The Telecommunications and Data Acquisition Progress Report 42-70

May and June 1982

N.A. Renzetti
Editor

August 15, 1982

NASA

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The Telecommunications and Data Acquisition Progress Report 42-70

May and June 1982

N.A. Renzetti
Editor

August 15, 1982

NASA

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Preface

This publication provides reports on developments in Earth-based radio technology with applications to several programs. In space communications it reports on activities of the Deep Space Network. In geodynamics it reports on the application of radio interferometry at microwave frequencies for geodynamic measurements. In the field of astrophysics the Deep Space Stations individually and in pairs as an interferometer have been applied to direct observation of celestial radio sources.

Each succeeding issue of this report will present material in some, but not necessarily all, of the following categories:

Radio Astronomy
Radio Interferometry at Microwave Frequencies

  Goedetic Techniques Development
  Spacecraft Navigation
  Orbiting Very Long Baseline Interferometry

Deep Space Network

  Description
  Program Planning
  Advanced Systems
  Network and Facility Engineering and Implementation
  Operations
  Spacecraft Radio Science
  Planetary Radar

In each issue, there will be a report on the current configuration of one of the seven DSN systems (Tracking, Telemetry, Command, Monitor and Control, Test Support, Radio Science, and Very Long Baseline Interferometry).

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

Contents

RADIO INTERFEROMETRY AT MICROWAVE FREQUENCIES
GEODETIC TECHNIQUES DEVELOPMENT
Mobile VLBI System-3

THE DEEP SPACE NETWORK
DESCRIPTION OF THE DSN

PLANETARY AND INTERPLANETARY MISSION SUPPORT
Advanced Flight Projects

ADVANCED SYSTEMS
Tracking and Ground-Based Navigation

Communications

Network Data Processing and Productivity

NETWORK AND FACILITY ENGINEERING AND IMPLEMENTATION
Network

OPERATIONS
Network Operations

ENERGY

Monitor and Control Equipment for the MV-3 Mobile Very Long Baseline Interferometry Station

R. W. Sniffin
Telecommunications Systems Section

A monitor and control system for the MV-3 mobile very long baseline interferometry station has been designed around a commercially available process controller. The significant features of the hardware design and the method by which this equipment will be operated by the MV-3 central computer are discussed.

I. Introduction

One of the goals of the MV-3 project (also known as the ORION project) was to utilize existing technology to the greatest extent possible and thereby reduce the amount of engineering required at JPL. An area in which it seemed probable that this goal could be achieved was that of monitor and control where recent advances in the field of industrial process control had provided an abundance of equipment and systems from which to make a selection. By pursuing this approach, JPL engineering was limited to performing the high-level system design, selecting the process control equipment, designing the interface to other subsystems and writing the software to operate the process control equipment from the MV-3 central computer.

II. MV-3 Monitor and Control System Design

There are four principal monitor and control functions at the MV-3 system level. These are execution of an end-to-end performance test, automation of frequently recurring activities, detection of failures which are not immediately observed by the end-to-end test and analysis of monitor data for malfunction isolation. The first function was delegated to th digital tone extractor of the Phase Calibration Subsystem. The remaining three were delegated to the Monitor and Control Subsystem to be accomplished either within the subsystem or through orchestration of other subsystems connected to the MV-3 central computer.

Much of the equipment selected for use in the MV-3 station had the capability for automatic control and monitoring via an RS-232C serial interface. This included the antenna servo equipment, the water vapor radiometer, the Mark III Data System and the hydrogen maser. The remaining equipment was either of new design or was not available with built-in monitor and control capability.

The original concept was to provide each major item of equipment in this second category with a small data acquisition system to perform the monitor and control functions. These would provide a standard interface to the MV-3 central computer — a Hewlett Packard 1000 System, Model 40, which is supplied with the Mark III Data System. The disadvantage of this approach was that a considerable amount of redundant

1

engineering would have been needed to design suitable mechanical, electrical and environmental provisions in each subsystem for the data acquisition equipment. The concept was therefore rejected in favor of a single microprocessor-based process controller on each transporter. Figure 1 illustrates the MV-3 monitor and control system, as implemented, showing the monitor and control interfaces as well as the end-to-end self test signal flow.

III. Selecting the Process Control Equipment

Selection of the process controllers began with a vendor survey to determine what capabilities were available in commercial equipment. The results were used to create a list of capabilities ranked by the number of vendors who could supply them. Discussions were then conducted with subsystem engineers to determine the minimum subset of capabilities which could fulfill all MV-3 monitor and control requirements. This subset, listed in Table 1, was incorporated into a specification for an "Analog Monitor and Control Assembly" and submitted for competitive procurement.

IV. Analog Monitor and Control Assemblies

The selected analog monitor and control assemblies are manufactured by Analog Devices of Norwood, MA. under the trade name of MACSYM 20. One of their intended uses is to serve as the interface between a host computer and up to 256 assorted analog or digital monitor or control points. In the MV-3 application, approximately 50-percent of this capacity is used.

As shown in the block diagram (Fig. 2), the analog monitor and control assemblies are designed around an 8-bit, Z80 microprocessor with 16K bytes of random-access read-write memory (RAM) and 16K bytes of read-only memory (ROM).

The microprocessor communicates with the host computer via an RS-232C serial interface and is interfaced to its signal conditioning cards via an analog/digital input/output (ADIO) controller. This controller contains all timing and control logic for addressing the individual signal conditioning cards and a 12-bit successive-approximation A/D converter. Input to the A/D converter is through a sample and hold amplifier, a programmable gain amplifier and a 16-channel multiplexer. This allows a single set of analog-to-digital conversion hardware to be shared by all signal conditioning cards.

In the case of analog and low-level digital signals, the monitor and control interface is by card-edge connectors at the rear of the signal conditioning cards. For digital monitor and control signals in excess of 24 volts, the interface is via an optical

isolation assembly. This assembly permits connection to circuits operating at 90 to 140 Vac and 10 to 60 Vdc. Selection of channel characteristics is made by choosing from among four module types for installation on the optical isolation assembly and by throwing a switch on the appropriate signal conditioning card.

The firmware supplied with the analog monitor and control assemblies provides the capability to interrogate or address the input and output channels either one at a time or in groups of from 2 to 16. This is done by issuing a brief command that includes the action to be taken and the identity and characteristics of the channel or channels. The assembly also has the capability to perform simple mathematical operations and to store an array of parameters and command or operation sequences for later execution by issuing the sequence number.

V. Subsystem Interfaces

The process of determining a minimum set of capabilities for the analog monitor and control assemblies resulted in the placement of requirements on the various subsystems for selection of transducers and signal conditioning. At the same time, two standard physical interfaces were adopted to decrease inter-subsystem cabling costs. These were a 9 twisted-pair cable with an overall shield to be used for all signals of 50 volts or less and a 12 conductor cable to be used for all signals in excess of 50 volts.

Within each subsystem, monitor points were selected to permit malfunction isolation to replaceable subassemblies. In general this was accomplished by measurement of subassembly outputs — either directly (where the output was a voltage) or using an appropriate transducer to convert the quantity being measured into a voltage. In certain applications, where the subassembly output could not be easily verified, parameters such as input power supply currents were monitored from which subassembly performance could be inferred. Selection of control points was straightforward, with the only decision which needed to be made relating to control status in the event of a momentary failure of the monitor and control equipment. Where it was desirable to keep a function operating in the event of such a failure, a two-signal (switch on, switch off) control scheme was adopted. For other cases a single signal was used.

At the analog monitor and control assembly end, an interface assembly was designed to connect the twist-lock connectors on the inter-subsystem cabling to the printed-circuit card edge connectors required by the signal conditioning cards and the screw terminals on the optical isolation assembly. This

wiring established the relationship between the functions being monitored or controlled and the physical address of that function. This assembly also provided a convenient place to separate analog monitor, digital monitor and digital control signals arriving in one cable from each other.

