1108 software. The reflector and alidade model was iteratively designed and analyzed by the JPL/IDEAS program, which minimized the distortion RMS with respect to the structural weight. Curves of values describing the optimizing processes are presented, functional aspects of the structural elements are defined, and detail descriptions of the design equations for stress calculations are included. Computed data used for calculating the RF boresight error and natural frequency answers are also included.

I. Introduction

The computer analyses of the 60-, 80-, and 100-meter reflector structural assemblies and their corresponding alidades were supplied to our cost analyzing contractor. The computer analyses were output from the JPL/IDEAS and NASTRAN programs.

This study was a follow-on of work done in Refs. 1 and 2 with the goals of supplying usable details for calculating performances and the manufacturing and erection costs of the antennas. This article describes what design considerations were used to generate the 64-m models and control the designing cycles through the IDEAS program to result in performances that satisfy the specifications.

At the time of this study, the design objective was to maintain or surpass the 64-m MARS' performance with lower costs.

With the emphasis on building more cost effective Deep Space Station antennas, we concluded that the design concepts pioneered by the Bonn antenna designers, using the axial symmetric reflector structure and the efficient four wheel alidade, most satisfactorily fulfills the performance requirements of the antenna at the lowest fabricated cost as discussed in Ref. 1. The elimination of the master equatorial pointing system makes the symmetrical design possible.

The JPL/IDEAS (Ref. 3) program was used to iteratively design the reflector and alidade truss structures. The design objective of the reflector structure was to minimize, by changes in the cross-sectional areas of the truss bar members the RMS distortions of the surface panels connecting nodes for wind and gravity loadings. For the alidade, the objective was to minimize RF boresight changes from the wind loading.

To emphasize the importance of accurate surface panels for X-band antennas, the accuracies and manufacturing steps are described for the surface panels installed on the 64-m DSS-14 antenna.

The NASTRAN (Ref. 4) structural program was used in the alidade analyses where rotations of structural members were

desired at the node points. That is, beam elements with bending formulation were needed instead of pin jointed bars used in IDEAS analyses. However, the design for optimum cross-sectional areas of the alidade members was done by using the "point" mode in the IDEAS program. Here, the compliances of the two mounting points of the reflector were minimized with constraint on the total structural weight of the alidade, with the objective of minimizing the RF boresight shift.

As done before (Ref. 1 and 2), a data generating program was used to prepare the reflector model. The cassegrain primary feed cone, the center hub portion of the reflector, the reflector back-up cone, the "stick" quadripod, the elevation wheel structure and the alidade were hand generated.

II. Reflector Configuration

The primary function of the reflector's truss assembly is to provide stable support points for the paraboloidal reflective surface panels under the environmental loads. Ideally, the best or most uniform deflection is obtained when the support points of each surface panel are on main rib trusses only. However, this truss arrangement would be the most expensive. The trade-off for lower costs is to create support points in the hoop truss and use intermediate ribs to provide the midpoint support points (Fig. 2). Of course, the additional deflections of these nonmain rib supports must be a reasonable fraction of the total desired distortion RMS.

The surface area or size of a surface panel maximized for the desired deflection under the environmental loads results in the minimum number of main ribs. The panels fabricated for the 64-m antenna at DSS 14 used aircraft-type beams for the framework, which resulted in lightweight and stiff panels. Although the surface sheets were riveted to the beam framework, spot welded fastenings or epoxy cemented joints are preferred to eliminate joints in the reflective surfaces that may result in RF noise bursts.

As the size of the surface panels increase, the distortions from the manufacturing processes, and the gravity and air pressure loads will also increase. Some points on the curve "of size vs accuracy" from manufacturing are illustrated by the surface panels built for the 64-m DSS 14 antenna (Table 1 and Fig. 1).

A static load test was made using bricks for weights. Specifically, panel No. 4 (Table 1) was supported top down from an angle iron framework using the normal supporting screws. Bricks equivalent to 1340 N/m^2 (28 lb/ft^2), the expected upward pressure at a 53.6-m/s (120-mph) wind speed, were placed on the panel. The maximum recorded deflection was 4.37 mm (0.172 in.) with the bottom cross straps on, and 5.08 mm (0.200 in.) with the straps off. For the 13.4-m/s (30-mph) air speed load, the maximum deflection was 0.48 mm (0.019 in.). The 50% porosity of the reflective surface was accounted for in the pressure difference calculations. Also, the zero returns of the dial indicators showed that the panels were not stressed to yield for the stow air loads.

The manufacturing process used a fixture with an egg-crate type of frame contoured on top to fit the panel's surface curvature, and a flat sheet of 2.36-mm (0.093-in.) thick aluminum was placed on the frame. Next, the Z-shaped aluminum beams, contoured to the correct curvature by rolling, were placed on top to force the flat sheet to the curvature of the fixture as the beams were pressed against the fixture. With clamps holding down the beams, pilot holes for the rivets were drilled and filled with temporary rivets. The panels were then moved to the automatic riveting machines for applying the rest of the rivets.

Each manufactured panel was placed in a contour checking fixture and held in place at the six fastening points to the reflector structure. The contour error for every 645 cm^2 was read and recorded on IBM punched cards. These cards were input to a computer; the root-mean-square (RMS) evaluations are shown in Table 1. The overall manufacture RMS error for all of the panels was 0.89 mm (0.035 in.).

Inspection of the error pattern of all panels showed that the error was not randomly distributed. It was a result from stressing of the beams by the surface sheet as it tried to return to its originally flat surface. In other words, the curvature readings showed a general flattening of all panels or a definite high point near the center. The vendor has improved his manufacturing techniques where more accurate panels were required by either over rolling the beams to a smaller radius curvature, or stretch-press forming the surface sheets to the correct curvature before the assembly process.

From the foregoing, it was concluded that surface panels up to 8.00 m^2 could be manufactured to 0.5 mm RMS accurately. Also, their stiffness from air loading tests showed that negligible distortion RMS needed to be added for air speeds up to 13.4 m/s (30 mph).

The radial rib type of reflector structure easily satisfied the axial symmetric configuration required for minimum gravity loading distortions. As shown in Fig. 3, with the two supporting points ② and ③ on the symmetric axis, all points on a particular hoop should deflect the same amount for gravity loading in the direction of the axis, if axial symmetry is satisfied. By proportioning the cross sectional areas of the main rib truss and accounting for the hoop-member areas, the

top nodes could be made to deflect to another parabola of longer focal length for the symmetric gravity loading case. Thus, the distortion from the symmetric loading case can be made negligible.

The antisymmetric loading produces opposite direction distortions in the top and bottom halves of reflector if gravity loading at horizontal look is considered. Here, the depth of the ribs has a major effect on the magnitude of the displacements, but some minimization is done in the IDEAS program arranging or changing the radial distribution pattern of the cross-sectional areas at the same time the minimization is done for the symmetric load case.

The number of main ribs of the low-cost 64-m antenna was decreased by a factor of two over the 64-m antenna at DSS 14, and the mounting points for the surface panels were created by adding subtrusses as shown in Fig. 2, Section GG, and in Fig. 3, Sections AA and BB.

The reflector structure (Fig. 3) was completed by adding a back-up cone consisting of bars from the connection node ③ to alternate main ribs at the bottom of the fifth hoop, and a bar from nodes ② to ③. Node ③ thus becomes a very "hard" node and capable of becoming a support point for both axial and lateral directions. The balancing counterweight of the reflector about the elevation axis is located here at node ③.

The computer model of the reflector's truss structure is a half (1st and 4th quadrants) model with normally symmetric constraints on the symmetry plane. For the supporting nodes, node ② is constrained laterally (X and Y of Fig. 3) only. Node ③ is constrained both laterally and axially (Z of Fig. 3).

III. Elevation Wheel/Alidade Configuration

The two base or mounting nodes of the reflector are fastened to nodes ② and ③ of the elevation wheel/alidade assembly shown in Fig. 4. Nodes ② and ③ are in the center part of the inverted-pyramid-shaped truss of the elevation wheel assembly with the elevation bearing shafts at nodes ⑤ and ⑥. The elevation gears (shown as dotted lines) are fastened to a truss assembly attached to nodes ③, ⑦, and ⑧. Nodes ⑤ and ⑥ with ⑦ and ⑧ can support a quadripod separated from the reflector structure (shown with dashed line to node ⑨).

The base of the alidade is a square frame with two cross diagonals and the azimuth radial bearing is located at the intersection of the diagonals. Attached to each corner are a set of wheels that rides a circular rail. To the two sides of the square base, triangular trusses are attached (⑤, (50), (52) and ⑥, (54), and (55)) that support the elevation bearings located at ⑤ and ⑥. The lateral supports for the two triangular trusses are from bars ⑤ – (51) and ⑥ – (51). With only lateral constraints supplied by a cylindrical-type roller bearing in the azimuth radial bearing at (51) and axial fixity at the elevation bearings as supplied by spherical roller bearings, the elevation wheel/alidade assembly becomes a stable determinant truss with special characteristics.

The torsional rigidity of the alidade assembly itself about the X and Y axes are low even with the pentahedron shaped truss (50), (52), (54), (55), and (53) in place. The basic rigidity of the alidade about these axes results from the azimuth bearing wheels at the four corners in contact with the azimuth rail with preloads available from the dead weights. The components of forces on nodes ② and ③ in the Y direction are transferred through the elevation bearings, the side triangular trusses of the alidade and then to the radial bearing, with components to the wheels at the base corners. Forces in the X direction are also directed through the elevation bearings, but now into the inclined bars ⑤ – (51) and ⑥ – (51) and to the azimuth radial bearing with components into the side triangular trusses and the azimuth wheels. Since the radial bearing at (51) cannot resist vertical axial loads, the axial forces in the inclined bars must be equal, but opposite in sign and, hence, the axial thrust or X loads in the two elevation bearings will be equal. Also, thermal expansion of the elevation wheel truss with respect to the alidade truss does not induce thermal stresses. Compensation is made when the azimuth radial bearing moves axially.

The remaining portion of forces in the Y direction is passed through link ③ – (53), representing the elevation gear-boxes, into the pentahedron shaped truss, and then to the base frame of the alidade.

By separating the antenna into two truss assemblies connected at only two nodes, the reaction forces from the reflector assembly analysis can be used as loads for the elevation wheel/alidade assembly analysis. However, it is also possible to attach the elevation wheel assembly to the reflector and analyze the reflector with constraints at nodes ⑤ and ⑥ with fixed cross-sectional areas for the wheel members.

By the first method, the reflector members can be designed for minimum RMS distortion of the reflector geometry while maintaining structural strengths for the maximum wind loads in the IDEAS program. Then the elevation wheel/alidade assembly can be designed for minimum compliance of the reflector's mounting nodes ② and ③ using the "point"

mode in IDEAS. However, the forces transmitted between the reflector and elevation wheel must be separated as analyzed. In other words, all axial or Z forces should be transferred at node ③ while maintaining axial rigidity of only the bar ② – ③ used for the reflector analysis.

Another solution is to attach the elevation wheel to the reflector and check analyze the axial rigidity requirements of bar ② – ③. Bending compliances of the elevation wheel beams must be accounted for in the final solution.

IV. Structural Details Discussion

The 64-m reflector truss assembly was configured for eight rows of surface panels, with the fifth hoop at the 50% diameter. The outer panel has a hoop at its midpoint to maintain an efficient triangular truss configuration (Fig. 3). Several preliminary IDEAS analyses indicated that any increase of the center depth of 3.81 m (150 in.) and the edge height of 1.52 m (60 in.) showed minimal decrease in wind distortion for the extra size.

To minimize wind torques about rotation axes, the vertical distance between the vertex of the paraboloid and the elevation bearing axis should also be minimized. A 2.08-m centerline distance below the reflector was defined to allow sufficient space for maintenance work on the elevation bearing assembly.

The bottom chord of the main girder was configured to be flat and horizontal. This places node ② above the elevation axis centerline. If a box beam is utilized in the elevation wheel, its depth would necessarily place the connection node ② above the centerline of the elevation bearing axis, even for a straight box beam. If necessary, the bottom chords of the main ribs could be dropped in the center for better match up of the pieces. In any case, a bending moment as well as the bending stiffness must be accounted for in the top members of the elevation wheel truss as to their effects on the gravity load distortion of the reflector. This check between the computer models and the actual field members needs to be carefully studied for members joining node ②.

Another node having physical interference problems is at the elevation bearing regions. At node ⑤ (as at node ⑥) bar ⑤ – ⑦, bar ⑤ – ③ and bar ⑤ – ⑧ of the elevation wheel (Fig. 4) interferes with bar ⑤ – (51) in close proximity to the node. The Bonn engineer's solution was to offset bar ⑤ – (51) at the elevation bearing and modify the bar ⑤ – (51) into a shallow truss to account for the resulting end moment from the offset. Since this end moment exists only for wind forces in the X direction and for vibration modes in the same direction, this design trade-off seems to be the correct solution.

The azimuth wheel assemblies must be off-center mounted at nodes (50), (52), (54), and (55) with respect to the forces developed at the wheel/rail contacts. These moments must be accounted for in the practical design.

The elevation gear box compliance was simulated by bar ③ – (53), and the azimuth gearboxes by bar (50) – (70) (typical).

The side clearance between the elevation bearing pillow blocks and the bottom support cone bars of the reflector will determine the azimuth bearing diameter. This clearance must be studied together with clearance for bar (93) and the same cone bars in a layout.

V. Reflector Analysis Description

The iterative design procedure for the reflector structure started first with the effort to minimize the wind distortion RMS as the overall structural weight was revised to a designated intermediate structural weight number through designing cycles in the IDEAS program. Then, with the program constrained not to reduce any cross-sectional areas, weight was added to the final desired weight figure to reduce the gravity distortion RMS. This procedure was described in more detail in Ref. 1 and the resulting 64-m design figures are delineated in Fig. 5.

The distortion RMS was computed within the IDEAS program by best fitting the deflections of the work points supporting the reflective surface panels with a paraboloid (Ref. 8). Each deflection number was weighted by a function of the surface panel area the node represented. This should represent the uniform amplitude objective of the shaped reflector surface.

The symmetry of the reflector's configuration was maintained by "grouping" the bars required to be of the same cross-sectional areas as those in the same "group" in the IDEAS program.

Throughout the designing cycles that minimized the RMS distortions, the IDEAS program was checking the structural quality of each bar as well as adjusting the cross-sectional area for minimizing RMS. Based on the axial loading, and the input stress limitations and design equations recommended in Ref. 5, the cross-sectional area of each group was revised, if necessary, through each cycle to comply with the requirements. The amount of the change was limited to the selected ten maximum percentage for each design cycle.

The end result of the IDEAS cycles was a structural design of a reflector assembly of the desired structural weight with minimum wind distortion RMS for the selected wind load, and minimum gravity loading RMS. The wind loading used was for the back wind direction at 60 deg elevation that produced the largest pitching moment and the maximum surface distortion also. If the objective design weight was large enough, then all of the bars would pass the structural quality tests without compromising the wind distortion RMS and gravity loading RMS.

The largest axial loadings of the bars normally results from the limit wind loadings. One limit loading occurs at stow configuration with the wind at maximum speed. The other occurs at horizontal look configuration with the wind directly into the back of the reflector, and with the wind at maximum speed allowable for this case.

The stow wind speed for design was determined to be 44.1 m/s (100 mph) (Ref. 6) for the Goldstone location. It was based on statistical analyses of wind data throughout the United States using the probability of occurrence of once in 100 years with gust factors added per Ref. 6.

With the reflector at operating altitudes other than the stow case, the hoops are compressively loaded to their maximum with the reflector at horizon look and with wind into the back side. This is the result of the erection procedure where the hoops are the last to be installed between the ribs. Hence, at the erection altitude of zenith look, the hoops are at zero load. Consequently, the rotation to horizon look applies a negative gravity load that loads the hoops compressively. This compression loading is then added to the compression loading resulting from the wind force into the back of the reflector.