Table 2 provides a summary of MV-3 monitor and control points which are assessed via the analog monitor and control equipment. Included in this list are the location arm position and meteorological sensor data which are not monitor data but rather part of the station data record. As noted in the table, the list does not include monitor and control data obtained directly by other subsystems.

VI. Analog Monitor and Control Assembly Operation

In order to address a functional monitor or control point it is necessary to know the physical address in the analog monitor and control assembly that is the card and channel number to which the function is connected. It is also necessary to know whether the function is a monitor or a control point, whether it is analog or digital, and, for analog channels, the sensitivity needed to make the measurement.

Fortunately, the command structure for the analog monitor and control assemblies is such that this information can be retrieved from the parameter array by calculating one or more indices into the array. Thus a single integer number corresponding to a function can be transmitted to the analog monitor and control assembly followed by an instruction sequence number. By following the instruction sequence previously

stored, the physical address and characteristics of the function can be determined and the appropriate operation performed. Values of monitor points are returned to the central computer by using additional instruction sequences.

The process of initializing the analog monitor and control assemblies to perform these operations consists of downloading two text files from the central computer. The first file contains command sequences whereas the second file contains the values to be stored in the parameter array. The process is performed once when the central computer software is initialized and again if an expected response is not received. The central computer software is notified of a failure if no response is received after reinitialization.

As data are received they are put into a common area of storage in the central computer. It takes between 15 and 20 seconds to completely update this area with data — the variation being caused by the data values themselves. This technique provides easy access by applications programs which are performing such activities as performance monitoring, station automation and malfunction isolation.

VII. Conclusion

A monitor and control system has been successfully designed and implemented using two industrial process controllers. The amount of engineering effort to support this activity has proven to be in close agreement with initial estimates. Additional benefits are expected during the development of applications software because of the versatility of the command language firmware provided with these controllers.

Bibliography

1. Brunn, D. L., et al., "ORION Mobile Unit Design," in *TDA Progress Report 42-60,* Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1980.

2. Sigman, E. H., and Parks, G. S., "Use of a Digital Tone Extractor for Real-Time Phase Analysis," in Proc. of the 1979 American Geophysical Union/NASA Conference on Radio Interferometry Techniques for Geodesy.

Table 1. Minimum capability list for MV-3 monitor and control

Parameter	Required capability
Analog input	
Number of channels	64
Configuration	Differential
Sensitivities	0 to ±10 Vdc and 0 to ± ; Vdc[1]
Resolution	12-bit
Digital input	
Number of channels	50
Logic levels supported	TTL, 26 Vdc/0 Vdc, 115 Vac/0 Vac
Simultaneous read capability	16 channels
Digital output	
Number of channels	27
Logic levels supported	115 Vac @ 1.8 Aac, 26 Vdc @ 1.8 Adc
Physical	
Configuration	Rack mount with slides
Maximum dimensions	2f ↓ mm (10.5") H × 521 mm (20.5") D[2]
Environmental	
Operating temperature	0 to 45°C
Storage temperature	−45 to 66°C
Operating altitude	Sea level to 3048 m (10,000 ft)

(1) Selected unit has sensitivities of ±10, ±5, ±7.5, ±1.25, ±0.625 Vdc, etc., to ±4.88 mVdc

(2) Selected unit vertical dimension is 222 mm (8.75")

Table 2. MV-3 monitor and control points

Subsystem and assembly	Monitor points	Control points
Microwave Subsystem		
Low-noise amplifiers	5	4
Cryogenics	6	7
Power supplies	2	–
Miscellaneous	3	–
Receiver Subsystem		
Down-converters	9	–
Power supplies	6	–
Phase Calibration Subsystem		
Cable stabilizer	6	–
Comb generators	2	2
Noise adding radiometer	–	1
Power supplies	8	–
Antenna Subsystem[1]		
Locator arm	9	2
Emergency stop/movement warning	1	2
Facilities Subsystem		
Power generation	5	–
Power distribution	2	10
Environmental control	22	–
Data Acquisition Subsytem[2]		
Mark III Data System	9	–
recorders	22	–
Frequency and Timing Subsystem[3]		
Distribution amplifiers	4	–
Cesium standard	3	–
Instrumentation	1	–
Meteorological equipment		
Sensor data	4	–
Power supplies	4	–
Water vapor radiometer[4]		

NOTES.

(1) Monitor and control of antenna controller and servos are by self-contained equipment.

(2) Configuration monitor and control of Mark III Data System and recorders are by self-contained equipment. Listed monitor points are power supply voltages.

(3) Monitor and control of hydrogen maser, frequency standard room temperature monitor and part of the frequency standard performance monitor are by self-contained equipment.

(4) Monitor and control of water vapor radiometer is by self-contained equipment.
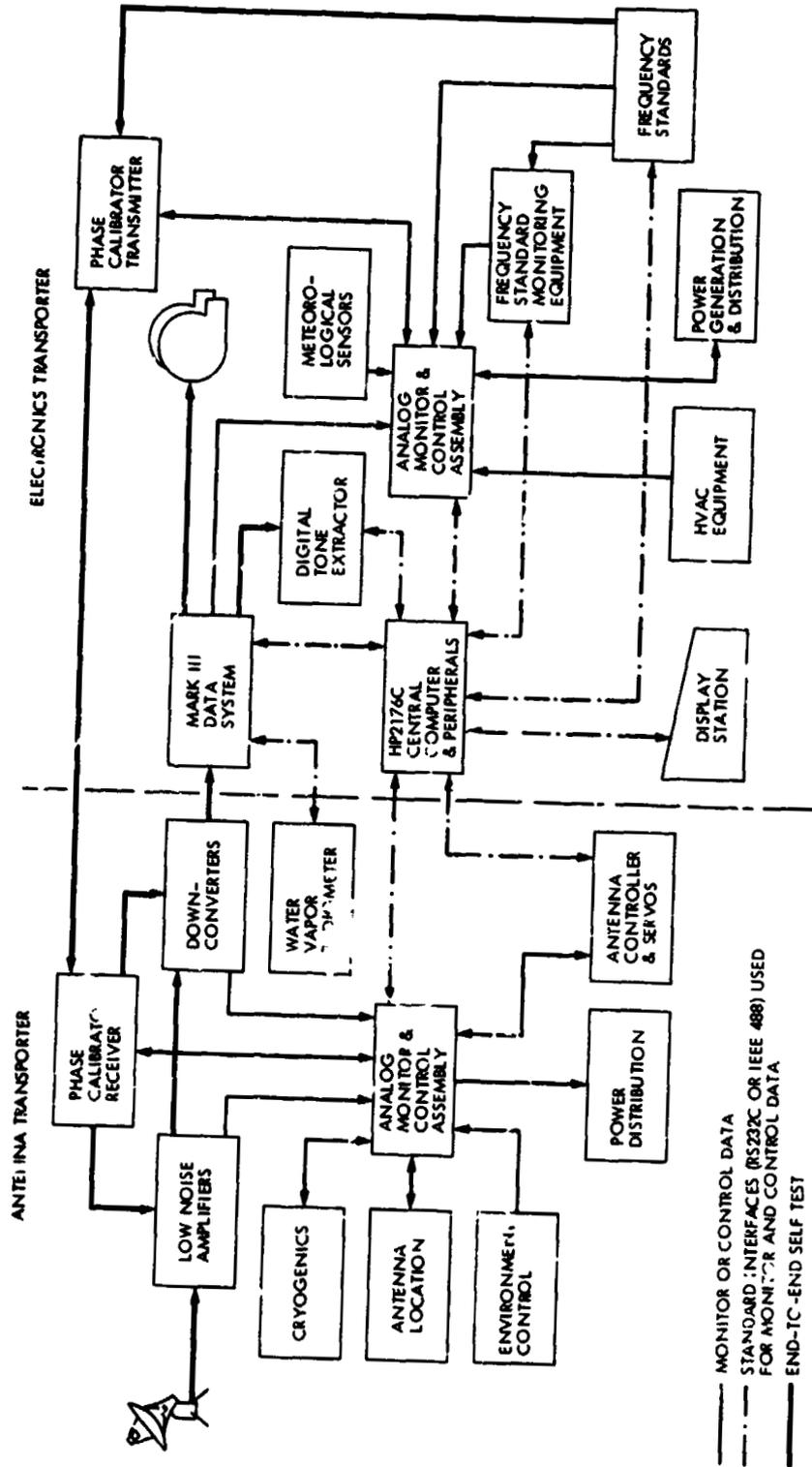
Fig. 1. MV-3 equipment/MV-3 monitor and control interconnections

PHASE CALIBRATOR TRANSMITTER

ELECTRONICS TRANSPORTER

FREQUENCY STANDARDS

FREQUENCY STANDARD MONITORING EQUIPMENT

POWER GENERATION & DISTRIBUTION

METEORO-LOGICAL SENSORS

ANALOG MONITOR & CONTROL ASSEMBLY

HVAC EQUIPMENT

DIGITAL TONE EXTRACTOR

MARK III DATA SYSTEM

HP2176C CENTRAL COMPUTER & PERIPHERALS

DISPLAY STATION

ANTENNA TRANSPORTER

PHASE CALIBRATOR RECEIVER

DOWN-CONVERTERS

WATER VAPOR RADIOMETER

ANTENNA CONTROLLER & SERVOS

LOW NOISE AMPLIFIERS

ANALOG MONITOR & CONTROL ASSEMBLY

POWER DISTRIBUTION

CRYOGENICS

ANTENNA LOCATION

ENVIRONMENT CONTROL

MONITOR OR CONTROL DATA

STANDARD INTERFACES (RS232C OR IEEE 488) USED FOR MONITOR AND CONTROL DATA

END-TO-END SELF TEST

Fig. 2. Analog Monitor and Control Assembly block diagram

DSN Command System Mark IV-85

H. C. Thorman

TDA Engineering Office

Modification of the currently operational DSN Command System Mark III-80 in 1982 consisted of upgrading Command System monitoring functions in the Network Operations Control Center. DSN Command System Mark IV-85 functional design is described for the Mark IV-A Network, which is planned for 1984-1985 implementation.

I. Present System

The Mark III-80 configuration of the DSN Command System, which was described in Ref. 1, is the currently operational configuration. To provide computer-controlled prepass data transfer and validation tests and revised displays, Network Operations Control Center (NOCC) Command Subsystem software modifications were completed in May 1982 as part of a general upgrade of the Network Operations Control Area (NOCA) to reduce operations costs.

II. Mark IV-85 System

A. Mark IVA Network Implementation

The Mark IVA Network implementation, to be completed in 1985, will provide one Signal Processing Center (SPC) at each of the three Deep Space Communication Complexes (Goldstone, California; Canberra, Australia; and Madrid, Spain). The Ground Communications Facility (GCF) will provide communications between JPL and each SPC. The Networks Consolidation Program (Ref. 2) provides for the Mark IVA Network to support high-apogee earth-orbital missions in addition to the deep space missions.

Figure 1 is a block diagram of the DSN Command System Mark IV-85, as previously described in Ref. 3. Each

Deep Space Communications Complex (DSCC) will have a 64-meter antenna with deep space uplink, a 34-meter antenna with both deep space and earth orbiter uplinks, and a 9-meter antenna with earth orbiter uplink. (The Goldstone and Canberra complexes will each have an additional 34-meter antenna for downlink only.)

During the past year the Mark IVA Network implementation schedule has been modified to provide earlier completion at the Canberra complex. Additional support requirements have enlarged the mission set. The Command System performance requirements, functional description, and subsystems configurations, however, are the same as presented in Ref. 3.

DSCC Command Subsystem prototype testing and software design are progressing on schedule. A contract for fabrication of the 16 new command modulator assemblies was awarded in June 1982.

B. Implementation Schedule

The Mark IVA Network implementation plan calls for an interim configuration to be installed at all three complexes in early 1984, and a final configuration at Goldstone by February 1985, Canberra by May 1985, and Madrid by August 1985. The interim configuration will include new command

equipment for support of high earth orbit and deep space missions, while retaining portions of the present Mark III configuration. The final configuration at each complex will include the 9-meter antenna and associated front end equipment, at least three strings of DSCC command subsystem equipment, and the new DSCC Monitor and Control Subsystem, as shown in Fig. 1.

C. Mission Set

The Mark IVA Network baseline requirements have been extended to provide capabilities to support all of the flight missions listed below:

(1) Current deep space missions:

    (a) Pioneers 6 through 12

    (b) Viking Lander

    (c) Helios

    (d) Voyagers 1 and 2

(2) Planned deep missions:

    (a) Galileo

    (b) International Solar Polar Mission (ISPM)

    (c) Giotto (backup support)

(3) Current high elliptical earth orbital missions:

    (a) International Sun-Earth Explorer No. 3 (ISEE-3)

(4) Future high elliptical earth orbital missions:

    (a) Active Magnetospheric Particle Tracking Explorer (AMPTE); 3 spacecraft

    (b) Origin of the Plasma in the Earth's Neighborhood (OPEN); 4 spacecraft

(5) Other earth orbiter missions:

    (a) TDRS (launch and emergency support)

    (b) Space Telescope (emergency support)

    (c) GOES G, H, I (backup support)

D. Performance Requirements

Support of the Mark IVA mission set will require Command System performance characteristics compatible with the NASA standard transponders, which are to be used on future spacecraft, and also compatible with current inflight spacecraft and certain planned spacecraft that do not use the standard transponder. Some of the required capabilities are listed below.

(1) Data rates. Data rates from 1 to 2000 bits/sec will be provided

(2) Subcarrier frequencies. Sine-wave and square-wave subcarriers will be generated at frequencies of 100 Hz to 16 kHz.

(3) Subcarrier data modulation. Selection will be provided for phase-shift-keyed (PSK) or frequency-shift-keyed (FSK) modulation of the subcarrier by the pulse-code-modulated (PCM) command symbol stream. An option for amplitude modulation (AM) of the FSK subcarrier will also be provided.

(4) Carrier modulation. The command-modulated subcarrier will be phase-modulated on an S-band carrier for radiation to the spacecraft. Control of modulation index angle will be provided over a range from 0.1 to 1.8 radians.

(5) Carrier frequencies. Generation of the uplink carrier at S-band frequencies assigned for deep space missions will be provided at the 64- and 34-meter antennas. S-band frequencies assigned for earth orbit missions will be provided at the 34- and 9-meter antennas.

III. System Functional Description

As discussed in Ref. 4, many of the spacecraft supported by the DSN have onboard storage and sequencing capabilities that permit command sequences to be sent well in advance of the actions to be taken by the spacecraft. Thus, fewer direct action (real-time) commands are needed. Ground system capabilities providing massive storage of spacecraft commands, multimission operating functions, and standardized protocol were incorporated in the DSN Command system in 1978 (Ref. 5). These capabilities will be continued in the Mark IV-85 system configuration.

A. Operational Functions

End-to-End spacecraft command operations are represented functionally in Fig. 2. Command sequences for one or more spacecraft are generated and stored at a Mission Operations Center (MOC). Commands for a particular spacecraft are selected from the command files, formatted into messages, and stored for transmittal to a specified link of a DSCC. Command data are extracted from the message received and are stored and queued until radiated. Finally, the commands arrive at the spacecraft and are either executed immediately or stored onboard for later execution.

The functions of the DSN Command System in this process include the following:

(1) Establishing the DSCC configuration for the specified spacecraft.

(2) Receiving and storing command data at the DSCC.

(3) Queuing command data to be radiated to the spacecraft.

(4) Radiating the command data to the spacecraft.

(5) Monitoring and reporting system status and events.