The wind forces on work points of the reflector structure were computed by a 1108 program that generated the NASTRAN force card images. Pressure difference coefficients, derived using the method described in Ref. 7, were input to the program with geometry data. These coefficients were then multiplied by the surface panel areas applicable to each workpoint to create the force card data. A 1.15 factor was used in the force card to account for the wind drag of the reflector structure in the vicinity of the workpoint. The drag forces on other parts of the reflector's structures were not accounted for in the distortion calculations.

The structural quality checks were made for all loadings with no additional load or safety factors with exceptions as follows:

(1) The compressive yield stress was input at 2.48×10^5 kN/m^2 (36,000 psi) for use of the ASTM A36 structural steel. The tensile yield stress was input at 2.07×10^5 kN/m^2 (30,000 psi) for a 20% safety factor. Since compressive stability governs the design because of long members used in reflectors, the tensile input values appeared to have little effect on the design.

(2) The standard air density was used with no factor used for Goldstone's elevation.

(3) Load reversal was assumed. This meant that all tensile forces were checked as compressive forces.

For all wind loads other than the stow wind case, 31.3 m/s (70 mph) wind was used to develop the wind forces.

With the servo drives designed to move the antenna at any altitude in winds of 23 m/s (50 mph), the maximum loading condition at 23 m/s (50 mph) is possible. For this case, the factor of safety computes to two based on the differences of wind velocities.

The design equations recommended in Ref. 5 are incorporated in the IDEAS program, which computes the maximum permissible compressive stresses with no safety factor included. The equations and their applicable scopes are described in more detail in the appendix. A limit of the length to the radius of gyration (L/r) of 200 is made in the IDEAS program, which is equivalent to stating that all bars are considered to be pin ended. Recommendations in Ref. 5 allow partially restrained connections with multiple bolt connections for bracing members with L/r ratios to 250. With the projected use of welded square tube members, stress calculations should produce answers on the conservative side.

The main effort in the reflector design is focussed on attaining low wind distortion RMS; in other terms, it is a stiffness design. The stresses from just gravity loadings should be low. Also, if the final structural weight is large for minimizing wind distortion, the structural quality calculations will increase the allowable wind speeds from 44.7 m/s (100 mph) for the stow case as well as for other altitudes cases.

VI. RF Boresight Pointing Analysis

The environmental loads on the antenna structures normally produce changes of the RF boresight directions, which result in pointing gain losses aside from the gain losses due to distortions of the reflective surfaces. The pointing errors of the primary reflector's distortion may be calculated using the computed parameters of the best fit paraboloid (Ref. 8) to the distortion vectors, and the displacements of the phase centers of the RF feed. Ray tracing as shown in Fig. 7, defines the change in the RF boresight direction.

With the primary reflector supported on nodes ② and ③ of the alidade (Figs. 4 and 6), additional pointing errors were calculated from computed displacements of these nodes in the alidade model by loadings of reaction forces from the reflector analysis.

With the elevation angle encoders mounted at nodes ⑤ and ⑥ measuring the differential rotations of the alidade and the elevation bearing shafts of the elevation wheel truss, two presently uncorrectable errors occur with respect to the elevation angle measurements to ground. The first error is the twist in the elevation wheel truss between the rotations of nodes ② and ③ and the elevation bearing shafts. The second error occurs as the alidade joints or nodes ⑤ and ⑥ carrying the encoder's housing rotate slightly from displacements from the loading.

Using the 64-m computer model delineated in Table 2, the best fit paraboloid data to many wind load distortions are given in Table 3 to suit labels shown on Fig. 7. The largest distortion RMS and boresight error results from back wind to elevation 60 deg, which is the same for YAW 120 deg at elevation 0 deg. Table 4 lists the rotations of the alidade members previously described for four wind loads that determine the RF boresight error values.

In addition to the gain losses from pointing error, the gain loss from offset or separation of the RF feed's phase center and the focus of the best fit paraboloid must be added. This offset gain loss can be computed by use of the radiation program (Ref. 9) resulting in straight line functions between the offset and the equivalent gain losses in distortion RMS values. The functions shown on Fig. 8 are for the F/D value of 0.4235 used for our studies and the equivalent RMS loss values are converted to gain losses by the Ruze equation.

VII. Alidade Analysis Description

The elevation wheel/alidade model shown in Fig. 4 was laterally (X and Y) constrained at node (51), the azimuth radial bearing position; vertically (Z) constrained at the four corner nodes (50), (52), (54) and (55), the azimuth thrust bearing positions; and rotationally constrained about the Z axis by four azimuth gearbox links to ground nodes.

Nodes ② and ③ were loaded with CONM2 weights of the reflector assembly for analyzing the alidade for the effect of gravity loadings. The counterweight at ③ was also modeled with a CONM2 weight card. Then, the reactions resulting from the wind loadings of the reflector were added as force cards at nodes ② and ③. For the horizontal look configuration, the elevation wheel was rotated with node ⑧ above node ⑦. Two bars were added to the elevation wheel to connect nodes ③ and ⑦ to the horizontal elevation gearbox link as shown in Fig. 6.

The design cycle of the elevation wheel/alidade model will be illustrated with computed values from an earlier study with the same configuration as shown in Fig. 4. First, the model was analyzed using the stress designing mode of the IDEAS program, where each bar is reduced to the minimum area satisfying the stress constraints of the design equations previously discussed. With the elevation wheel at zenith look, the reaction forces from the reflector model for the stow wind loads in the X and Y directions (Fig. 4), and the reaction forces for the 60-deg elevation with wind from the back were used as loads on the model. The computed displacements of nodes ② and ③ for the stow wind loads after the stress designing mode is shown by the left most values of Fig. 9 plotted against the structural weight.

Next, the model was analyzed using the point designing mode with moment-type dummy forces applied at nodes ② and ③ and using, first, the stow wind reaction loads in the -Y direction as the real loads. The IDEAS program then minimized the combined displacements of nodes ② and ③ resulting from the real loads by appropriate changes in the cross-sectional areas, with the total weight as a constraint. Thus, as the total weight increase for five design cycles, the displacement results are shown as solid lines in Fig. 9 to the right of the first values computed by the stress mode.

The combined displacements of nodes ② and ③, or the rotation of the line connecting the two nodes, is shown to decrease in the plane of the real and dummy loadings. However, the rotation in the normal plane or XZ plane was also reduced with the increased areas, but to a lesser degree.

To decrease the rotation in the XZ plane, design cycles using the dummy and real loads in the -X direction, with the cross-sectional areas constrained from any reduction, produced the result shown by the dotted lines in Fig. 9 with a slight weight increase.

Figure 10 illustrates the individual bar area increases throughout the previously described design cycles.

VIII. Natural Frequency Analysis

Natural frequency analyses using the IDEAS program were made on an earlier computer model with the same configuration, but with slightly different dimensions to the model described in this article. The results shown in Table 5 should be well within 10% of the values for the described model.

IX. Conclusions

The basic design problem in the reflector structure is compressive failure of the bars in the truss. Nominally, the structure at stow is calculated to just fail on occurrence of a maximum wind velocity in a 100-year mean recurrence interval period. However, there are in-built safety factors on the plus side consisting of (1) limit of length to radius of gyration ratio of 200 for all bars, (2) pin ends on all bars, (3) load reversals assumed in the computations leading to tension loads calculated as compressive loads, (4) sea-level dynamic pressures assumed for wind loading calculations, (5) the 100-year mean recurrence interval period recommended in Ref. 10 for use with permanent structures (the 50-year period is used for transmission towers in Ref. 5), (6) the bars that reach failure mode first are inside the highly redundant truss where redistribution of forces can occur, (7) the designing process in the IDEAS program for minimizing the RMS wind distortion adds areas to the critical bars that are also critical at stow wind condition. This may result in added safety factors for the stow case. The addition is a function of the final structural weight of the reflector.

It turns out that if the reflector structure is loaded only by its own weight and if all cross-sectional areas are reduced by the identical ratio to the original areas, the gravity distortion figure and the longitudinal stresses remains the same as before. The wind loading stresses will increase inversely by the ratio and, hence, the distortion RMS figure.

If a lighter weight reflector is desired, with the increased wind distortion RMS of a finished design, the IDEAS program will proceed with the redesign reducing the cross-sectional areas while maintaining the stress constraints for all loads and the gravity distortion RMS figure, if there are excess areas for wind loads.

At this point in design, a "stick" quad was mounted with its height about equal to the focal length, and the deflection figures for the node representing the apex of the quad was used as the deflection of the hyperboloid's focus point. Further design effort is needed in this area to upgrade this displacement number of the RF feed system. With strong emphasis on quad legs, and with minimum shadowing of the reflector that results in low vibrational frequencies of the whole quad, stability problems in a wind gusting atmosphere may be introduced without careful study of this problem.

Appendix

Structural Design Equations

The self-supporting steel transmission tower is a truss ssembly of steel angles using bolts for connection means. The ritical loadings are compressive instabilities of the truss nembers. Since these characteristics can be applied similarly o reflector structures, the design guides for the transmission owers (Ref. 5) should provide a sound basis for the structural lesign of reflector trusses. The recommendations of Ref. 5 eflect the experience of good practice and actual full-scale lestruction tests on towers.

The two basic strength formulas with no safety factors :alculates the allowable unit stress on the cross section of ıxially-loaded compression members.

Short column when $KL/r \leqslant C_c$

$$F_a = \left[1 - \frac{\left(\frac{KL}{r}\right)^2}{2\,C_c^2}\right] F_y \qquad (1)$$

$$KL/r = 60 + \frac{0.50\,L}{r} \text{ for } \frac{L}{r} \text{ up to and including } 120 \qquad (2)$$

Long column when $KL/r \geqslant C_c$

$$F_a = \frac{*1.97 \times 10^9}{\left(\frac{KL}{r}\right)^2} \qquad (3)$$

$$KL/r = \frac{L}{r} \text{ for } \frac{L}{r} \text{ from } 120 \text{ to } 200 \qquad (4)$$

$$C_c = \pi\sqrt{\frac{2E}{F_y}} = 126.1 \text{ for steel (A36)} \qquad (5)$$

where

K = restraint factor
L = length of column, m
r = radius of gyration, m
F_a = unit compressive stress, kN/m^2
F_y = modulus of elasticity, kN/m^2
E = modulus of elasticity, kN/m^2

*1.97×10^7 (kN/m^2) = 2.80×10^8 (psi)

The use of Eq. (2) in Eq. (1) accounts for normal framing eccentricities at both ends of a "short" column. Equation (2) reflects significant reductions in strength of the compression member that occurs in the intermediate range of slenderness from eccentricities, crookedness, and residual stresses. However, this reduction diminishes for still longer columns, and becomes relatively insignificant for L/r in excess of C_c.

For members with L/r in excess of C_c, Eq. (4) used in Eq. (3) signifies that both ends are unrestrained against rotation. This results in more conservative results as compared to the guide (Ref. 5), which allows partial restraints against rotation for joints made by multiple-bolt connections to a member considered to have adequate flexural strength to resist rotation of the joint. However, the members are not analyzed for drag load from the stow wind that produces bending stresses, which should be added to the axial stresses. Since the summed stresses for the above loading in a typical member of the reflector structure does not exceed the yield stress, the use of Eq. (3) and (4) is considered to be logical.

References

1. Levy, R., "Conceptual Studies for New Low-Cost 64-M Antennas," in the *Deep Space Network Progress Report 42-33*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 55–61, June 15, 1976.
2. Katow, M. S., "Radio-Frequency Boresight Analysis of the Low-Cost 64-Meter Antenna," in the *Deep Space Network Progress Report 42-35*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 128, 130, August 15, 1976.
3. Levy, R., "Iterative Design of Antenna Structures," in the *Deep Space Network Progress Report*, Technical Report 32-1526, Vol. XII, Jet Propulsion Laboratory, Pasadena, Calif., pp. 100–111, December 15, 1972.
4. *The NASTRAN User's Manual* (level 16.0), NASA SP-222(03), March 1976.
5. "Guide for Design of Steel Transmission Towers," *ASCE*, 1971. Prepared by the Task Committee on Analysis and Design of Structures of the Structural Division of the American Society of Civil Engineers.
6. Katow, M. S., "Proposed Modifications to the Wind Load Specification Applied to Ground Antennas," in the *Deep Space Network Progress Report 42-32*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 65–69, April 15, 1976.
7. Katow, M. S., "Aerodynamic Static Differential Pressure Values for the 50 Percent Porous Reflector Dish," in the *Deep Space Network Progress Report 42-29*, pp. 60–65, Jet Propulsion Laboratory, Pasadena, Calif.
8. Katow, M. S., and Schmele, L. W., "Antenna Structures: Evaluation Techniques of Reflector Distortions," in *Supporting Research and Advanced Development, Space Programs Summary 37-40*, Vol. IV, pp. 176–184, Jet Propulsion Laboratory, Pasadena, August 31, 1966.
9. Bathker, D., "Radiation Pattern Programs," in *Computer Programs for Antenna Feed System Design and Analysis*, Edited by Ludwig, A., TR 32-979, Jet Propulsion Laboratory, Pasadena, Calif., April 15, 1967.
10. Task Committee on Wind Forces, "Wind Forces on Structures," *Trans. ASCE*, Vol. 126, Part II, 1961, pp. 1124–1198.

Table 1. 64-m DSS 14 surface panel data

Panel No.	Dimensions			Area, m^2 $(ft)^2$	Surface Accuracy, mm (in.)
	A, m	B, m	C, m		
2	2.10	1.87	4.01	7.96 (85.7)	1.08 (0.043)
3	1.87	1.59	4.73	8.18 (88.1)	1.07 (0.042)
4	1.59	1.35	3.94	5.79 (62.3)	1.00 (0.039)
5	1.35	1.11	3.86	4.75 (51.1)	0.88 (0.035)
6	2.22	1.82	3.15	6.36 (68.5)	0.81 (0.031)
7	1.83	1.28	4.23	6 58 (70.8)	0.83 (0.033)
8	1.28	0.77	3.92	2.66 (28.6)	0.46 (0.018)
9	1.54	0.81	2.79	2.26 (24.3)	0.43 (0.017)

Table 2. 64-m computer model data

SEQ	Description	Weight	
		kg x 10^{-4}	(kips)
1	Reflector with back-up cone, with Cassegrain cone, and quad-structural weight	29.48	(650.0)
2	Additional Cassegrain weight	1.74	(38.0)
3	Additional quad weight	0.64	(14.0)
4	Surface panels weight	5.73	(126.4)
5	Alidade structural weight	68.21	(1503.7)
6	Counterweight	13.37	(295.0)

Major missing weights = elevation gear and its supports and base stiffeners for alidade

SEQ	Load/wind description	Wind velocity		Distortion RMS	
		m/s	(mph)	mm	(in.)
1	Zenith look gravity off/on	–	–	0.15	(0.0058)
2	Horizon look gravity	–	–	0.20	(0.0078)
3	Zenith look gravity – set 36 deg	–	–	0.17	(0.0067)
4	Horizon look gravity – set 36 deg	–	–	0.10	(0.0038)
5	60 deg elevation – back wind	13.4	30	0.31	(0.0123)
6	90 deg elevation – side wind	13.4	30	0.21	(0.0081)
7	0 deg elevation – back wind	13.4	30	0.05	(0.0019)
8	0 deg elevation – into face	13.4	30	0.06	(0.0022)

Table 3. 64-m best fit data

Best fit data	Wind velocity = 13.4 m/s (30 mph)			
	El 60 deg back wind or El 0 deg yaw 120 mm (in.)		El 90 deg – stow mm (in.)	
Distortion – RMS	0.31	(0.012)	0.21	(0.008)
A – Z offset	0.05	(0.002)	0.00	(0.000)
B – axis rotation	0.000269 – rad		0.000221 – rad	
C – focal length – best fit	27.0993 – m		27.0984 – m	
D – Y offset	12.80	(0.504)	10.90	(0.429)
E –	7.29	(0.287)	5.99	(0.236)
F – feed deflection	0.43	(0.017)	0.28	(0.011)
G –	5.51	(0.217)	4.90	(0.193)
H –	5.94	(0.234)	5.18	(0.204)
I –	4.75	(0.187)	4.14	(0.163)
J –	2.54	(0.100)	1.85	(0.073)
K – boresight error	0.000094 – rad		0.000068 – rad	

Table 4. 64-m alidade boresight error data from wind loadings

SEQ	Wind direction	② − ③ rotation × 10^4 rad	Elevation shaft rotation ⑤ × 10^4 ⑥ rad		Encoder housing rotation ⑤ × 10^4 ⑥ rad		Uncorrected elevation error ⑤ × 10^4 ⑥ rad		Elevation boresight error ⑤ × 10^4 ⑥ rad		Azimuth boresight error × 10^4 ⑤1 rad
1	Yaw 0 deg Elevation 60 deg Back wind 31.3 m/s 13.4 m/s	9.06 1.66	8.65 1.59	8.65 1.59	3.06 0.56	3.06 0.56	0.41 0.08	0.41 0.08	3.47 0.64	3.47 0.64	0.0
2	Yaw 120 deg Elevation 0 deg 31.3 m/s 13.4 m/s	−4.03	2.33	2.58	0.76	2.10	1.70	0.36	1.70 0.31	0.36 0.07	−2.10 −0.39
3	Yaw 0 deg Elevation 90 deg 44.7 m/s 13.4 m/s	10.78 0.97	10.65 0.96	10.65 0.96	4.45 0.40	4.45 0.40	0.13 0.01	0.13 0.01	4.58 0.41	4.58 0.41	0.0
4	Yaw 90 deg Elevation 90 deg 44.7 m/s 13.4 m/s	−3.76	−0.08	−0.31	1.87	−2.28	−0.11	0.11	1.76 0.16	−2.17 −0.20	0.0
	44.7 m/s 13.4 m/s								−3.76[a] −0.34	−3.76[a] −0.34	

[a]Cross-elevation error.