B. Operational Procedure

On-site configuration inputs to the DSCC Command (DCD) Subsystem specify the flight project name and the spacecraft identification number. These inputs cause the Command Processor Assembly (CPA) software to transfer a specified configuration and standards and limits table from disk storage to memory, and to configure the DCD Subsystem according to the table. Changes may later be made by messages from NOCC via the GCF (or by keyboard entries at the Link Monitor and Control Console, in an emergency).

Prior to the beginning of the scheduled spacecraft track, the control of the DSCC command functions is transferred to the NOCC. Configuration standards and alarm/abort limits can be updated by GCF transmission of messages from the NOCC Command Subsystem (NCD) real-time monitor processor. The standards and limits are derived from files compiled in the NOCC Support Subsystem. Spacecraft-dependent parameters, such as symbol period, subcarrier frequency, alarm limits, and abort limits, are established via these messages. After the proper configuration standards and limits have been established, test commands are transmitted through the system to ensure that the system can accept spacecraft commands via GCF, temporarily store the commands, and confirm radiation. During this test the transmitter output is radiated into a dummy load. After the Network Operations Control Team (NOCT) has established that the system is operating properly, the station operator switches the transmitter to space radiation, and the NOCT transfers command data control to the flight project's MOC for loading of actual spacecraft command sequences to be radiated to the spacecraft during the track period.

At the time for radiation of each command element, the subsystem advances to the active mode (see Fig. 3 for description of the various modes) and command data are transferred to the Command Modulator Assembly (CMA) for immediate radiation via the Receiver-Exciter, Transmitter, Microwave, and Antenna Subsystems.

C. Command Data Handling

The DCD Subsystem design allows mission operations to prepare large files of spacecraft commands in advance and then to forward several files to the DSCC link at the beginning of a spacecraft track. The design also provides real-time system status monitoring and control. For protection of

data integrity, every message block to or from the CPA contains a block check sum, in addition to the GCF error detection provisions.

1. Command files. Each file may consist of up to 256 high-speed data blocks. The content of each data block is a file element. The first block in a file contains the header element and each subsequent block contains a command element. Each command element may consist of up to 800 bits of spacecraft command data. Up to 8 files for a given mission can be stored by the CPA. Thus the available storage is over 1.6 million command bits.

The header element contains file identification information, file processing instructions, and a file checksum. The file processing instructions include optional file radiation open and close window times, and an optional file bit 1 radiation time. File open and close window times specify the time interval during which command elements in the file may begin radiation (i.e., a mission sequence may demand that specific commands not be sent before or after a certain time). The bit 1 radiation time allows the project to specify the exact time at which the file is to begin radiation to the spacecraft. The file checksum is created at the time of file generation and is passed intact to the CPA. It adds reliability to insure that no data were dropped or altered in the transfer from one facility to another. (This is in addition to the previously mentioned block checksums.)

The command elements each contain command bits, file identification, element number, element size, and an optional "delay time" (interval from start of previous element). If delay time is not specified, the element will start radiating immediately after the end of the previous element.

2. Receiving and storing command data at a DSCC. Normally, the files of commands to be radiated to the spacecraft will be sent from the MOC to the specified DSCC link at the beginning of a spacecraft track period. However, files may be sent to the DSCC link at any time during the spacecraft track period. The first step in receiving and storing command data at a DSCC is the process of opening a file area on the CPA disk. The MOC accomplishes this by sending a header element, which serves as a *file-open* directive. After the CPA acknowledges receipt of the header element, the MOC sends the remainder of the file (up to 255 command elements) and follows it with a *file-close* directive. The CPA acknowledges the file-close instruction and indicates whether the file loading was successful or unsuccessful. If the file loading was unsuccessful, the acknowledge message contains the reason for the failure and from what point in the file the command elements are to be transmitted. When the file is successfully

closed, the MOC may proceed to send additional files, up to a total of eight.

3. **Queuing the command data for radiation.** After the files are stored at the CPA, the MOC then sends a *file-attacn* directive for each of up to five f'e names to be placed in the radiation *queue*. The Mission Control Team determines in which order the files are to be attached. The order in which the file-attach directives are received at the CPA determines the sequence in which the files will be radiated: that is, first attached, first to radiate to the spacecraft.

4. **Command radiation to the spacecraft.** The first command element in the top (prime) file in the queue begins radiation to the spacecraft immediately after attachment or as soon as all optional file instructions (such as bit 1 radiation time) are satisfied. The prime file status is defined to be *active* when the first command element begins radiation. Upon completion of radiation of the first command element, the second command element begins radiation either immediately or when the optional *delay time* has been satisfied. The process continues until all command elements in the file have been radiated. After the first file completes radiation, the second file in the queue automatically becomes the prime file and the command radiation process is repeated. After the second file completes radiation, the third file becomes prime, etc. This process is repeated until all files in the queue are exhausted. The MOC can attach new files to the queue whenever space is available.

Confirmations of prime-file command-element radiations are reported in *event messages* to the MOC and NOCC once per minute, or after five elements have been radiated, whichever occurs first. If a command element is aborted, or if an alarm occurs, an event message is sent immediately.

5. **Additional data processing.** The foregoing descriptions of the DSCC functions of storing the command files, attaching the files to the queue, and radiating the commands to the spacecraft assume nominal (standard) operation. Additional data processing functions are provided for worst-case conditions of non-nominal operations and failure recovery. Control of these functions is normally exercised remotely from the MOC. However, emergency control is also available at the Link Monitor and Control Console.

a File erase A file can be deleted from storage at the CPA by means of a *file erase* directive, if the file is not attached to the radiation queue

b Clearing the queue As previously stated, the order of file radiation to the spacecraft is dependent on the order of

files in the queue. To rearrange the order, a *clear-queue* directive must be sent, followed by file-attach directives in the desired order.

c. Suspend radiation. If the Mission Control Team desires to stop command radiation, a *suspend* message can be sent to the CPA. This message stops command radiation to the spacecraft upon completion of the current element. The file status then changes from active to suspended.

d. Resume command radiation. To restart radiation of a suspended file (either suspended intentionally or from an abort), a message can be sent to *resume* radiation *at* any specified unradiated element in the file. The suspend and resume-at directives can be used for skipping elements of the prime file, if desired.

e. Command abort. As each command bit is radiated to the spacecraft, numerous checks are made to insure validity of the command data. If a failure is detected during the radiation, the command element is automatically aborted, the prime file status is changed from active to suspended, and radiation is terminated until a resume directive is received.

In addition to the automatic abort function there is provision for the MOC to send an *abort and suspend* directive to terminate command radiation immediately without waiting for completion of an element.

f. Close window time override. If a close-window time is specified in a file header element, and the Mission Operations Team later decides to extend the permissible time for radiation of that file, an *override* message can be sent (after the file becomes prime) which instructs the CPA to ignore the close window time.

D. Data Records

All message blocks received by the CPA and all blocks sent from the CPA will be logged at the DSCC on the Original Data Record (ODR). In addition, the CPA has the capability to record a temporary ODR on disk if the ODR is disabled.

Message blocks from all complexes are recorded at the GCF central communications terminal (CCT). Command system message blocks from a Mission Operation Center to a DSCC are also recorded at the CCT.

The DSCC original data records and the CCT recording provide information for fault isolation in case problems occur in the Command System operation.

IV. Subsystems Configurations for Mark IV-85 System

Planned modifications and reconfiguration of subsystems for the DSN Command System Mark IV-85 (and Mark IV-84) are summarized below.

A. Antenna Mechanical Subsystem

At Canberra and Madrid all antennas will be located in the vicinity of the SPC. At Goldstone, the 64-meter antenna and the 9-meter antenna will be located near the SPC. The Goldstone 34-meter transmit-receive antenna will remain at the present DSS 12 (Echo Station) site, but control will reside at the SPC.

B. Antenna Microwave Subsystem

For the 9-meter antenna, the microwave subsystem will provide uplink signal feed at S-band frequencies assigned for earth orbital missions (2025-2110 MHz). For one 34-meter antenna at each complex, the microwave subsystem will be required to handle S-band uplinks over the range of earth orbital and deep space missions (2025-2120 MHz). For the 64-meter antenna the microwave subsystem uplink capability will be unchanged (S-band 2110-2120 MHz).