Table 5. 64-m reflector plus alidade natural frequencies

SEQ	Elevation angle, deg	First mode rotation motion axis	Natural frequency, Hz
1	90	Azimuth axis	1.28
2	90	Elevation axis	1.18
3	0	Reflector symmetric axis	1.11
4	0	Azimuth	1.37

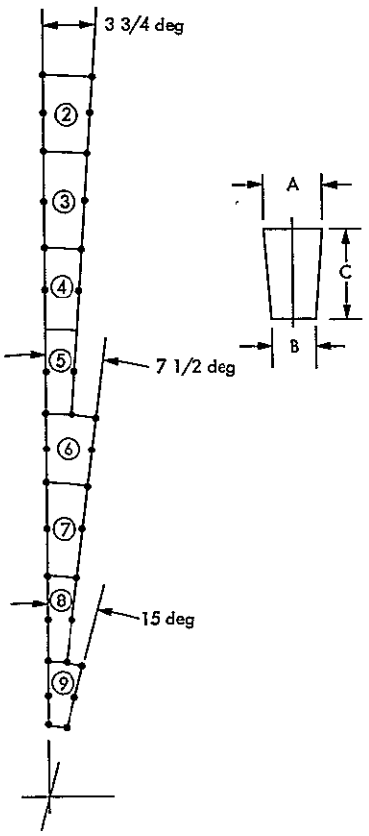

Fig. 1. 64-m antenna surface panels configuration, DSS 14

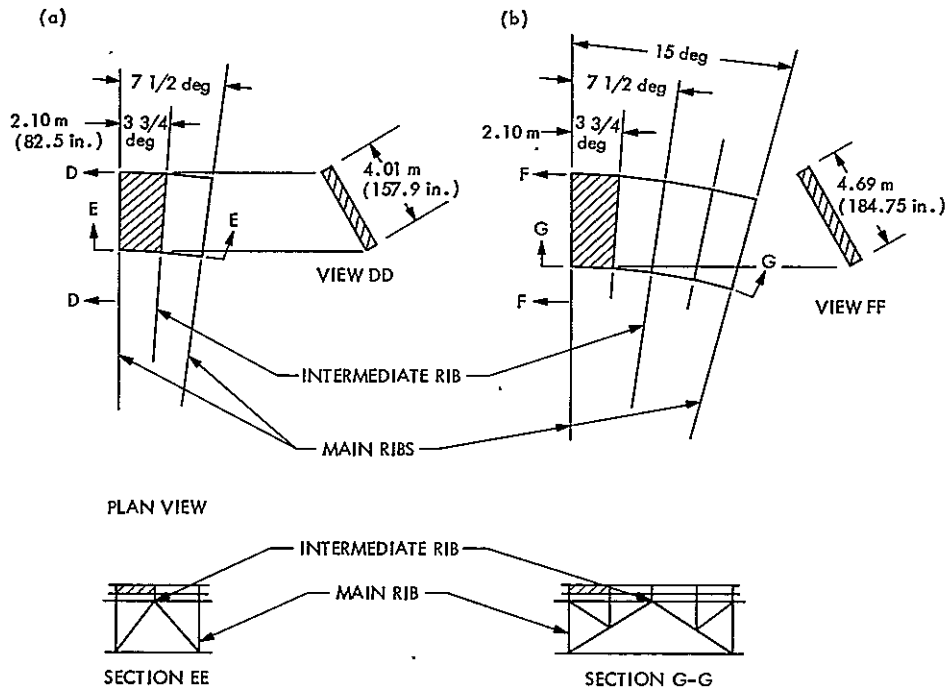

Fig. 2. Surface panel sizes: (a) 64-m, DSS 14; (b) 64-m, low cost

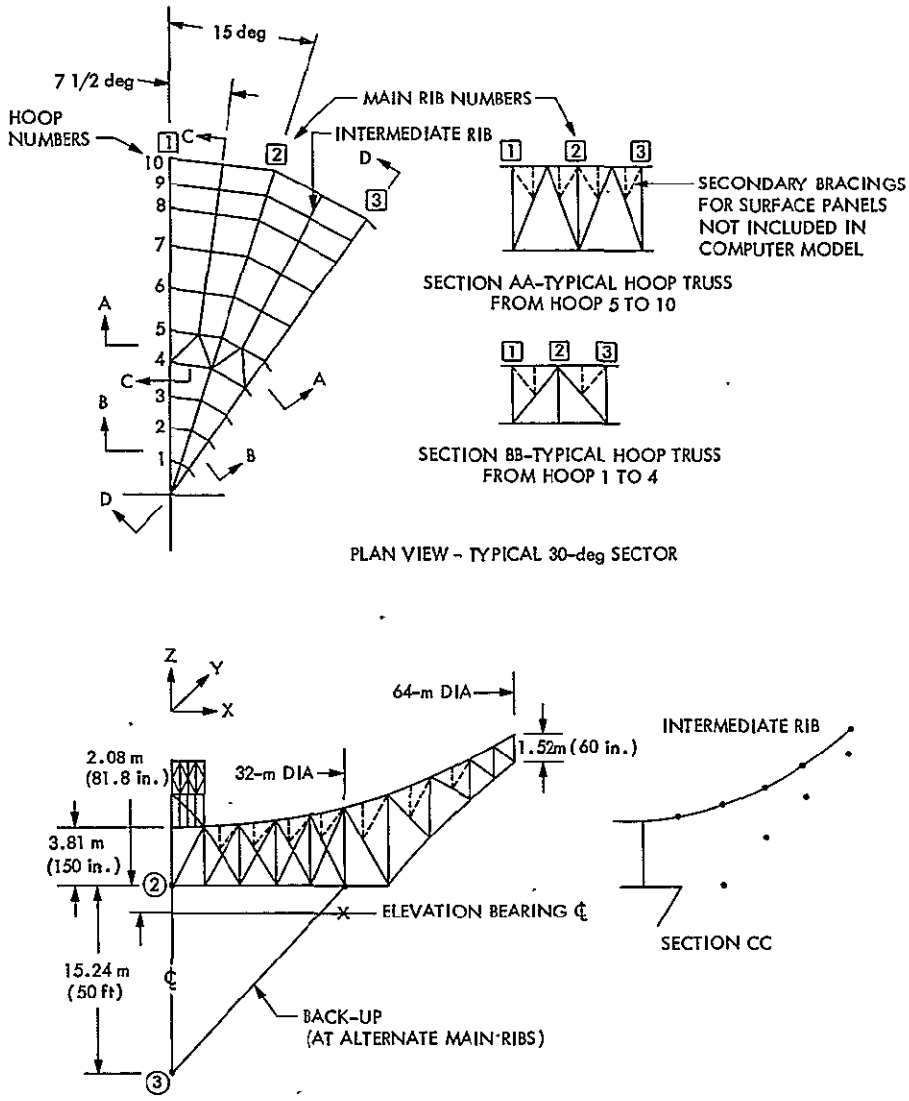

Fig. 3. 64-m low-cost reflector structural model

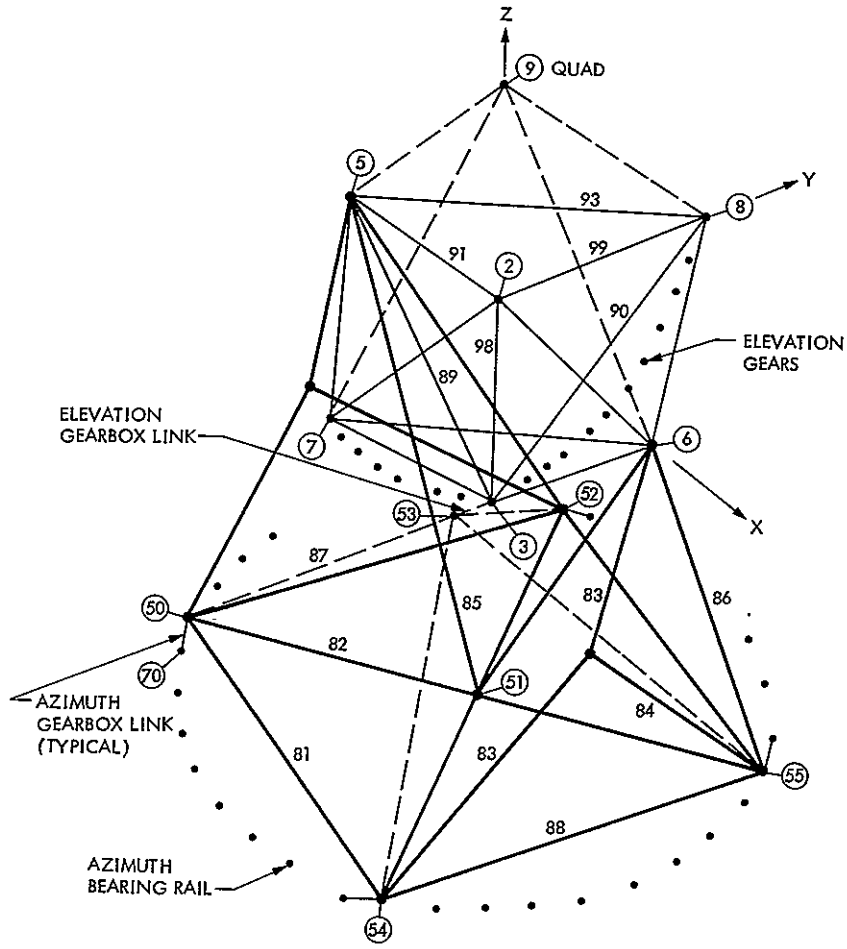

Fig. 4. Elevation wheel/alidade model

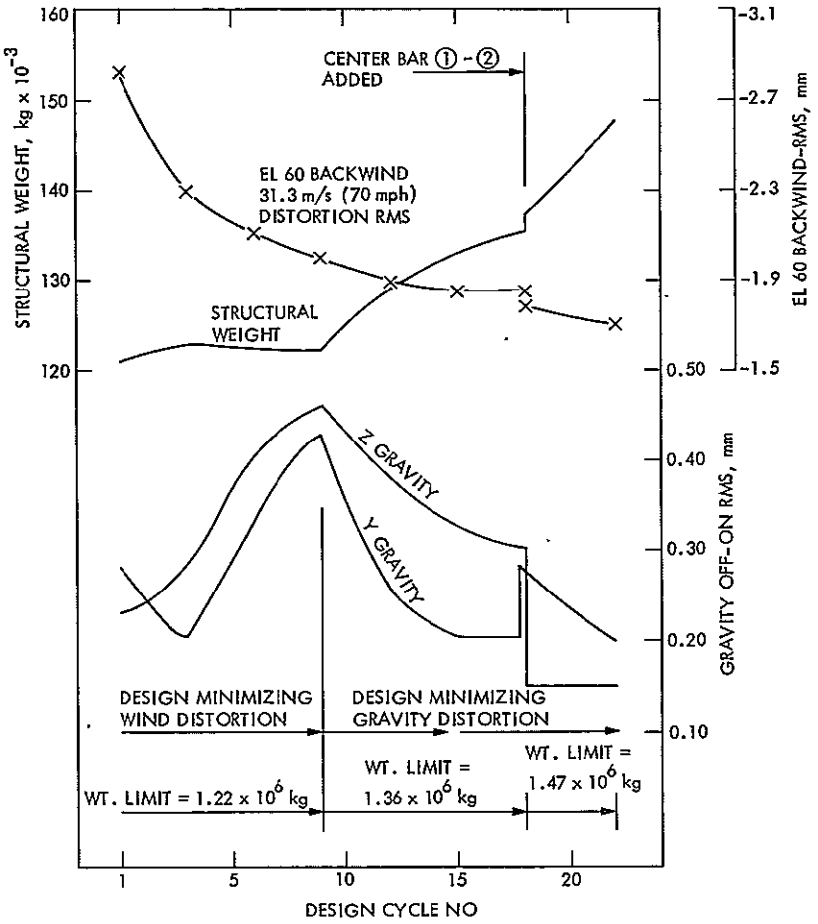

Fig. 5. 64-m IDEAS iterative design data

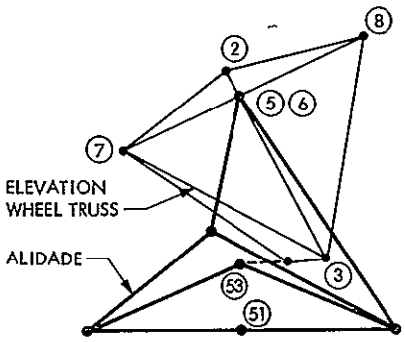

Fig. 6. Elevation wheel/alidade at 60-deg elevation

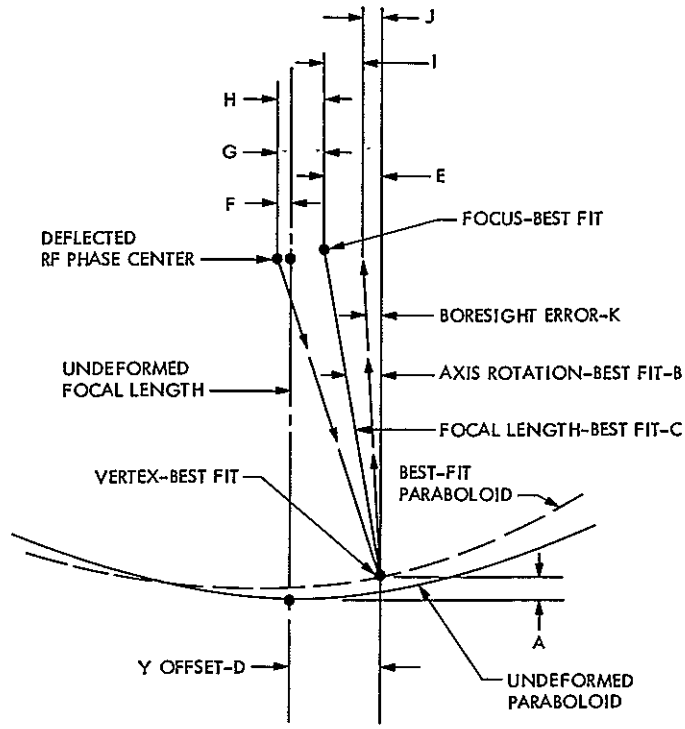

Fig. 7. 64-m reflector boresight error data

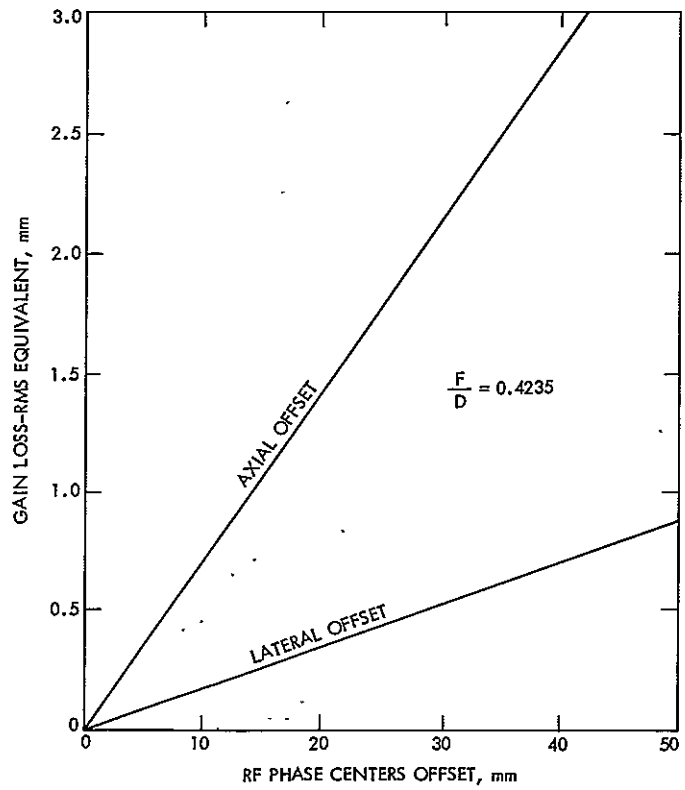

Fig. 8. Phase centers offset gain loss

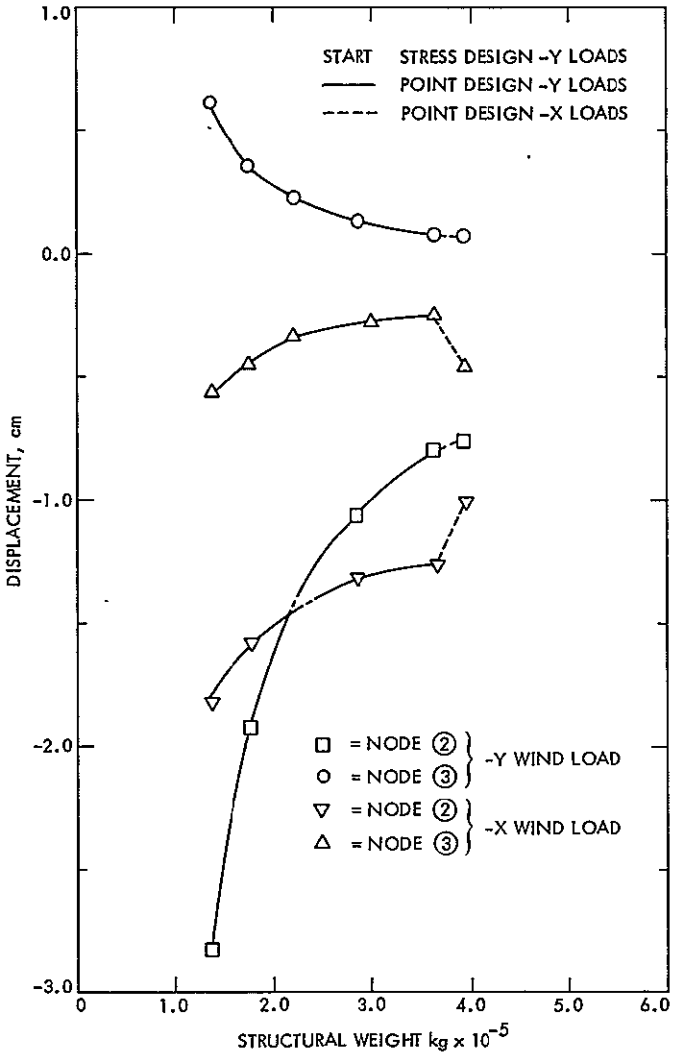

Fig. 9. 64-m alidade displacements vs design cycles

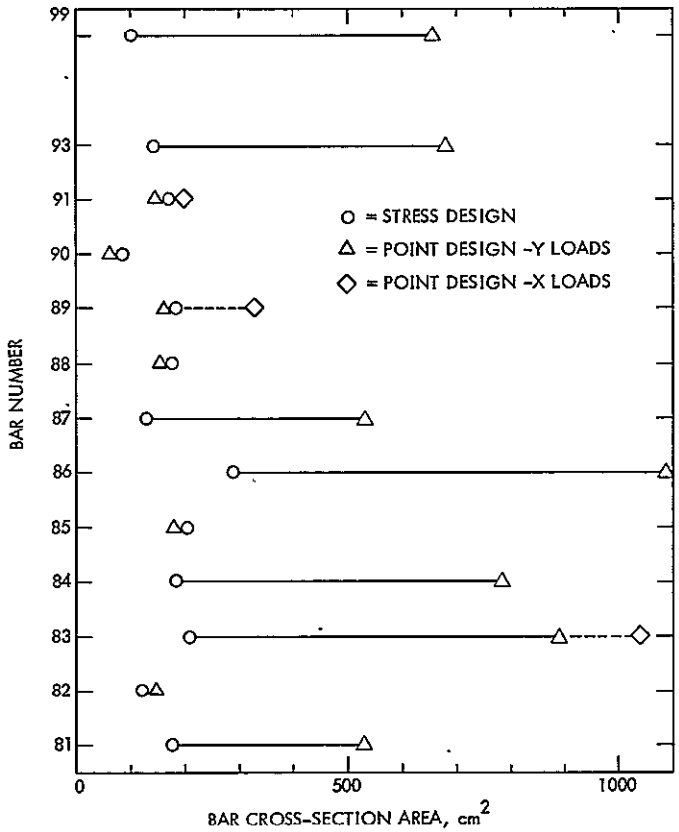

Fig. 10. Bar area changes - 64-m alidade design

CCM Implementation Status

R. A. Mancini
DSN Data Systems Section

Intensive activity has been invested in the preparation of planning the Control and Computational Module (CCM) implementation into the DSN. A de facto standard has evolved from work performed by the Communications Systems Research Section. The DSN Data Systems Section is currently coordinating the total effort involved with the CCM implementation. Part of this effort is the selection of a Standard CCM family. A CCM Selection Recommendation Committee has been formed of representatives of the Telecommunications Science and Engineering Division and the DSN Engineering Section. A support CCM Policy Committee has also been formed of broad representation to develop policies to govern use of CCMs in DSN application. Every effort is being taken toward the establishing and support of a CCM Standard family for use in designing and implementing applicable equipment for the DSN.

I. Introduction

Because of the widespread availability of Large Scale Integration (LSI) programmable modules and their support chips,[1] the DSN has for some time been interested in establishing a standard family of chips to be used for applicable new designs for implementation into the network. The Control and Computational Modules (CCMs) refers to a family of IC chips consisting of a CPU (central processing unit) Group, Memory Group, Input/Output chips, Special Function chips, and Support chips. See Table 1 for definition of CCM Family. The LSI CPU chips are generally referred to as microprocessors. In the latest technology where the CPU, memory, and I/O are on a single chip, the proper term is "microcomputer" (this configuration is not at this time being considered for inclusion in the CCM Standard). The strongest driver for going to a CCM standard is reduction of life cycle costs in the maintenance and operation area. There has been design activity (using microprocessors) in progress for severa years at JPL and an initial effort to establish a standard wa carried on in the Communications Systems Research Sectio culminating in an initial microcircuit set (CCM design review September, 1976). This set is being considered a de facto standard until a standard can be selected.

Involvement by the DSN Data Systems Section (338 engineers in helping to establish and support a standard CCM set began with FY'78 (Oct. 1, 1977). This effort has beer directed toward the establishment of a DSN standard and ha involved the forming committees made up from severa sections. One committee has worked on establishing a CCM policy and another on the selection of a DSN standard CCM Meetings have been held in gathering data about the available chips. Surveys have been made to gather information; one to determine internal requirements and another to determine

available industry capabilities. All efforts are focused on establishing a CCM standard family and supporting its use in the DSN.

II. Background

In preparation to the effort beginning FY'78, a memo was written by the manager of the DSN Data Systems Section (338) that discussed: (1) CCM selection; (2) development of specifications; and (3) support effort. The support effort included: Hi Rel packaging; Hi Rel ICs (current SSI SMI Standard), Applications Data, establishment of documentation standards, main frame packaging, parts failure analysis, and ROM and PROM programming and configuration control.

A CCM implementation schedule was prepared by the Command Systems and Computational Module Group and reviewed by the Telecommunications Science and Engineering Division (33). The purpose of the review was to show the schedules intended to accomplish the CCM implementation. The areas of activity were broken out as in the aforementioned memo. Schedules of the activities were proposed, and an outline expansion of the activities was included. The plan is to make the CCM selection first and overlapping this closely by an update of the standard IC Hi-Rel program. For the Hi-Rel ICs the plan is to survey the potential users to determine the needs. Other effort would be directed to determine the applicability and the completeness of the current list of standard ICs, taking into account the newer chips available from industry. The plan also calls for making recommendations for updating the standard Hi-Rel IC list and for establishing procedures to accomplish such an update.

Support services were put on the schedule to show when they would be available. It is intended that Section 338 will be the focal point of the gathering and disbursement of applications information. It is expected that as successful applications are documented, other designers can benefit from such experience. In providing this service, this Section will set up a file and a distribution system.

Packaging support will be provided to assist designers in establishing their configurations, or they can use the documented designs of existing equipment. Also, a prototype wire wrap service will be established to provide the capability to receive a wire list (in prescribed format) and deliver wrapped boards.

Parts testing from SSI to LSI (CCMs) will be provided. Section 338 currently has a tester to test the standard line of chips. The plan is to procure a tester to test the CCM chips that will be selected as standard for DSN applications. In order to keep track of the reliability of the parts supplied to the DSN, parts failure analysis program will be set up to look for trends that can provide information to correct any failure tendencies. This program will require coordination with the DSN operations to make certain that adequate data will be available from which to draw credible conclusions.

Development system will be necessary in the support of program generation for the new CCM CPU chip. After the standard CCM family is established, a development system will be procured. An additional feature designed in the development system will be the capability to program ROM and PROM modules as the last step of program development.

III. CCM Standard Selection Process