The 9-meter and 34-meter antenna microwave subsystems provide selection of right or left circular polarization. The 64-meter antenna microwave subsystem provides selection of linear polarization or right or left circular polarization.

C. Transmitter Subsystem

The 9-meter antenna will have a 10-kW transmitter operating in the earth orbital mission S-band frequency range. A 34-meter antenna will have a 20-kW transmitter operating over the range of earth orbital and deep space mission S-band frequencies. The 64-meter antenna will have 20-kW and 100-kW transmitters for the deep space mission S-band frequency range, as now.

D. Receiver-Exciter Subsystem

An S-band exciter for the earth orbital frequency range will be provided for each 9-meter antenna. The DSN exciter for the 34-meter antenna will be upgraded to cover earth orbital and deep space mission S-band frequencies. The present DSN S-band exciter will be retained in the 64-meter antenna link.

Functions of the exciter include receiving the command-modulated subcarrier signal from the DSCC Command (DCD) Subsystem, phase-modulating that signal on the uplink carrier, returning a demodulated signal to the DCD subsystem for confirmation, and sending modulation on or off indications to the DCD subsystem.

E. DSCC Command Subsystem

In the final Mark IVA Network configuration (in 1985), the DCD Subsystem in the SPC at each complex will be implemented as shown in Fig. 1. A new Command Switch Assembly (CSA) will permit any of the exciters to be connected to any of the Command Modulator Assemblies (CMA) under control of the Complex Monitor and Control console. New CMAs will be implemented to accommodate the Mark IVA mission support requirements. The CPAs will use existing Modcomp II-25 computers with core memory increased to maximum capacity. CPA software will be upgraded to satisfy new mission support requirements, to modify the CMA interface functions, and to provide required functions for interfacing with the new DSCC Monitor and Control Subsystems.

F. DSCC Monitor and Control Subsystem

New equipment will be implemented for the DSCC Monitor and Control Subsystem (DMC) at each complex in the final Mark IVA Network configuration. Assignment of command equipment (antenna, transmitter, exciter, and command modulator-processor combinations) to a given "link," for each scheduled spacecraft pass or for a scheduled test, will be accomplished by the DMC along with telemetry and tracking equipment assignments. Prepass countdown will be controlled by inputs at the Link Monitor and Control Console.

The DMC will receive antenna pointing and uplink frequency predictions and will relay them to the appropriate subsystems. The DMC will send link status information to the CPA, and the CPA will send Command Subsystem status information to the DMC for link console displays and for incorporation into the monitor data that the DMC sends to the NOCC.

In the interim configuration, the Monitor and Control Subsystem will be limited to the existing Data System Terminal (DST) and Digital Information Subsystem (DIS) functions.

G. GCF Subsystems

In the final Mark IVA Network configuration, the GCF Digital Communication (GDC) Subsystem will replace the

present GCF High Speed Data and GCF Wideband Data Subsystems. Command data blocks will be communicated at a line rate of 56 kb/s, instead of the present 7.2 kb/s rate, between the Central Communications Terminal at JPL and the Area Routing Assembly at each DSCC.

At the Goldstone DSCC the GCF Intersite Analog Communications Subsystem will communicate the CMA output signal from the SPC to the DSS 12 exciter and the confirmation signal from the exciter to the SPC.

H. NOCC Command Subsystem

The NOCC Command Subsystem (NCD) Real-Time Monitor (RTM) software will be upgraded to accommodate new destination codes, spacecraft identifiers, standards and limits tables and test command tables for the interim and final configurations. The NOCC Support Subsystem will be expanded to provide capability for Command System performance records and analysis and additional capacity for test command tables.

References

1. Thorman, H. C., "DSN Command System Mark III-80," in *The Telecommunications and Data Acquisition Progress Report 42-57*, pp. 33-42, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1980.

2. Yeater, M. L., and Herrman, D. T., "Networks Consolidation Program," in *The Telecommunications and Data Acquisition Progress Report 42-65*, pp. 19-24, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 15, 1981.

3. Thorman, H. C., "DSN Command System," in *The Telecommunications and Data Acquisition Progress Report 42-64*, pp. 53-60, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1981.

4. Stinnett, W. G., "DSN Command System Mark III-78," in *The Deep Space Network Progress Report 42-43*, pp. 4-8, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1978.

5. Thorman, H. C., "DSN Command System Mark III-78," in *The Deep Space Network Progress Report 42-49*, pp. 11-18, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1979.
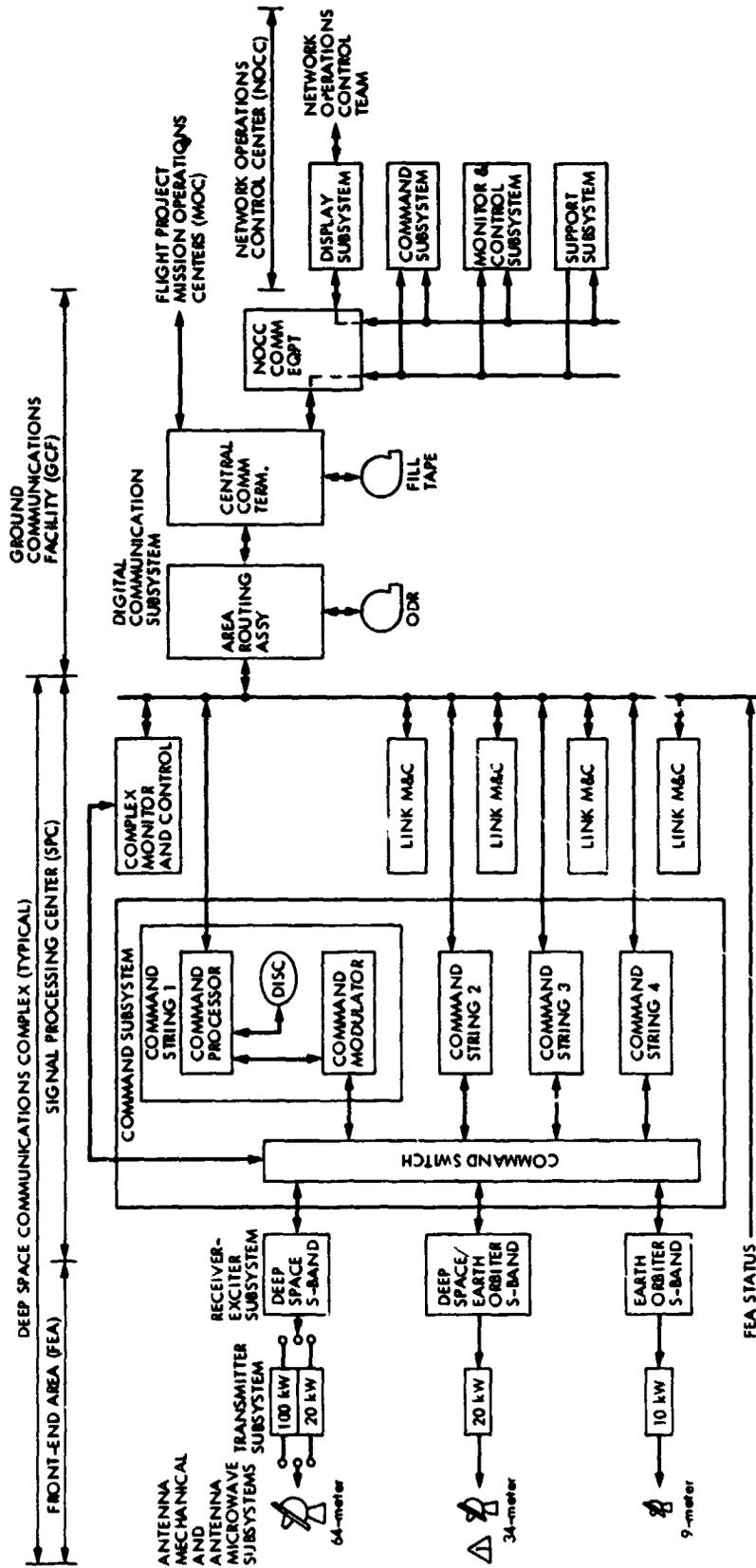
Fig. 1. DSN Command Subsystem Mark IV-85 block diagram

14

MISSION OPERATIONS CENTER	DEEP SPACE COMMUNICATIONS COMPLEXES	SPACECRAFT

MISSION B

MISSION A NON-REAL-TIME COMMAND GENERATION

B

A

REAL-TIME PROCESSING

MADRID DSCC

TIDBINBILLA DSCC

GOLDSTONE DSCC REAL-TIME PROCESSING

IN TRANSIT seconds-hours

SPACECRAFT B

SPACECRAFT A STORAGE BEFORE EXECUTION, seconds-days

STORAGE BEFORE ENTRY INTO REAL-TIME SYSTEM, days

STORAGE BEFORE FORWARDING TO DSN, hours

STORAGE BEFORE RADIATION, seconds-hours
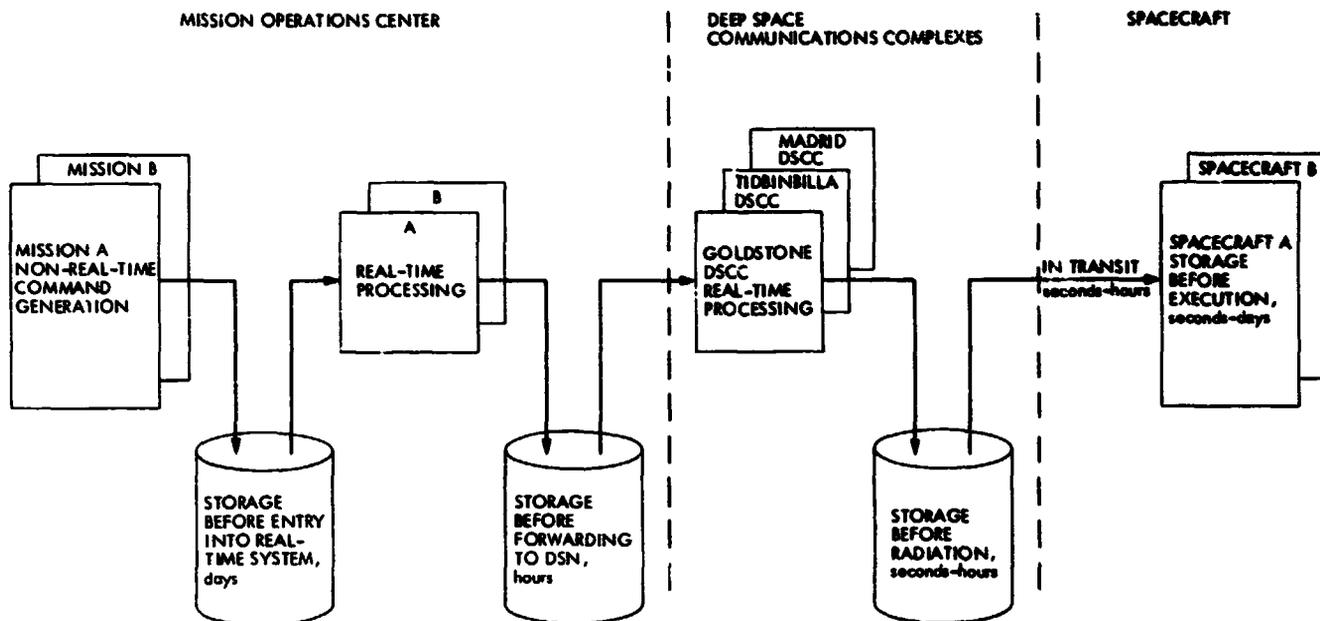
Fig. 2. End-to-end command data flow—typical storage times

SUBSYSTEM MODE	PRIME PURPOSE	DATA ACCEPTED VIA HIGH SPEED	COMMENTS
CALIBRATE 1	ALLOWS DSS OPS PERSONNEL TO PERFORM INITIALIZATION TASKS	CONFIGURATION, STANDARDS AND LIMITS, MODE CONTROL, RECALL	UPON RECEIPT OF A STANDARDS AND LIMITS AND CONFIGURATION HSD BLOCK, THE CALIBRATE 2 MODE WILL BE ENTERED
CALIBRATE 2	STANDARD MODE FOR UPDATING STANDARDS AND LIMITS AND CONFIGURATION DATA VIA HIGH SPEED PRIOR TO SPACECRAFT ACQUISITION	CONFIGURATION, STANDARDS AND LIMITS, MODE CONTROL, RECALL	MULTIMISSION STANDARD PROCEDURES SHOULD STATE THAT THIS MODE IS TO BE ENTERED FOR CHANGING SUBCARRIER FREQUENCY, BIT RATE
IDLE 1	SAFE MODE - CANNOT COMMAND. PROVISION FOR IDLE/ACQUISITION SEQUENCE. ALLOWS CONFIGURATION AND STANDARDS AND LIMITS TO BE CHANGED	CONFIGURATION, STANDARDS AND LIMITS, MODE CONTROL, RECALL	ALL STANDARDS AND LIMITS AND CONFIGURATION DATA PARAMETERS WILL TAKE IMMEDIATE EFFECT IN THIS MODE
IDLE 2	ALLOWS ENTRY INTO ACTIVE MODE. PROVIDES IDLE/ACQUISITION SEQUENCE DURING COMMAND PERIODS	RECALL AND MODE CONTROL DATA ONLY	MODE CONTROL DATA CONTAINED IN THE COMMAND ELEMENT BLOCK
ACTIVE	COMMAND TRANSMISSION	RECALL AND MODE CONTROL DATA ONLY	
ABORT	PROVIDES ABORT INSTRUCTION TO CMA	RECALL AND MODE CONTROL DATA ONLY	

(labels on flow diagram: BIT VERIFY ERROR; SELECTABLE ABORT RETURN (IDLE 1 STANDARD))

NOTES: 1. COMMAND DATA MESSAGES WILL BE ACCEPTED IN ALL MODES
2. ALARM MESSAGES/ALARM DATA WILL BE TRANSMITTED TO THE MOC IN ALL MODES EXCEPT ABORT

Fig. 3. DSCC Command Subsystem modes

15

N82 32532

The DSN VLBI System Mark IV-86

W. D. Chaney

TDA Engineering

This article describes the DSN VLBI System Mark IV-86, Wide Channel Bandwidth (Block II). It covers the system requirements, description and implementation plans. Narrow Channel Bandwidth VLBI (Block I) was described in prei ·us articles (Refs. 1,2).

I. Introduction

The DSN Very Long Baseline Interferometry (VLBI) System Mark IV-86 will be implemented to provide the radio source catalog and baselines maintenance for Galileo delta differential one-way range (delta DOR). The primary stations to be used for radio source catalog maintenance are the 34-m deep space listen-only stations that are planned for Goldstone, California, and Canberra, Australia. However, the 64-m deep space stations at each complex can also be used for Wide Channel Bandwidth (WCB) VLBI.

The Galileo precision requirement for radio source catalog is 25 nanoradians (15 cm), and 30 cm for baseline determination.

II. System Description

A. Definition

The Deep Space Communications Complex (DSCC) VLBI System is the assemblage of various subsystems at a specific complex which form an instrument for receiving and obtaining necessary VLBI data in conjunction with at least one other complex and, together with elements involved with the monitoring and control and data processing functions, comprise the DSN VLBI System.