Meetings were held with JPL personnel to become familiar with the previous work performed in evaluating microprocessors. Using this information, an initial version of selection criteria was generated. Because the standard family is to be used by Division 33 and the DSN Engineering Section (355), a CCM Selection Recommendation Committee was called together of representatives selected by the managers of those sections who felt a need for a vote in selection of a standard line. The first of many meetings of the committee was held 9 March, 1978, with subsequent meetings being held weekly. Much of the first meeting and parts of subsequent meetings was spent in discussing why CCMs were selected and what configuration they should take. Many question were raised on what policies should govern their use. As a result of these concerns and to allow the committee to proceed with the business necessary to make a selection for a standard CCM family, another smaller committee (with representatives from several sections) was formed to work out a policy governing the use and application of CCMs within the Division. When completed, the CCM policy will be reviewed at the division level before approval is granted. With the formation of the second committee, work has proceeded in preparing for a selection of a standard.

In order to gather data for the selection process, an engineering survey of requirements and an industry survey of capabilities were initiated. The Engineering Survey was made of projected needs within Division 33 and Section 355. The survey was in the form of a memo with attached response forms. It was sent to all group supervisors of Division 33 and Section 355. The memo requested a list of which applications were anticipated (dedicated controllers, or data processors, or computations), the word length needed, the operating speed requirements, the microprocessor characteristics desired, and the software support needed.

There were 15 responses to the engineering survey from 35 requests that went out to all group supervisors of the involved

sections. The responses did provide a rather broad set of requirements. The respondents indicated: an intent to use the CCMs either dedicated controllers or for performing data processing or compilation; reserved opinions that an 8080 would be fast enough to do the job, with some definite opinions that it would not; most indicated that an 8-bit word would be satisfactory, and two indicated a need for a 16-bit word. No definite conclusion could be reached regarding functional characteristics priorities. It was generally considered that the software support was at least as important as the choice of hardware characteristics offered.

The industry survey is being sent to 12 manufacturers. The type of information being requested includes; chip availability (of the line considered most likely candidates to date, i.e. 8080A, 6800, Z80 and 6502); capability to meet various screening levels; flexibility in meeting JPL/vendor interfacing needs; vendor history and future planning; availability of software support tools; availability of test equipment (could be from other than CCM manufacturer, i.e., Tektronix); description and outline of courses of instruction provided by vendor; and expected deliveries of chips. This survey went out to industry on 4-18-78 and responses are due by May 11, 1978.

Consultants have been contacted and have been and will be used to provide additional insight into certain aspects of the selection criteria that need their expertise. The group supervisor of the Parts Engineering Group of the Electronics Parts Engineering Section (365) and two other members of his group came to one of the CCM selection recommendation meetings and explained the spectrum and ramifications of screening options available from industry (MIL-M-38510D, Class B to Commercial, manufacturers Spec Sheet).

Members of the committee have come with various backgrounds to apply to the selection process. Some members have direct design experience; others are acquiring experience from manufacturers courses and seminars; and some are gaining experience by purchase use and study of kits available from the various manufacturers. Insight into the technology is also being gained by investigating, by studying, and by experiencing demonstrations of test equipment that are available.

The proposed CCM Selection Criteria have been discussed at great lengths at the several meetings. The main topics addressed are:

(1) Projected Availability

> To be assured of multiple sources and gain confidence in the vendors market strength and expected longevity.

(2) Vendor Capability and Responsiveness

> To feel confidence in the ability of a vendor to meet JPL's QA and support needs as they might arise.

(3) Device Applicability/Adaptability

> To provide a measure to assess how well the microprocessor meets the JPL requirements including the M and O aspects of long-term use.

(4) Availability of Support Tools

> To be sure that the products are sufficiently supported in the testing and program development areas.

(5) Energy Conservation Considerations

> To compare how much power is needed to support the competing candidates.

For a complete copy of the criteria, see Table 2. The purpose of the selection criteria is to provide a common reference of comparison to support the justification of the final selection, when made.

In order to use the criteria in the selection process, a CCM selection Evaluation form is being developed. The purpose of the evaluation sheets is to provide an agreed upon reference for making the evaluation comparison. In effect the evaluation sheet defines how to implement the criteria in the selection process. The evaluation Sheets include weighting factors for each item in order to provide a resultant number for comparison among competing candidate microprocessor lines.

The plan is to complete the CCM selection by June 1, 1978 so that CCM standard can be published in order that use (as governed by the CCM policy) in new designs can commence at that time.

IV. CCM Policy Committee

As stated before, this committee grew out of the need for answering the basic questions concerning the establishing and implementing of a standard CCM family. The topics addressed by this committee were: (1) Why CCM? (2) What are CCM? (3) CCM Committee membership and responsibility, (4) Selection criteria, (5) Screening requirements, (6) Implementation of selected standard, (7) Use of CCMs before the standard,

deviation from standard, (8) Addition to Standard, and (9) Support services.

The CCM Policy Committee is made up of members of several Division 33 sections. Four meetings have been held to date to fill in the outline of topics provided above. A memo describing the topics will be published after the concurrence and/or approval of the Division 33 Office.

V. Conclusion

The effort to establish a DSN standard CCM family is essentially on schedule. Involvement in deriving this standard has been extended to all interested sections within Division 33 and also Section 355 to assure the widest possible acceptance of the standard and to insure the best possible fit of need (requirements) with capability of the standard set.

Table 1. Definition of a CCM family

CPU group	One to three chips.
CPU ALU REGs Control unit Bus interface Clock System bus	The CPU contains the usual computer functional capabilities; Arithmetic Logic Unit, ALU. Registers; and Control Unit. By their nature bus structures are necessary for interfacing. For our purpose the address bus should accommodate memory up to 64K bytes. A clock is necessary (this is included in some chip types). The Instruction Set should be one that can efficiently support DSN requirements. This will require some bench mark comparisons.
Memory group	
ROM RAM EPROM	The memory will be implemented, according to need and cost, from a selected set of the various configurations available: PROM, EPROM, ROM and RAM.
I/O	Single chip per function except perhaps SIA.
Peripheral controller Interrupt Floppy disk CRT Cassette DMA controller RS 232 IEEE 488 DSN 15-line SIA Floating point Peripheral interface adaptor Bidirectional bus driver (I/O support)	These functions are self explanatory and are the various interfaces that should be made available to support the expected needs of the DSN applications.
Special functions	May be a group of chips.
Display AD/DA FFT Timers (programmable)	Displays will probably be needed for stand alone troubleshooting of the CCM built equipment. For interfacing with analog equipment, AD/DA converters will be necessary. For Fast Fourier Transforms applications the FFT chip will save on hardware and the software development efforts for this special computation. Timers can provide a range of controlled timing signals to reduce the burden on memory and reduce operating time.
Support	
SSI and MSI ICs	The above-mentioned chips will not be sufficient to implement the circuit functions that will be needed. Therefore, the standard IC line will be used to supply this need.

Table 2. Proposed CCM selection criteria

- Projected availability
  - Multiple sources
    - Basic CPU
      - Identify sources who produce using identical masks
      - Identify sources who produce on cross license
    - Peripherals for DSN system requirements
      - (Same source question asked under CPU for each DSN identified peripheral)
  - Current history (for each device in Part I)
    - How long a producer
    - Principal user (which segment of industry)
    - Current % of sales to military applications
    - Currently producing to MIL-STD-883/B or not
  - Future projection (for each device in Part I)
    - What is the expected market life
    - What is expected date of next generation
    - Vendor estimated delivery time cycle
    - Cost to reproduce line (when obsoleted by mfg.)
  - Financial health of vendor
- Vendor capability and responsiveness
  - Adequate process controls (QA survey of candidates)
    - Permit JPL visibility into process
    - Supply desired test data
  - Responsiveness
    - Interest in JPL
    - Provide special device symbolizing
  - User publications
    - Accuracy
    - Readability
    - Completeness
- Device applicability/adaptability
  - Requirements
    - DSN applications
      - Dedicated controller
      - Data or file manipulation
      - Computation
    - DSN HD/SW compatibility (with existing equipment)
  - Device functional capability (to be evaluated)
    - Speed
      - Clock rate
      - Clock cycles/instruction
      - Number instructions/real-time function (using bench mark tests)
      - I/O capability
      - Specific characteristics
    - DMA
    - Interrupts
    - Universal interface
    - Compatible bus structure
    - Software/hardware applications adaptability
      - Software/hardware tradeoffs
      - Board real estate needs (chip count)
        - Level of integration of family
        - Completeness of set
      - Interface/debug simplicity
        - Timing
        - Interface controls
        - Interrupt
        - Single-step operation
    - Life cycle cost considerations
      - Future expansion of compatible family
      - Initial cost HW/SW vs. support cost
- Availability of support tools
  - Software development aids
    - Cross assemblers
    - Assemblers/translators
    - Editor
    - Debug/monitor
    - Linking loader
    - Simulators
  - Software packages
    - Development system
      - Operating system
      - Utility routines
  - Hardware development aids
    - Development system
      - Operating system
      - Emulators
  - Personnel training aids
    - Self teaching
    - Courses
    - Consultants available
    - Manuals
  - Device testing aids
    - Test equipment available
    - Degree characterized
- Energy conservation considerations
  - Number of power supplies required
  - Total device family power requirements

Implementation of the Radio Science Subsystem in the DSN

K. R. Kimball
DSN Data Systems Section

The DSN is implementing a new subsystem for support of radio science data acquisition requirements beginning in late 1978. The article describes the functional characteristics of this equipment and utilizes the factors that were of major importance in the design and implementation approach.

I. Introduction

The Radio Science Subsystem becomes operational in mid-1978 at the 64-m Deep Space Stations and at the DSN's Compatibility Test Area (CTA-21). This is a new subsystem at these locations, and it is intended to bring radio science data acquisition and processing functions into a more operational and automated environment. The initial application of this equipment will be to acquire and record open-loop receiver data during the Pioneer Venus probe entry in December 1978, and during the Pioneer Venus orbiter planetary occultations.

II. Functional Overview

A high-level block diagram of the Radio Science Subsystem equipment is shown in Fig. 1. At the 64-m DSS sites, the subsystem is composed of two functionally independent elements known as the Digital Recording Assembly (DRA) and the Occultation Data Assembly (ODA). At CTA-21 the subsystem consists of a DRA and a bandwidth reduction filter (BFR).

The DRA is a general-purpose wide-band digital recording system that can accommodate data recording rates from 150 kbits/s to 5×10^6 bits/s. The DRAs at the DSS sites will be used primarily for data recording, whereas the DRA at CTA-21 constitutes a centralized playback facility. The recording medium for the DRA is 1-in. wide instrumentation tape.

The ODA similarly performs a data acquisition and recording function, but operates in the data rate range of 2 kbits/s to 640 kbits/s. These rates permit recording of data on standard computer tape. The ODA contains a standard DSN minicomputer (Modcomp II/25) that acts as the control and buffering device for the data acquisition/recording process. A primary impetus for implementation of the ODA is to transfer the on-site recording function from analog recorders to digital recorders; analog recorders have been used for recording occultation data for Viking and previous flight projects. The machines that have been used are becoming obsolete and unmaintainable, and the data flow to the end product has been cumbersome. By placing the original data on computer-type magnetic tape, the ODA implementation will streamline the data processing and handling necessary to produce a deliverable to the radio science data users.

The bandwidth reduction filter (BRF) located at CTA-21 has been implemented specifically for the purpose of processing the Pioneer-Venus probe entry data. It operates in

conjunction with bandwidth reduction software, which is to reside in the CTA 21 telemetry processor assembly. The bandwidth reduction hardware/software combination allow selected narrow "bands" of the wideband spectrum recorded on the DRA for extraction and recording on computer-compatible tape, which is the deliverable to the data user.

III. Key Functional Characteristics of the Radio Science Subsystem

A. DRA Functional Characteristics

The DRA has been designed as a general-purpose digital recording device (i.e., its capabilities are not tailored to the near-term radio science applications). A functional block diagram of the DRA is shown in Fig. 2, and a summary of functional performance characteristics of the equipment is as follows:

(1) Digital data can be recorded at any discrete rate between 150 kbits/s and 50×10^6 kbits. The average bit error rate of the overall recording/reproduction process is less than 1×10^{-6}.

(2) A dual-transport configuration (at the DSS sites) provides the capability for continuous (overlapped) recording. This configuration also permits tapes to be copied on site.

(3) Data is recorded on 1-in. wide instrumentation tape. A 28-track longitudinal recording format is used; 18 tracks are assigned to user data recording, two tracks are used to record redundant data time tags to a resolution of 1 μs, and one track is used for recording a servo reference tone. The remaining seven tracks are not presently used (electronics for these tracks have not been included in the equipment).

(4) Real-time visibility into the quality of the data recording is provided (i.e., data reproduction occurs in a read-after-write mode).

B. ODA Functional Characteristics

The ODA performs a set of functions which are more diversified and automated than those of the DRA. A functional block diagram of the ODA is shown in Fig. 3. The functions performed by the hardware/software combination of the ODA are summarized below:

(1) Narrowband analog output signals (up to 4) from the open-loop receiver can be digitized and recorded at a maximum aggregate sampling rate of 80K samples/s (e.g., 80K samples/s on 1 signal or 20K samples/s on 4 signals). Lower sampling rates at binary submultiples are also available. Samples are time-tagged and recorded on computer-compatible tape.

(2) The ODA exercises real-time control over the programmed local oscillator in the receiver subsystem; the local oscillator frequency is controlled such that the desired signals are translated into the baseband spectrum which the ODA has the capability of recording. The information necessary to perform this control function is received by the ODA via the high-speed data line (HSDL). These time-tagged predicts are transmitted from NOCC prior to the pass and stored on the ODA disc.

(3) Following the recording of the data, the tapes can be played back via the HSDL to NOCC, where they are incorporated into the network data log. Alternatively, tapes can be copied on site and mailed to NOCC for incorporation into the data records system.

IV. Design Considerations Related to Operability/Maintainability

The design of the radio science hardware and software has been influenced strongly by operability and maintainability considerations. Special attention was given to the man-machine interface in the design review process. Being that both the DRA and ODA involve extensive use of the magnetic tape recorders, the operational impact of such an implementation has been a legitimate concern to operations organization personnel. Given that the use of tape recorders is a "necessary evil" in the effort to implement the subsystem functional requirements, the design goal has been to minimize the amount of time required by operations personnel for control, monitoring, and maintenance of the equipment.

A. Equipment Control/Monitoring/Checkout Approaches

An important design objective for all new equipment being implemented in the DSN is to allow remote operation of the equipment from a single centralized location at the station; the long-term objective is to have a single operator capable of controlling all station functions. In the case of equipment where tape recorders are used, such as the DRA and ODA, this goal cannot be fully achieved in that an operator is required to periodically change tapes. A second design objective is to provide automated checkout capability in all new equipment. Both of these objectives have been partially achieved in the initial configuration of the radio science subsystem.

The ODA can be fully controlled and monitored from the DSS data system terminal (DST). Tape changing is the only operator action required at the ODA itself. There are no

operational controls or indicators built into the ODA special-purpose hardware. For pre-pass equipment checkout, the ODA operational software program is capable of performing an automated checkout sequence which verifies that all hardware is functioning properly.

The capability for remote control and monitoring of the DRA was included in the original equipment design, but the implementation was postponed due to (1) budgetary constraints and (2) the fact that the initial operational use of the DRA (for the Pioneer Venus probe entry) will be infrequent and of short duration. Implementation of a remote control/monitor capability, along with an automated checkout capability, is now planned to start in October 1978 and be operational by December 1980. The DRA presently contains the necessary front panel controls and indicators which allow the operator to easily set up and monitor the assembly performance.

B. Conformance to Operational Software Standards

During the design phase of the ODA software, a separate effort was initiated to establish a standardized set of operator input/output interfaces for all minicomputer-based assemblies at the DSS. These standards, which deal primarily with Operator Control Instructions (OCIs) and the format of status displays, were defined in time to allow their incorporation into the initial ODA software release.

C. Utilization of DSN Standard Hardware Devices and Components

The radio science subsystem hardware has incorporated standard DSN hardware devices and components wherever applicable. In the case of the ODA, the hardware consists of a standard DSN minicomputer (Modcomp II/25), standard DSN peripherals (disks, mag-tape units, keyboard-printer, and input/output hardware), and a single special-purpose hardware subassembly known as the occultation converter. This subassembly uses standard DSN packaging hardware, and approximately 95% of the components of the subassembly are standard DSN Hi-Rel parts.

A large portion of the DRA equipment consists of commercially procured hardware. Although it was not considered appropriate or feasible to impose the utilization of DSN standard hardware for this equipment, it was specified that certain critical elements be fabricated with commercially available Hi-Rel electronic components (screened to MIL-STD-883, Level B). Those portions of the DRA which were designed and developed at JPL use standard DSN packaging and components, except in certain areas where very high-speed logic operation was required.

V. Key Aspects of Implementation Approach

The hardware and software implementation activities for the radio science subsystem have generally proceeded smoothly and on schedule. The implementation has involved major contracts with system integration contractors. The implementation schedule and contractor performance have received a high degree of management attention throughout the implementation phase.

A. DRA Hardware Implementation

The DRA is a composite of commercial and non-commercial hardware elements. An outline of the significant aspects of the implementation approach is summarized below:

(1) A decision was made early in the implementation phase that the basic digital record/reproduce equipment should be procured commercially from a *single* systems contractor.

(2) A set of prototype DRA equipment was built and tested. Martin-Marietta Corp. (Denver Division) was selected to build the prototype set of the digital record/reproduce equipment. Martin selected the Honeywell Model 96 tape transport for use in the system. At the completion of the prototype evaluation phase, several functional design changes were specified to enhance the flexibility, operability, and maintainability of the equipment.

(3) A contract was let in January 1977 to Martin-Marietta for delivery of the production version of the digital recording equipment. The first system was delivered in August 1977, which was very close to the originally planned schedule.

(4) Those portions of the DRA equipment which are peripheral to the basic recording function were designed at JPL; an integration contract was established with Sperry-Rand Corporation (Univac Defense Systems Division) for documentation, fabrication, test, and integration of this equipment with the Martin-Marietta equipment. The fully assembled DRA's were tested and transferred to the DSN operations organization at the contractor's facility.

(5) As of April 1978, all DRA equipment was installed and operating at the respective DSS sites.

B. ODA Hardware Implementation

The ODA hardware implementation was much more straightforward than that of the DRA. Key aspects of the implementation activities are summarized below:

(1) All hardware elements except the occultation converter subassembly were standard DSN devices and were commercially procured.

(2) The occultation converter subassembly was designed at JPL, but no prototype was built. The documentation, fabrication, and test of the subassembly was accomplished under the same integration contract as the DRA equipment; all cables for the ODA were also supplied under this contract.

(3) Final integration of the complete ODA's took place at JPL, where the equipment was transferred to the DSN operations organization.

C. ODA Software Implementation

The ODA software implementation effort has also proceeded smoothly and on schedule. Key aspects of the implementation approach have been:

(1) The software was designed and implemented in accordance with DSN software methodology standards.

(2) A high-level program language (Fortran) was used wherever possible.

(3) Existing software elements from other DSS subsystems were incorporated as much as possible; specifically, much of the input/output software was transferred from the Metric Data Assembly with very little modification.

(4) The implementation effort was organized into a work breakdown structure, which was used to monitor progress toward completion.

(5) Software verification testing was performed in a hardware environment which matched that of the final DSS configuration as closely as possible.

The software is scheduled to become operational in June 1978. It is planned to reverify performance in the station environment at that time.

VI. Future Plans for Augmenting Subsystem Capability

Several efforts to enhance and expand the initial capabilities of the radio science subsystem are also well underway. A partial list of these plans is presented below:

(1) It is planned to increase DRA data rate capability by 33% by December 1980. This increase is necessary for support of VLBI recording functions, which require a 32M b/s recording rate at a tape speed of 60 in./s.

(2) The DRA will also be upgraded by December 1980 to provide remote control/monitor capability and automated checkout capability. This upgrade will involve a connection to the ODA computer for control/monitor data flow.

(3) The ODA will be upgraded (by October 1978) to provide the capability to monitor the actual frequency output of the programmed local oscillator in the receiver subsystem.

(4) Special-purpose hardware will be added to the subsystem by April 1979 which will accommodate the digitization and recording of medium-band occultation data from the Voyager Saturn ring encounter.

(5) Special-purpose hardware will be added to subsystem for acquisition and recording of VLBI data. This addition is also scheduled for April 1979.