B. Description

Functionally, the DSN VLBI System (Fig. 1) comprises the DSCCs, which individually receive the RF signal and downconvert segment bandwidths of the RF spectrum to videoband frequencies, which are then digitized and formatted by digital equipment. The digital data is then recorded on wideband recorders and shipped to the JPL/Caltech Correlator Facility for processing. Many of the functional blocks in Fig. 1 are common to both the Wide Channel Bandwidth (Block II) and the Narrow Channel Bandwidth (Block I) versions of the VLBI System. (See Refs. 1 and 2 for Block I description.)

The Antenna Subsystem is pointed to the appropriate signal source at the proper time by the Antenna Pointing Subsystem, which obtains pointing information (predicts) from the Network Data Processing Area (NDPA) of the Network Operations Control Center (NOCC) via the DSS Monitor and Control Subsystem (DNC) and the Ground Communications Facility (GCF).

The Antenna Microwave Subsystem (UWV) receives the signal flux gathered by the antenna. After amplification by the FET or TWM, the signal is sent to the Receiver-Excit Subsystem, which heterodynes this signal to an intermediate frequency (IF).

The WCB system configuration will have 14 channels of 7 pairs (expandable to 28 of 14 pairs) of adjacent lower and upper sideband spectra selectable from 0.25- to 2.0-MHz bandwidths. These paired channels may be allocated as desired between S- and X-band frequencies, and each channel may be individually set to any place within the bandwidth, limited only by the front end amplifier and RF-IF downconverter. These frequencies are further downconverted from IF to video for the final acquisition and recording processes.

The Frequency and Timing Subsystem (FTS) provides the station local clock, using a very stable hydrogen maser as the primary standard. Reference frequencies and timing signals are derived from the clock for distribution to other subsystems. Similarly, a reference signal from the coherent reference generator (CRG), which distributes the reference signals, will drive phase calibration generators (PCG) as part of the Frequency and Timing Subsystem (FTS) via a coaxialcable, phase-stabilization assembly which effectively translates the station's clock frequency stability to the comb generators in the PCG. The comb generator provides comblike, phase-stable, line spectra S- and X-band microwave frequencies, which are injected into the respective Microwave Subsystems prior to the input circuitry of the FETs or TWMs.

These phase-stable reference signals are amplified by the receiver and are down-converted simultaneously with the received signals. These reference signals will be used to calibrate out phase variations (which occur within the receiver, down converter, and digital subsystems) during the crosscorrelation and data processing procedure. Since the comb signal encounters the received signal for the first time at the injection point, this point is established as the instrument's RF reference point for the DSCC VLBI System. This is the point at which the cross-correlation and postcorrelation estimation calculations refer the resultant Earth parameters, station location and clock offset and rate information relative to the other instruments.

The reference is used to relate other station references such as the station's location reference point (intersection of antenna axes or equivalent) and the Epoch reference point at the FTS CRG output located within the control room. The cable stabilizer effectively translates these points with a known time delay for interstation clock synchronization purposes. The clock Epoch reference point in turn will function as the reference for all subsystems and assemblies within the respective stations.

The data acquisition and recording subassembly records the data at rates up to 112 Mbits/sec. The tapes generated are shipped to the JPL/Caltech correlation facility for processing.

The sample of data from each radio source is transmitted via GCF wideband data line to the NOCC VLBI Processor Subsystem (VPS) for validation of fringes from the baseline pair of stations. The results are displayed to the Network Operations Control Team, and also transmitted to the stations for display.

The Deep Space Communications Complex (DSCC) Monitor and Control Subsystem (DMC) sends control and configuration information to the DSCC VLBI Subsystem (DVS) from data received from NOCC via the GCF. It also collects various calibration and configuration data which is provided to the DVS for recording with the VLBI data and via the GCF for monitoring.

At NOCC (Fig. 1), the NDPA use data for real-time monitoring-display functions in the work Operations Control Area (NOCA). The NOCA provides the control information to the DSCC, via the GCF, to the DMC.

The JPL/Caltech VLBI Processor receives tapes and performs the cross-correlation of the data from the observing stations and, with further postcorrelation and estimation processing, radio source catalog data and baseline data are generated.

III. Implementation

A. General

Implementation of Wide Channel Bandwidth (WCB) VLBI will provide a capability to maintain the radio source catalog and baseline distances. The WCB VLBI supplements the Narrow Channel Bandwidth (NCB) VLBI. The NCB VLBI is used for direct navigation support in determining relative clock and clock rate offsets, Universal Time 1 (UT1), polar motion (PM) and delta differenced one-way range (delta DOR).

A simplified block diagram of WCB VLBI is shown in Fig. 1. The WCB VLBI will be implemented on a 34-m listen-only antenna subnet. Effective bandwidth will be 400 MHz at X-band and 100 MHz at S-band. Sampling rates and recording will be at 14 Mbits/sec to 112 Mbits/sec. Tapes will be shipped to the JPL/Caltech correlator for processing.

B. Functional Performance Requirements

Parameter	X-band	S-band
System temperature	60 K	50 K
Bandwidth	400 MHz	100 MHz
Frequency range	8200-8600 MHz	2200-2300 MHz

Parameter	X-band	S-band
No. of channels	14 (expandable to 28)	
Channel bandwidth	2 MHz	
No. of tones/channel	3 (minimum)	
No. of sampling rate	14-112 Mbits/sec	

C. Modifications for WCB VLBI

Modifications to a standard 34-m listen-only Front-End Area (FEA) and Signal Processing Center (SPC) to provide WCB VLBI are given below.

1. **Microwave.** Wideband field effect transistors (FET) at both S- and X-band frequencies will be added with a bandwidth of 100 and 400 MHz, respectively. System temperature will be less than 50 K at S-band and 60 K at X-band. The frequency range will be at least 100 MHz at S-band from 2200-2300 MHz and 400 MHz at X-band from 8200-8600 MHz.

An insertion port for both S-band and X-band will be supplied as far forward in the microwave as possible in order to insert calibration tones for VLBI system calibration.

2. **Receiver.** An RF-to-IF conversion assembly will be added for both S- and X-band. The S-band will have a bandwidth ⩾ 100 MHz and range of 2200-2300 MHz. The X-band will have a bandwidth ⩾ 400 MHz and a range of 8200-8600 MHz.

3. **Phase calibration generators (PCG).** A PCG assembly will be added that will transfer the stability of the hydrogen maser frequency standard to the antenna microwave (UWV). At the UWV a coherent comb generator subassembly generates a comb of frequencies across the frequency range. This comb frequency divisor is selected so that at least three tones are in one 2-MHz VLBI channel.

4. **WCB VLBI Data Acquisition Assembly.** The Data Acquisition Assembly is located in the SPC and provides the following functions:

(1) Selects the IF signal.

(2) Selects frequency synthesis channels.

(3) Provides IF-video conversion.

(4) Provides image rejection and low pass filtering.

(5) Provides data sampling, formatting and recording.

(6) Provides selective data transmission via WBDL for validation.

Initially the Data Acquisition Assembly will provide 14 channels of 7 pairs of adjacent upper and lower sideband spectra, selectable from 0.25 to 2.0 MHz. The number of channels will be expandable to 28 channels of 14 pairs.

Data sampling and recording will be at 14 to 112 Mbits/sec. The format for recording will be compatible with the Mark III Haystack Observatory. Data transfer will be magnetic tapes shipped to the JPL/CIT correlator facility.

5. **Water Vapor Radiometer (WVR)** The WVR measures water vapor content along the line of sight in order to calibrate the VLBI data for this error source. The Advanced-Systems WVR models will be upgraded and integrated into the DSCC technical facilities. Data will be relayed from the technical facilities subsystem to the DSCC VLBI subsystem for incorporation into the VLBI data stream. These data will be used by the Block II correlator to calibrate VLBI observables for water vapor content along the line of sight.

The WVR measures the brightness temperature at two frequencies – 20.7 and 31.4 GHz. These brightness temperatures are used to determine the water vapor content along the line of sight. The WVR is slued with the main antenna; and periodic calibration is done by dipping in elevation at each 90° in azimuth.