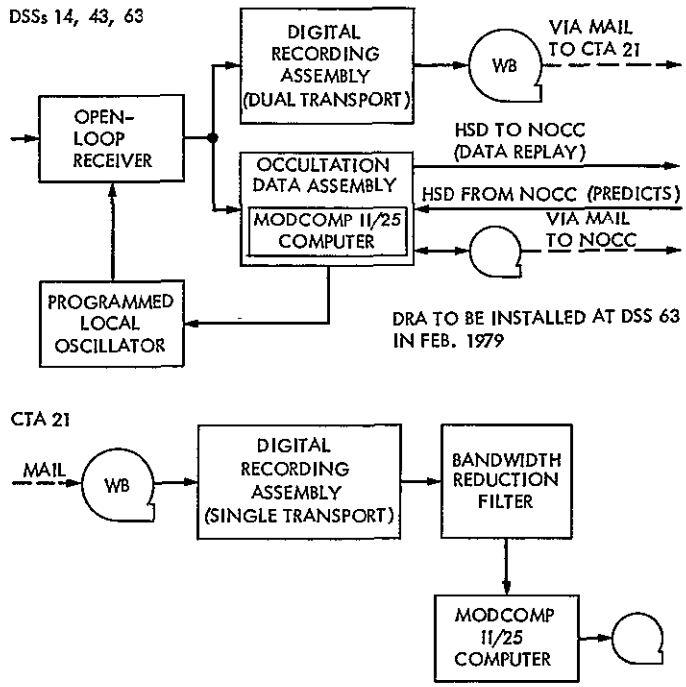

Fig. 1. Radio science subsystem equipment configuration (bold outlines)

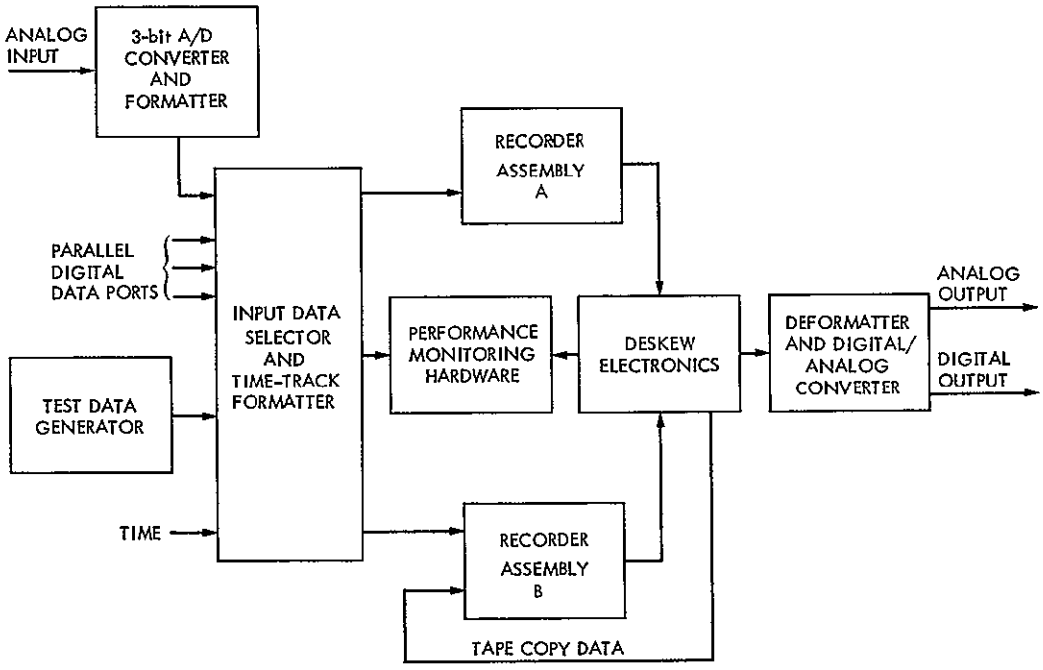

Fig. 2. Digital recording assembly functional block diagram

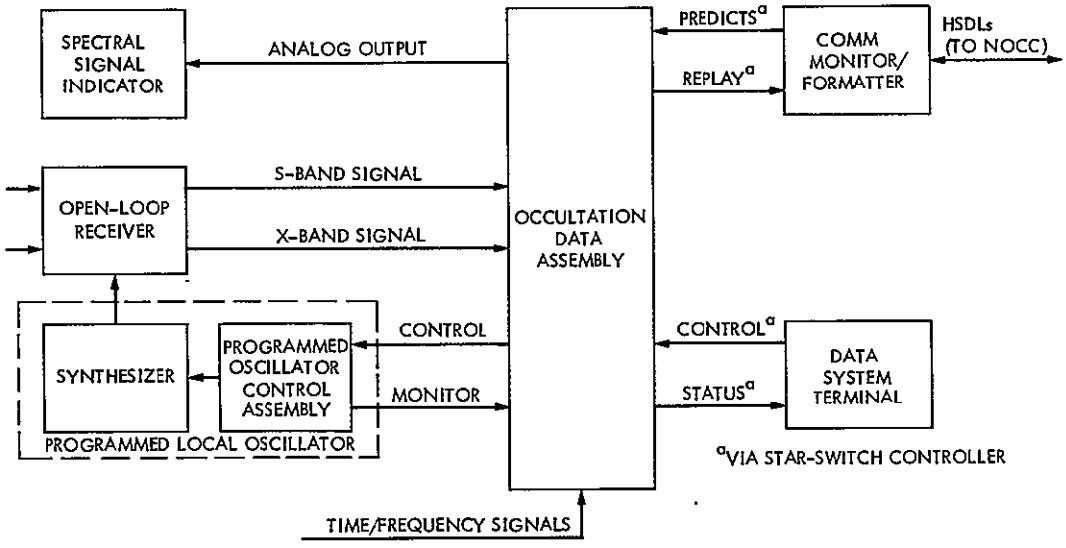

Fig. 3. Occultation data assembly functional interface

Energy Consumption Program—A Computer Model Simulating Energy Loads in Buildings

F. W. Stoller, F. L. Lansing, V. W. Chai, and S. Higgins
DSN Engineering Section

The JPL Energy Consumption Computer Program has been primarily developed as a useful tool in the on-going building modification studies in the DSN energy conservation project. The program simulates building heating and cooling loads and computes thermal and electric energy consumption and cost. It is a very low-cost code compared with other sophisticated programs such as NECAP (costs 1/200 of NECAP) or with other commercial ones such as ECUBE, TRACE, etc. The accuracy of computations are not sacrificed, however, since the results lie within ±10% margin compared to those read from energy meters. The program is carefully structured to reduce both user's time and running cost by asking minimum information from the user and reducing many internal time-consuming computational loops. Many unique features were added to handle two-level electronics control rooms not found in any other program.

I. Introduction

It is commonly conceived that buildings can be designed for minimum energy consumption if their thermal insulation is increased, window air leakage and lighting levels decreased, shading devices properly installed, heating and cooling equipment adequately designed and maintained, and their capacity fully utilized. These energy saving ideas and others should be coupled before implementation with cost of add-on materials, maintenance and operation costs, conforming to building codes, life styles, esthetics, etc. Design and operation of heating and cooling systems based upon conventional steady-state peak summer hour or peak winter hour, usually results in over-sizing of equipment and consequently overheating or overcooling of the space to be controlled. Over-designed systems, while they are occupying more space, always operate at lower efficiency and, in turn, require more energy to function.

The only effective way to study and design heating or cooling systems to minimize their energy consumption is to simulate the building hermal performance as accurately as possible. With the advent of high-speed digital computers, the above simulation process on an hourly basis for a full year and summation over many zones or buildings became feasible. In the last ten years, several thermal energy loads programs have been developed which vary in cost, availability to the user, program structure and assumptions used in computations, as indicated in Refs. (1), (2). Most of these programs, whether they are public or proprietary, are applicable to new buildings or to add-on or retrofit systems. The user may access these programs by: (1) purchasing public source codes, (2) through time-sharing by input data only to proprietary source codes or (3) input data only to the developer when dealing with complete proprietary codes. The disadvantages to the user in types (2) and (3) above are the lack of awareness about assumptions

and limitations made by the developer and the inability of the user to modify or improve the codes. Proprietary programs are usually written by architect/engineering consultants, heating, ventilation and air conditioning (HVAC) equipment manufacturers or utility companies. The cost to the user is usually included in their service.

As part of the Deep Space Network (DSN) energy conservation project (an extension of the old Goldstone Energy Project), approximately 23 buildings out of 50 were identified by a first phase study as major energy consumption buildings and were further put under investigation for a second building modification study. Both studies and energy-saving recommendations were made by architect/engineer firms in cooperation with the DSN engineering section. Furthermore, several energy-saving suggestions and proposals regarding building modifications have been presented by DSS Operations technicians and engineers as part of the personnel Energy Conservation, Awareness and Recognition Program, (ECARP). It has therefore become essential for economic and technical reasons to develop an in-house tool for assessment and evaluation of all building modifications.

To accomplish the task of finding this tool, a survey was made among approximately 14 codes for heating and cooling load simulation. The major programs among them are: (1) ECUBE 75, which was developed by the American Gas Association, (2) NBSLD, the National Bureau of Standards load calculation program, (3) NECAP, the comprehensive and expensive NASA's energy cost analysis program, (4) TRACE, the Trane Air Conditioning and Economics Program and (5) USPS, the United States Postal Services program developed by the General American Transportation Corp. Two architect/engineering firms, Keller and Gannon Consulting Engineers and Burns and Roe Inc., used ECUBE 75 in the initial phases of the building modification study.

Most of these programs tend to be of the same structure and are divided into four parts: (1) calculation of space (zone) loads for heating and cooling, (2) simulation of fan-coil (air handler) systems to meet above space loads, (3) simulation of primary equipment such as chillers, boilers, engines or heat pumps and (4) economic analysis of owning and operating HVAC systems. The Energy Consumption Program (ECP) is, however, mainly divided into the first three parts above, and the fourth part is replaced by the cost of thermal and electric energy as will be discussed further in the following sections.

II. Program Description

The calculation of energy requirements for heating or cooling in any building involves three major consecutive steps. First, the heat loss or heat gain to the space which is heated or cooled is computed. Second, the heating or cooling loads imposed on the heating or cooling coils inside the air-handlers are determined. Third, the energy input to all of the primary equipment or components constituting the air-conditioning system, such as compressors, heat pumps, boilers, engines, etc., is calculated. Each of these calculation steps may be carried out with various degrees of complexity and sophistication if more accuracy or refinement are required. The following describe the methodology used in each of the programming procedure and calculation steps.

A. Weather

Under this heading, only outside air dry-bulb temperature, cloud cover factors and wind speed are needed. For the outside air dry-bulb temperature, twelve representative days (one for each month) per year indicate the year's weather pattern. For each of these representative days, 24-h values of dry-bulb temperatures are needed. Cloud cover factors are determined by the ratio of actually accumulated daily solar energy incident on the location to that accumulated theoretically by ASHRAE model (Ref. 3). If no solar irradiation data were available, the cloud-cover factor would be estimated from past experience by the percentage of clear sky area on the selected representative day of the month. Twelve cloud-cover factors are required, corresponding to 12 months of the year. Wind speeds are required hourly for a complete day representing the yearly average wind speed in the location. All the weather data above represents the minimum information needed in order to achieve reliable results.

B. Transmission Loads Through Walls and Roofs

Some methods have been developed in the past such as "degree-days" method or "bin" method for proportioning the design load to provide hourly loads, and were found unacceptable with their gross approximations. On the other hand, sophisticated and time-consuming methods, such as using the transfer functions or response factors, are unfortunately very expensive and cannot be justified on the basis of the random changes of many other parameters in the system. This means that it is illogical to increase the accuracy of only one of the many sources of heating/cooling loads (i.e., the heat transmission through walls and roofs) while the other sources are subject to random changes with very large errors.

The methodology used in ECP is the Sol-air temperature method or total equivalent temperature method (TETD) as described in Refs. (3) and (4). It is a comprehensive yet easy to apply method. In the TETD method, the effects of outside air temperature and solar radiation intensity are combined into a single quantity. Walls and roofs are assumed homogenous with constant material properties that are determined in

advance by a special procedure. A special subroutine is written to determine analytically all the steady-state and transient heat transfer coefficients together with the phase angles needed to run the heat transmission computations.

C. Solar Heat Gain

Solar radiation values have been included in a subroutine using the well-known ASHRAE model described in Ref. (3) with ground reflectivity assumed at 0.2. The cloud-cover factors will modify these radiation values at a given month, day, or hour to yield the solar radiation values specific to the location. This procedure was found to be effective in reducing the solar radiation data or cloud data to be processed by the user.

D. Infiltration/Exfiltration Loads

The air change method was preferred over the crack method. For any space under consideration, the effect of repetitive opening and closing of doors or windows, leakage of outside air to or from the space through cracks, clearances, etc., is averaged by assuming an outside air change rate of 1.2 changes per hour. The heat loss or gain to the space is calculated accordingly, assuming quasi-steady-state conditions.

E. Internal Heat Load Profile

Only sensible heat gain is considered in ECP since the latent loads are often less than 10% of the total heat gain. The internal heat load is commonly composed of people heat gain, light, electronic equipment, mechanical/electrical equipment heat gain and other miscellaneous sources of heat gain such as process steam, etc. Each of these loads are calculated on an hourly basis for two-day types representing repetitive events for the whole year. Daytype (1) represents all working weekdays (approximately 251 days) and daytype (2) augments all weekends and holidays (approximately 114 days) to include 10 official holidays. The number of persons occupying the room, kilowatt rating of electrical, electronic equipment and lights are listed every hour for 24-h. Since some of the data collected under this internal load calculations section are considered "best" estimates, with varying degrees of accuracy, it was decided that the total equivalent temperature difference method, TETD, previously described under II-B is in fact an adequate methodology. No transient effects are considered for internal heat loads due to people, light equipment and hourly values were assumed causing quasi-steady state effects.

F. Architectural Data

The architectural and physical characteristics of the building (or zones) under study play an essential role in the first load calculation step. The data include building orientation, latitude, elevation, wall areas, glass areas, space volume, physical dimensions, outside wall solar absorptivity and the cross section description of layers constituting walls and roofs. The data are grouped by zone; the maximum number of zones per building is not allowed to exceed 8 in the program. If there are more than 8 zones, grouping of several zones into macrozones having the same fan-coil (air handler) arrangement may be done.

G. Shading Effects

Shading due to overhangs, side projections or adjacent buildings is handled in ECP as a fraction between 0 and 1 given by the user based on average shading conditions. The shading factors are used to attenuate the solar radiation falling on particular wall(s) of the zone under study. To reduce the manipulation process, various analytical methods used to compute the shaded areas, and their variations with hour of the day or sun's angles were not considered in ECP.

H. Loads Due to Neighboring Areas

The effect of a neighboring zone kept at a temperature different from that of the zone under study has been taken into consideration in ECP structure. However, ground floors were not included and were assumed always well insulated with no heat exchange to or from a basement zone.

I. Fan-Coil (Air Handler) Type and Arrangements

There are nine types of air handler arrangements that have been incorporated into ECP as illustrated in Fig. 1. These are: (1) single cold duct with terminal reheat at the zone, (2) dual duct multizone with mixing boxes or single duct multizone with mixing hot/cold air at the air handler section, (3) single cold duct with bypass control around the cooling coil and terminal reheat at the zone, (4) heat pump with bypass control, or single duct with alternately operating cooling and heating coils with bypass control, (5) two-level room (plenum and comfort air) with cold plenum air and comfort air modulated with terminal reheat, (6) two-level room with cold plenum air and comfort air modulated by a mixture of cold air with by-passed mixed air and terminal reheat, (7) two-level room with cold plenum air and comfort air modulated by mixing cold and hot decks, (8) single cold duct with fixed by-passed return air with terminal reheat and (9) two-level room with constant volume cold plenum air and variable volume comfort air at fixed hot deck temperature. For each building, the maximum number of air handlers allowed is 10. The simulation of air handler configurations is the second step of calculations. The ratio of outside air to total circulating air and the setpoint temperatures of both cold and hot decks are required as input. Also, for two-level rooms the maximum allowable plenum air temperature and the ratio of comfort air

discharge to total air discharge are required from the user. ECP also uses outside air economizer systems with various flow and temperature control mechanisms.

J. Time Clocks

The inclusion of time clocks in air-conditioning systems to control the operation (on or off) of air handlers has been presented as an energy saving suggestion. The energy saved by a time clock control is optimum for buildings that operate on the regular 8:00 a.m. to 5:00 p.m. schedule. The time clock on/off control is not applicable to light bulbs or electronic equipment since these will directly appear under the internal load schedules data.

K. Primary Equipment Performance

The user will identify under this heading: (1) the type or heating or reheating systems (electric, gas-fired or heat pump), (2) the arrangement of air handlers with the electric driven compressor/chillers (vapor compression refrigeration units) and (3) compressor stages and sizes. The maximum number of compressors is 10 and each compressor is allowed to be made of two stages with the second stage giving the full unit tonnage when on. Electric heaters (or boilers) were assumed having a constant 80% efficiency at all loads. The coefficient of performance of heat pumps and the vapor compression refrigeration cycle were assumed as a fraction (~50%) of the ideal Carnot's working between the refrigerant's evaporator and condenser temperatures. Partial-load performance was assumed unchanged from 100% full load until 40% of the full load. Beyond the 40% full load limit, the electric energy consumed in refrigeration units or heat pump units was assumed constant.

L. Energy Consumed by Auxiliary Equipment

Auxiliary equipment are defined in ECP as that equipment outside the air-conditioned zone space, which is necessary for building operation but does not affect the heating and cooling load calculations. Auxiliary equipment includes air handler fans, condenser fans (air cooled), condenser pumps (water cooled), boiler pumps, building external lights, etc. The energy consumed by this equipment directly affects the watt-hour meter reading. Their load profile and schedule can be quite complex, and their consumption is modeled in ECP by name-tag capacities only. For all pumps or fans, no allowance was made for partial load performance, or flow pressure and discharge variations.

M. Energy Cost

The unit cost of both thermal and electric energy as purchased from a utility company or generated on site is used to estimate the monthly or yearly cost of energy consumed.

In addition to the above component description, many default values are assigned to fill unknown input data. The program is written in FORTRAN IV computer language using the EXEC-8 codes of UNIVAC-1108 machines at JPL. The program output results include: (1) input data "echo" for the user to see all the entered data in tabulated forms, (2) diagnostic messages, (3) itemization of zones heating/cooling loads, per hour for the two-day type and for each month, (4) air handler and primary equipment energy consumption each hour for the two-day types and for each month, (5) hourly supply air temperature to each zone, (6) maximum and minimum air temperature supplied to each zone per month per year, (7) final equilibrium room temperature, (8) mixed air temperatures at the air handler, (9) peak hourly heating and cooling loads per zone, (10) peak hourly heating and cooling loads per primary equipment, (11) summation of monthly and yearly energy consumption and cost. For newly constructed buildings, the program also prints the design cooling tons of refrigeration, the design supply air discharge and the design outside air to circulating air ratio needed for ventilation.

III. Flow Chart

Fig. 2 illustrates briefly a block diagram of ECP. Each block depicts one phase of the ECP program: (1) input data, (2) ECP computations, (3) design considerations and (4) output. The subtitles in each block give the reader a general idea about the sequence of events within each phase of the ECP program.

The program documentation package includes user instructions for entering data, detailed flow charts, algorithm descriptions, input/output format and a solved example for illustration. The user would expect a cost of CPU time of approximately $5/building zone, excluding printing cost. In this regard the program is considered inexpensive compared to other comprehensive codes, such as NASA's NECAP which costs approximately 200 times the cost of ECP. The accuracy of ECP, on the other hand, has been found very reasonable for engineering purposes, since predicted energy consumption values are only off the watt/hour meter values by a very narrow margin (±10%). This margin was tested at Goldstone Complex with zones that vary in number from one to fifteen per building.

References

1. *Technical Guidelines for Energy Consumption*, Final report No. AFCEC-TR-TI-12 for period Dec. 75 – March 77. Published Buildings Environmental Div., NBS Dept. of Commerce, Washington, D. C.
2. Chen, S. Y. S., "Existing Load and Energy Programs," *Heating-Piping and Air Conditioning*, December, 1975, pp. 35–39.
3. American Society for Heating, Refrigeration and Air Conditioning Engineers (ASHRAE) *Handbook of Fundamentals*, 1972.
4. Threlkeld, J. L., *Thermal Environmental Engineering*, Prentice-Hall Inc., New Jersey, 1962.

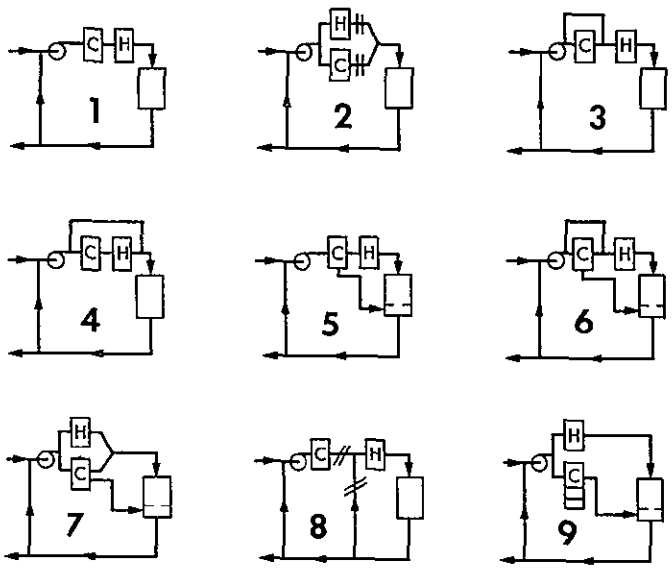

Fig. 1. Schematic of 9 air-handler-type arrangements

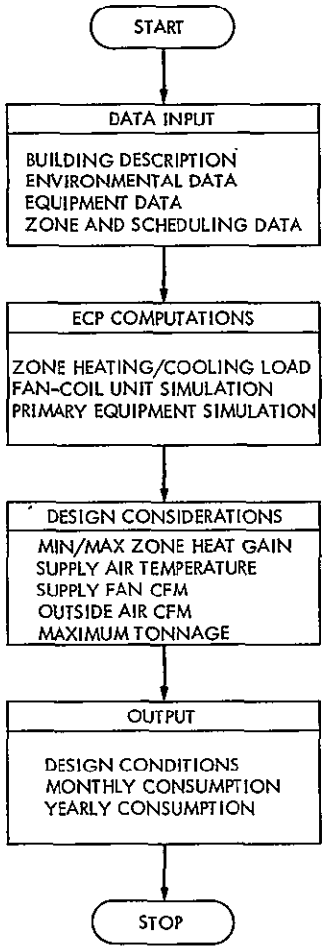

Fig. 2. ECP block diagram

Network Telemetry System Performance Tests in Support of the Mark III Data System Implementation

R. D. Rey and E. J. Nipper
Network Operations Section

This article presents a description of Network Telemetry System Performance Tests that were executed throughout the DSN in support of the Mark III Data System Implementation.

I. Introduction

System Performance Tests (SPTs) are executed throughout the DSN whenever a modification is made to the Network Telemetry System that affects its performance. This is the case when new software or hardware is added. For the Mark III Data System Implementation both new hardware and new software were added to the Telemetry System. The purpose of this article is to describe the effort that was undertaken in executing telemetry SPTs throughout the DSN after the implementation of the Mark III Data System.

The philosophy and the objectives of the SPTs will be discussed to demonstrate the benefits gained by performing these tests. The telemetry SPTs will be described along with the test procedure and test software. Finally, a summary of the results and the status of the Network Telemetry System will be presented.

II. Objectives of System Performance Testing

The development of test procedures, test software and the execution of the telemetry SPTs are performed to accomplish certain objectives. The overall objective is to verify that the Network Telemetry System meets specified operational capabilities. These capabilities are defined in Refs. 1 through 3. The SPT must verify that the telemetry system configuration and interface requirements are satisfied and must also evaluate that the Telemetry System meets specified performance criteria.

The telemetry SPT is designed so that it can be used to locate or diagnose system problems. The problems may occur during the installation of new equipment or software, or they may be the result of system failures uncovered during DSN operations.

Use of SPTs for Pre-Pass Preparation Tests (PPTs) required that the SPT procedures be modular. The modularity feature of the procedure enhances the capability to test modifications that may affect the capabilities of the Telemetry System.

An additional objective of SPTs is to provide some system level training of station personnel. The SPTs are prepared so that the configurations used are as near as possible to configurations used during spacecraft tracking operations. Thus, by using the test procedures station personnel can gain experience in operating the Telemetry System.

III. Test Configuration

Figure 1 shows the general telemetry test configuration for both a 64-meter and 26-meter station. Also shown in Fig. 2 are the 26-meter conjoint differences. Simulated data are generated in the Simulation Conversion Assembly (SCA). The simulated data called for in the tests are a pseudo-noise (PN) sequence at the required bit rate. The SCA modulates these data with a square wave subcarrier, which is set at a specific modulation index by using a wave analyzer at the Receiver (RCV). The data on the subcarrier phase modulate a carrier generated in the test transmitter, which is interfaced to the RCV through ambient load, and the test transmitter signal level is adjusted to obtain the required signal-to-noise ratio (SNR). The SNR is accurately measured at the Y-factor detector.

The telemetry data handling software (DSN Program Library Software No. DMT-5085-OP) resides in the Telemetry Processor Assembly (TPA), which is a general purpose minicomputer. The TPA software can be configured to process data using either of two telemetry channels. Acceptable data rates and coding schemes are as follows:

(1) Channel 1 (CH1) short constraint length convolutionally coded data, with data rates from 10 to 115.2K bits per second (BPS), are decoded in the Maximum Likelihood Convolutional Decoder (MCD) before being passed to the TPA for formatting. Long constraint length convolutionally coded data up to and including 4096 BPS are software sequentially decoded and formatted by the TPA. Block coded data (CH1) up to 2.0K are passed directly from the SSA to the TPA for software decoding and formatting. Uncoded data from the SSA are sent directly to the TPA for formatting.

(2) Channel 2 (CH2) block coded data in CH2 are decoded in the Block Decoder Assembly (BDA) at rates up to and including 16.0K BPS and sent to the TPA for formatting. Long constraint length convolutionally coded data (CH2) up to and including 4096 BPS are software sequentially decoded in the TPA and formatted for output. Uncoded data from the SSA are sent directly to the TPA for formatting.

High-Speed Data

The TPA formats the data into a High-Speed Data (HSD) block, which it then encloses in the data portion of a Sub-System Block (SSB) for routing through the Star Switch Controller (SSC) to the Communications Monitor and Formatter (CMF). The CMF then strips off the SSB to leave an HSD block. The HSD block is simultaneously transmitted to the GCF and recorded on a digital ODR at the CMF.

Wide-Band Data

The TPA, in order to accommodate data rates greater than 7.2K from the BDA and MCD, has the capability of formatting and outputting Wide-Band Data (WBD) to the GCF. These data can also be recorded on a High Rate Operational Data Record (HODR) at the TPA.

The transmitted data blocks are described in detail in Ref. 4. The specifications placed on the telemetry system by these requirements are rigorously tested and adhered to.

Interface information in the form of SSBs is transmitted from the TPA through the Star Switch Controller (SSC) to the Digital Instrumentation Subsystem (DIS). The SSBs are shipped via HSD blocks and validated at the System Performance Test Assembly (SPTA). The validation consists of checking subsystem status and alarms.

The SPTA is used to process the data for the tests using the Telemetry System Test Software, DMI-5124-TP and Performance Test Procedures, which runs on disc. HSD blocks are simultaneously transmitted to the SPTA via the CMF and are recorded on a digital ODR at the CMF. WBD blocks are simultaneously transmitted via WBD lines to the SPTA, and recorded on digital recorders at the TPA. To use the SPTA as the processing computer, the TPA HSD/WBD lines are patched at the station communications center to the SPTA HSD/WBD lines.

IV. Telemetry System Test Software

Telemetry SPT Program DMT-5124-TP resides in the back-up CMF, which will be called the SPTA. The Telemetry System Test Software and Performance Test Procedures are combined on disc with other test programs and procedures to make up System Performance Test Program DMI-5001-OS. Only DMI-5124-TP, the Telemetry Test Procedures, and the Executive Program, DMI-5122-TP, with which they interface, are described in this report.

The Executive Program interfaces with the Telemetry String to be tested through the HSD/WBD on-site Ground Communication Facility (GCF) equipment and the Star Switch Controllers (SSCs). These interfaces allow the Telemetry String and its operational software to be tested in an operational environment. In other words, the telemetry system is tested relative to the data as if they were leaving the station.

Telemetry data HSD/WBD blocks may be output on the line printer for visual inspection, although normal operation utilizes the software to output and calculate the following:

(1) Standard block header information consisting of source, destination, data-dependent type (DDT), user-dependent type (UOT), spacecraft identification, day of the year and the time of the block formation.

(2) Formatted configuration and lock status with receiver AGC or signal level in dBm, and bit SNR in dBm.

(3) TPA and channel number bit error rate (BER) or word error rate (WER), erasure rate and symbol error rate measured over a given period.

(4) Signal level for the y-factor calculation computed by the software relative to the input parameters.

In the bit error rate test, the data may be either hardware PN, software PN or fixed pattern. The test software synchronizes to the data and does a bit-by-bit comparison. The WER/BER accumulation may be over any chosen interval resulting in a grand total output, while statistics can be displayed as interim summaries at intervals as required. Also grand totals will be output immediately if test is terminated early.

As the test is being executed, the test software will detect errors and display error messages as follows:

(1) The Δt between telemetry HSD/WBD blocks not within specified tolerances.

(2) Incorrect data type, as evidenced by the inability to achieve synchronization.

(3) Binary time and millisecond clock differences.

(4) BSN errors.

(5) GCF errors.

(6) Message buffer overflow (inability to output error messages because of buffer overflow).

An additional capability of the SPT software is the real-time dumping of High Speed Data during operations. Both incoming and outgoing real-time data can be dumped to a line printer in either Hex, octal or english formats. This allows station personnel to perform diagnostics of system problems during tracking operations.

V. Test Procedure Format

Performance Test Procedures are combined on disc with other test programs and procedures to make up System Performance Test Program DMI-5001-OS (Refs. 5 and 6). The procedures on Disc are handled by the TLM and EXEC Test Programs in the same manner as operator inputs from any other I/O device. A list of these directives, normally referred to as Operator Control Inputs (OCIs), is contained in the Software Operators Manual (SOM) which accompanies the programs.

The programs will operate in any of the station MDS computers; however, because of interface requirements for testing, such as HSD/WBD, printer/plotter, etc., the procedures were designed for operating only in a Communications Monitor and Formatter Assembly (CMFA), which, when operating as the test computer, is the aforementioned System Performance Test Assembly (SPTA).

The overall test is divided into three main tests. Each test contains a number of subtests that are modular so that any test can be run independently. This allows the procedures to be used for pretrack readiness tests (PRTs), trouble-shooting and new equipment or software tests. It also allows for reduced testing when resources so demand.

The three tests are the Configuration and Interface Tests, the Telemetry Performance Tests and the Operational Data Record (ODR) Test. These tests all call out a Test Preparation procedure as needed.

A. Configuration and Interface Tests

Configuration and Interface Tests are performed on the two telemetry channels. They test the telemetry operational software and hardware interfaces using all operational configurations planned for project support. These tests also check HSD and WBD interfaces and HSD/WBD block configurations. The blocks are verified by inspection of the formatted block headers, and configuration and lock indicators are formatted by the test software for ease of interpretation. Monitor messages verifying that the initialization, status and calculation messages are correct is also accomplished in these tests.

Although the AGC/dBm conversion test and the signal-to-noise ratio (SNR) calculation test are separate sections of the test procedure, they are normally run with the configuration and interface tests. The AGC/dBm conversion test verifies that software can perform a correct conversion from AGC volts to signal level in dBm. The conversion parameters are entered in the program, and accuracy of the conversion is checked to see that it is within ±0.05 dBm. The SNR calculation test verifies the accuracy of the software SNR estimator routine. SNRs of varied dB are set up using the y-factor machine. The calculation is verified to be accurate within ±0.3 dB.

The important aspect of the configuration and interface tests is that they insure that the functional capabilities of the Telemetry System exist.

B. Telemetry Performance Test

The Telemetry Performance Tests determine the capability of the Telemetry System to meet the DSN support performance requirements. These tests are designed to evaluate telemetry performance at threshold SNRs with CH1 and CH2 configured in Project support modes.

The test measures the output data Word Error Rate (WER) for coded data and the Bit Error Rate (BER) for uncoded data given an input data signal-to-noise ratio. The results are compared against predicted BER/WER, and if they are within given tolerances, the performance is considered to be acceptable. The predictions are based on available mathematical models of the Telemetry System (Refs. 7 through 9). These models are evaluated using a computer program. This program is called the Telemetry Analysis Program, and is used primarily to generate performance parameters for the telemetry SPTs.

A strong signal test is always executed prior to a weak signal test. The strong signal test is run to detect gross errors in the HSD/WBD blocks that would invalidate the weak signal tests, which take much longer to perform. The strong signal test also confirms the set up configuration and interfaces.

The weak signal performance tests are run for all operational configurations and data rates. An accurate SNR is set using the Y-factor machine, and the WER/BER is measured and recorded. The SNR calculation is also checked for accuracy.

C. Original Data Record (ODR) Tests

Communications Monitor and Formatter (CMF) ODRs, TPA temporary ODRs (TODRs) and High Rate ODRs (HODRs) are verified in the following way. An ODR or HODR is generated, and a short bit error test (BER) is run to insure the quality of the data. Also TODR operational capability is exercised to ensure the ODR or HODR will contain TODR despooled data. The ODR or HODR is then played back, and a BER test is conducted to verify data content and quality.

Analog tapes are generated at a specific SNR and then played back. The tape output SNR is obtained and compared against the set SNR. The tape output SNR must be no greater than 1 dB, less than the set SNR.

Data reporting is accomplished by using data sheets which are automatically output by the test procedure. These are presented in Fig. 3.

Test results are analyzed and recorded by the telemetry System Cognizant Operations Engineer (SCOE).

VI. Mark III Data System Implementation Network Telemetry Test Status and Results

Execution of network telemetry SPTs for the Mark III Data System Implementation began November 1976 with a completion date of March 22, 1978. The tests were carried out at the following stations:

(1) Goldstone, California: DSS 11, DSS 12, and DSS 14.

(2) Tidbinbilla and Honeysuckle: DSS 42, DSS 43, DSS 44.

(3) Madrid and Cebreros, Spain: DSS 62, DSS 63, DSS 61.

Station personnel, while conducting the SPTs, uncovered problems and gained working experience and knowledge about the system. Upon completion of the SPTs, station personnel sent the results to the telemetry SCOE. The data were recorded and analyzed to evaluate the status of the Telemetry System.

VII. Summary

All Telemetry SPTs for the Mark III Data System have been completed throughout the DSN, and the performance as demonstrated by the Telemetry SPTs has been acceptable.

Also, SPTs were designed for continued use as System Maintenance and Operations Tests, which can be used to maintain station performance and diagnosis system problems during both tracking operations and maintenance periods. Furthermore, the SPTs can be used for station readiness tests and, with some modification, can be used to provide a computer-aided countdown capability.

References

1. *Deep Space Network System Requirements* – Detailed Interface Design, Document No. 820-13. Jet Propulsion Laboratory. Pasadena, Calif., Feb. 1, 1971 (JPL internal document).

2. *Deep Space Network System Requirements,* Document No. 821-2; Rev. A, Chg. 1. Jet Propulsion Laboratory, Pasadena, Calif., Feb. 1, 1977 (JPL internal document).

3. *Network Operations Plan for All Active Projects including Pioneer Venus* (JPL internal document.)

4. *Telemetry System Wideband and High Speed Data Interfaces,* Document No. TLM-3-5, listed in Deep Space Network System Requirements-Detailed Interface Design, Document No. 820-13. Jet Propulsion Laboratory, Pasadena, Calif., May 1, 1973 (JPL internal document).

5. *DSIF Standard Test Procedure Telemetry System Performance Test.* Document No. 853-96; 2A-22. Jet Propulsion Laboratory, Pasadena, Calif., Jan. 16, 1978 (JPL internal document).

6. *DSN System Requirements Detailed Subsystem Interface Design,* Document No. 820-14. Jet Propulsion Laboratory, Pasadena, Calif., Feb. 1, 1978 (JPL internal document).

7. *Deep Space Network/Flight Project Interface Design Handbook,* Document No. 810-5 (Rev. D) (JPL internal document).

8. Edelson, R. E., *Telecommunications Systems Design Techniques Handbook,* Technical Memorandum 33-571. Jet Propulsion Laboratory, Pasadena, Calif., July 15, 1972.

9. Tausworthe, R. C., "Communications Systems Development: Efficiency of Noisy Reference Detection," in *The Deep Space Network*, Space Programs Summary 37-54, Vol. III, pp. 195-201. Jet Propulsion Laboratory, Pasadena, Calif., Dec. 31, 1968.

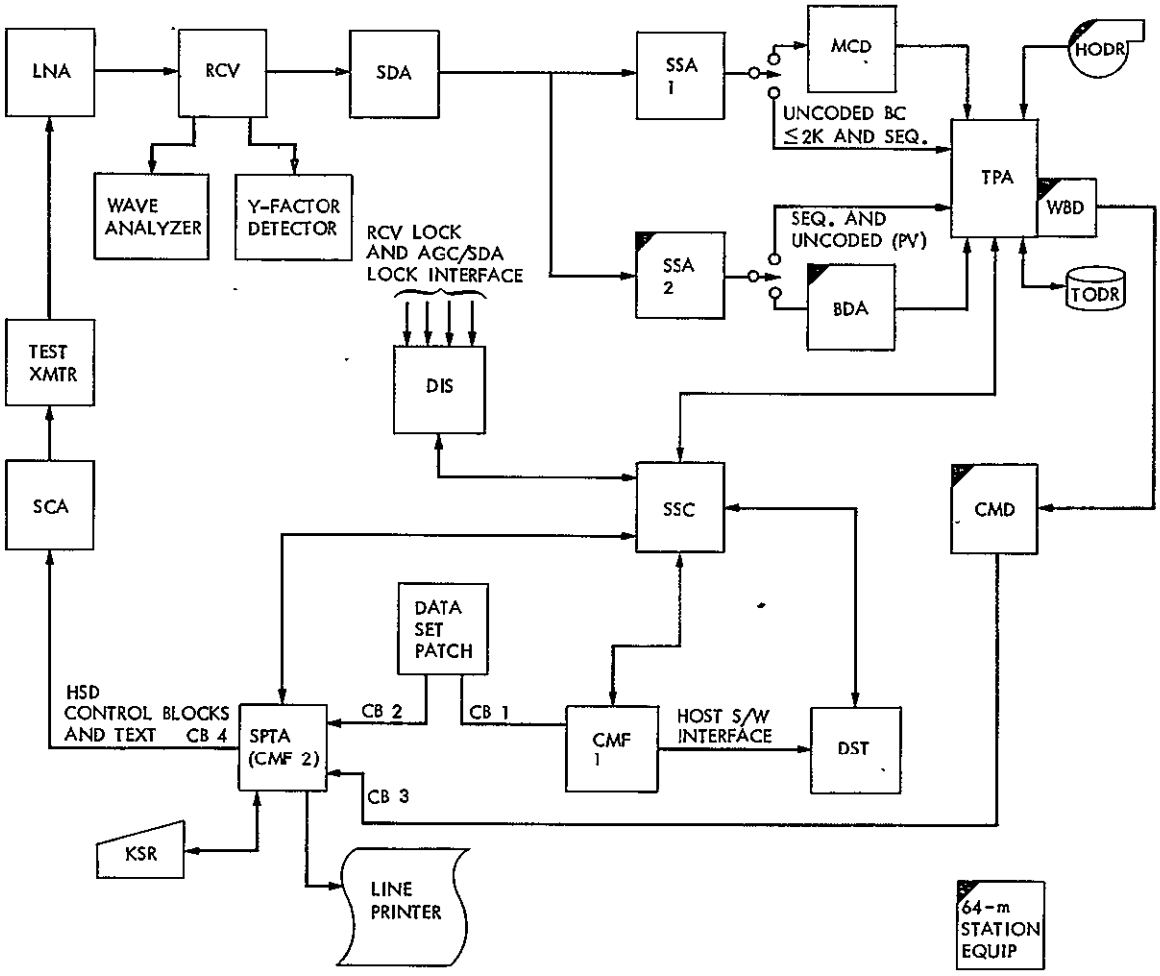

Fig. 1. 26/64-m telemetry SPT diagram (typical)

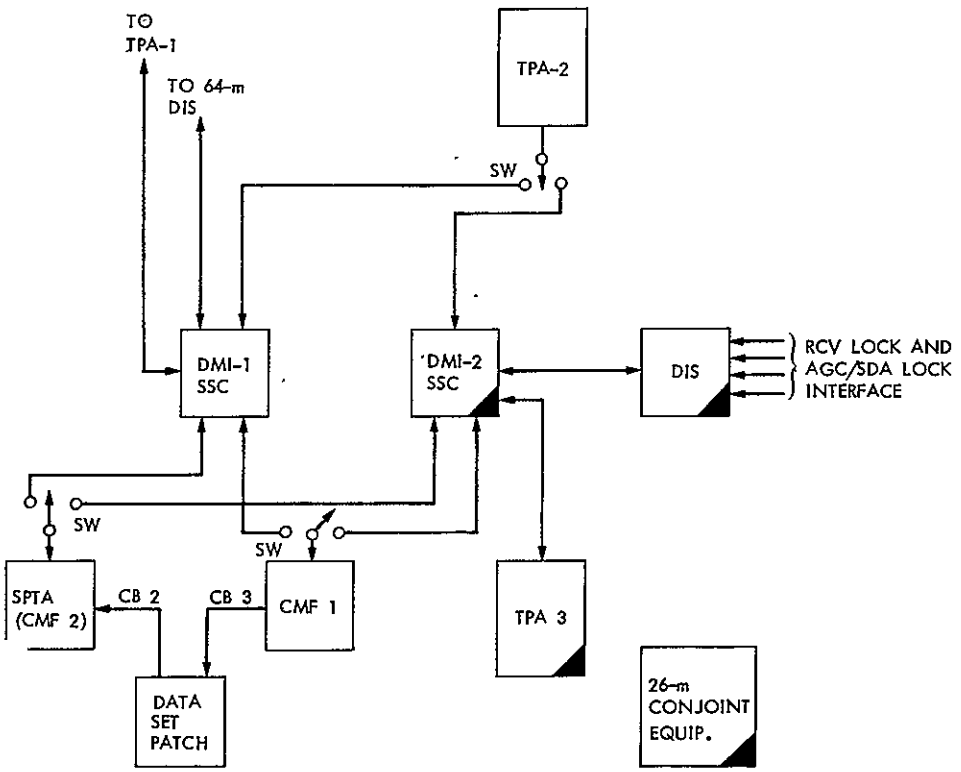

Fig. 2. 26-m conjoint interface differences

```
:1                PIONEER VENUS TELEMETRY TEST DATA SHEET
:0
:0    DSS-____  TEST CONDUCTOR_____________    TPA NO. 1.
:0                                             TIME:   ~
:0
```

TST NO.	BIT RATE BPS	SYMB RATE BPS	SYMB SNR IN	SPECIFICATION SNR	SPECIFICATION SER	SPECIFICATION ERAS RATE	ACTUAL RESULTS SNR	ACTUAL RESULTS SER	ACTUAL RESULTS ERAS RATE
1.	2048.0	4096.0	0.5	>-0.4	<0.084	<0.05			
2.	1024.0	2048.0	0.5	>-0.4	<0.084	<0.05			
3.	682.6	1365.3	0.7	>-0.4	<0.084	<0.05			
4.	512.0	1024.0	0.9	>-0.4	<0.084	<0.05			
5.	341.3	682.6	1.1	>0.4	<0.084	<0.05			
6.	256.0	512.0	1.4	>0.4	<0.084	<0.05			
7.	170.6	341.3	1.9	>0.4	<0.084	<0.05			
8.	128.0	256.0	3.2	>0.5	<0.060	<0.04			
9	64.0	128.0	2.9	>0.0	<0.060	<0.04			
10.	16.0	32.0	3.2	>0.0	<0.060	<0.04			
11.	8.0	16.0	5.2	>2.5	<0.030	<0.01			

Fig. 3. Data Sheet 1 — performance test data sheet (typical)

DSN "Load and Go" Pre-track Preparation for Voyager Support

T. M. Taylor
Deep Space Network Operations Section

The use of Level 4 Prepass Readiness Tests ("load and go" pre-track preparation) provides an efficient increase in Network productivity. Application of this method of operation was evaluated for Voyager support, so that increased tracking could be provided to all projects during a period of continued interproject conflicts.

I. Introduction

Pre-track preparation time is scheduled at each Deep Space Station (DSS) prior to direct spacecraft tracking operations. This time is used to configure and calibrate DSS equipment based on predefined "level of support" requirements. Reduction of this time increases the number of station hours available for spacecraft support, a measure of Network productivity.

Prepass Readiness Tests (PRTs), Levels 1 through 4, formally define pretrack preparation time and activities. Level 3 is normally used for prime mission cruise support, and Level 4, commonly called "load and go," is used for extended mission support. Due to DSS loading and interproject conflicts, the Voyager Project and DSN Operations mutually agreed to a trial period of Level 4 PRTs to evaluate the effectiveness and inherent risks involved during actual cruise phase support.

II. Procedure

The procedure followed was the same as that used for the Level 4 PRT evaluation for Pioneer, Helios, and Viking. This procedure was documented in *The Deep Space Network Progress Report* 42-34, May and June 1976. The key characteristics of the Level 4 PRT are reiterated here:

(1) No equipment calibrations are performed except for ranging. Use is made of the latest previous calibration data available.

(2) Subsystem software programs are not exercised as required for other levels of support. The programs are merely loaded and initialized to commence running at the beginning of the pass. However, prior to acquisition, two test commands are sent locally (at the DSS) to a dummy load to validate the on-site command system interface.

(3) No subsystem data transfer tests are run between the DSS and NOCC. Instead, the first 30 minutes of each pass are used for this purpose and to validate the data. The data obtained during this period are provided to the project.

(4) Discrepancy Reports (DRs) are handled differently in those cases where equipment calibrations are out of limits, equipment failures and procedural errors occur

during the first 30 minutes of a pass, and acquisition of spacecraft signal (AOS) is not accomplished on time.

On these types of problems, DRs are closed out on the basis of having resulted from a "load and go" countdown, with its attendant and recognized higher risk.

III. Performance Analysis

Level 4 PRTs were implemented on about 43% of the Voyager tracks during the month of January 1978. Significant data are summarized in Tables 1-4 and accompanying comments.

IV. Conclusion

The trial period ended satisfactorily with the Voyager Project utilizing about 21 h of the additional tracking time made available. The efficiency of this method of operation has clearly been established and, with pre-defined scheduling criteria based on activities programmed to occur during each pass, the Project and the DSN have agreed to schedule Level 4 PRTs during the cruise phase of the Voyager Mission.

Reference

1. Hatch, J. T., "Use of Load and Go Countdowns by the DSN Deep Space Stations", in the *Deep Space Network Progress Report 42-34,* May and June 1976, pp. 124-128, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1976.

Table 1. PRT schedule

DSS	Support requirement	Level 3 time, h	Level 4 time, h
26 m	Standard support	2.0	0.5
	Add ranging (0.5 h)	2.5	1.0
64 m	Standard support low-rate telemetry (HSDL), S-band ranging	2.5	1.0
	Add high-rate telemetry (WBDL) (0.5 h) or X-band ranging (0.5 h)	3.0	1.5
	Add high-rate telemetry and X-band ranging	3.5	1.5

1.5 h are saved in all cases above except the last one, where the time requirements are non-additive to each other for Level 4, and 2 h are saved.

Table 2. Level 4 DSS AOS record

No. of times		Time average, min	Total, min
Early	14	2.2	31
On schedule	11	–	–
Late	18	5.7	103

The numbers in Table 2 cannot be directly compared with Level 3 AOS records because of the flexibility allowed to make up lost time when delays occur during the longer preparation period. The average times, both early and late, for this trial were closer to schedule than the trial for the other Projects. The average during that trial was 6.5 min for early and 7.2 min for late AOS. The latest Level 4 AOS for Voyager was 16 min.

Table 3. Discrepancy report (DR) comparisons

Countdown	No. of DRs
Level 3	17
Level 4	12

The respective number of Discrepancy Reports is comparable to the trial period for the other Projects. The ratio of Level 3 vs Level 4 is about the same as the respective number of tracks and hours tracked. This indicates that there is no significant increase in risk for Level 4 tracks.

Table 4. DSS pre-track preparation and tracking time comparisons

	Tracks		Pre-track preparation, h	Total time, h
	No.	Hours		
Level 3	58	580	144.5	724.5
Level 4	43	442	32.0	474.0

Additional tracking time realized: 64.5 h.