6. **Block II correlator.** The Block II VLBI correlator is a joint JPL/Caltech implementation with the correlator located at the Caltech campus.

The Block II correlator will be implemented for the correlation and postcorrelation for 3 stations simultaneously (expandable to 7 stations). Data correlation will be to the rate of data acquisition. Data input will be VLBI data tapes.

In order to calibrate the VLBI data, known tone signals of constant frequency are injected in the microwave subsystem during a VLBI observation. During correlation, the correlator assembly will generate (with a local model) this same frequency and (by comparison to the injected constant frequency) measure phase change due to phase instabilities in the microwave and receiver subsystems.

Given a set of parameters, the software model calibrates the phase to within 10^{-5} cycle of fringe. Also, a record of the calculations along with their results are kept with a precision of 10^{-5} cycle of fringe. Output is available in both delay and frequency domain.

The software model constantly updates its computation of the required geometric time delay lag due to the Earth's rotation. Eight instantaneous lags (four preceding and four follow-

ing the nominal geometric delay) will be provided to determine the actual geometric delay. The maximum equivalent error of the VLBI Block II processor tracking the model delay (the error in keeping constant the point of maximum correlation) will be 0.01 lag.

The JPL/Caltech VPS will be able also to process data collected and recorded by Goddard Space Flight Center (GSFC)/Haystack Observatory Mark III VLBI System.

Postcorrelation functional requirements are as follows:

(1) Compute natural radio source and tone phase.

(2) Calibrate natural radio source phase for station instrument error.

(3) Compute calibrated natural radio source delay.

(4) Resolve cycle ambiguities.

(5) Calibrate for transmission media effects.

(6) Solve for natural radio source locations and baseline.

(7) Update radio source catalog and baselines.

References

1. Chaney, W. D., and Ham, N. C., "DSN VLBI System MK I-80," *TDA Progress Report 42 56*, Jet Propulsion Laboratory, Pasadena, Calif., April 15, 1980.

2. Chaney, W. D., "The DSN VLBI System Mark IV-85," *TDA Progress Report 42-64*, Jet Propulsion Laboratory, Pasadena, Calif., August 15, 1981.
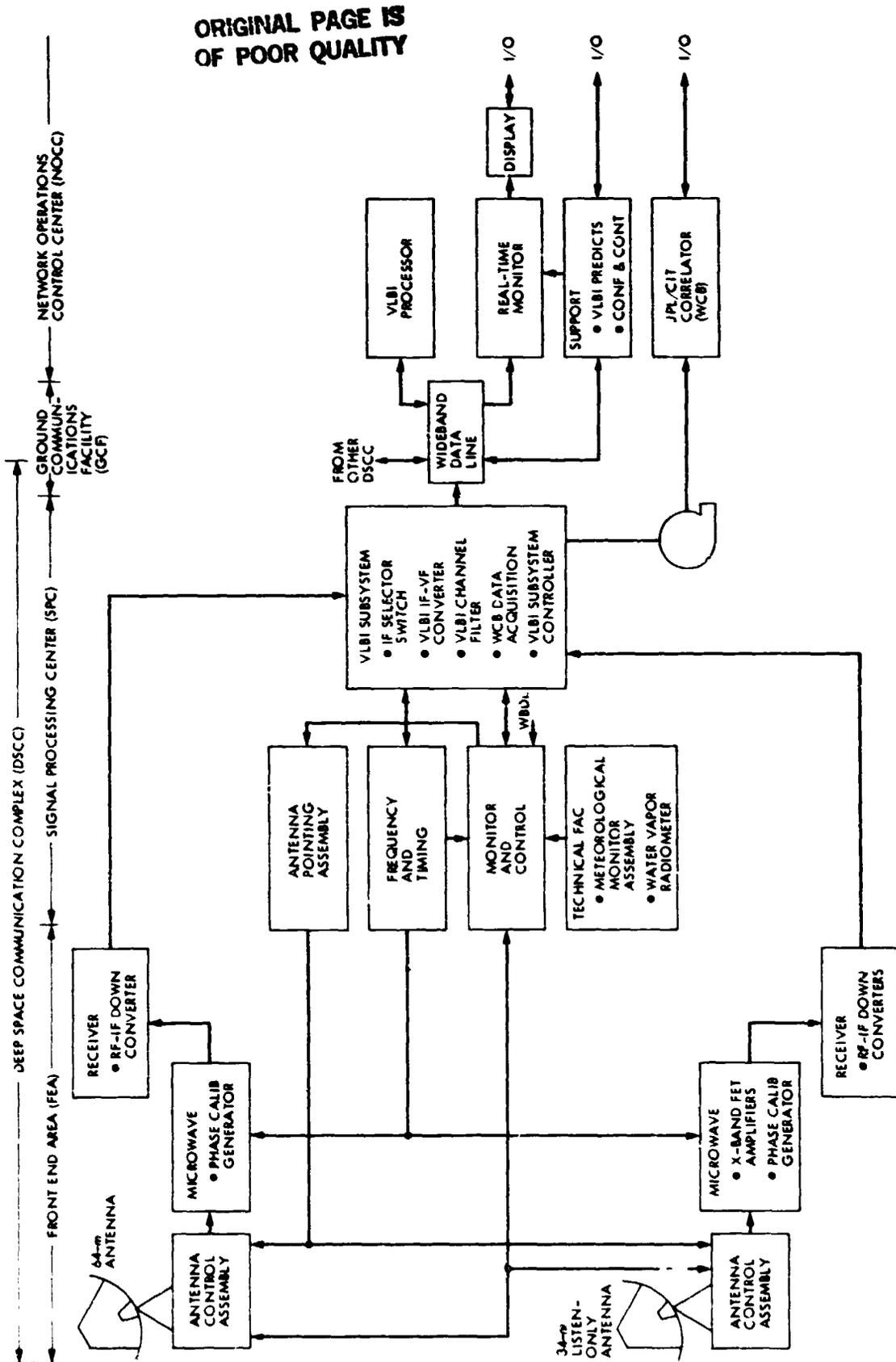
Fig. 1. The DSN VLBI System Mark IV-85 (simplified diagram)

D4-12

Deep Space Payload Launches via the Space Transportation System

A. L. Berman and W. E. Larkin
TDA Mission Support Office

J. C. McKinney
Mission Information Systems Engineering Section

Beginning with tne Galileo spacecraft launch in 1985, deep space payloads will be launched via the Space Shuttle. This change from the previous use of expendable launch vehicles will introduce large changes in procedures and data flow configurations for both the flight project and the Deep Space Network during the launch period. This article describes a typical Galileo launch period sequence of events and telemetry and command data flow configurations.

I. Introduction

Starting with the launch of the Galileo space: raft, now scheduled for 1985, deep space payloads will be launched via the Space Shuttle vehicle of the Space Transportation System (STS), in sharp contrast to all previous launches of deep space payloads via expendable launch vehicles. This very significant change in method of launch will result in large procedural changes for both flight projects and the Deep Space Network (DSN), in its capacity as lead support network for all deep space missions. In a previous article (Ref. 1), the Space Shuttle impact on the DSN initial acquisition was described: in this article, procedural differences for the flight project and DSN during the pre and postlaunch periods are examined. In particular, major subphases of the pre and postlaunch period are identified, and telemetry and command data flow configurations are presented for each subphase. The Galileo mission is used to exemplify a typical launch period sequence of events and data flow configurations.

Major differences for the flight project and DSN in the upcoming Space Shuttle era are the larger number of subphases in the launch period (e.g., Shuttle on-orbit phase of several hours, for which there was no similar phase in the expendables era) and direct launch involvement of additional NASA centers (e.g., Johnson Space Center) and non-NASA facilities (e.g., Air Force Satellite Control Facility).

Section II describes the nominal Galileo sequence of events during the launch period, Section III identifies major launch period subphases, while Section IV identifies the telemetry and command data flow configurations for each of the launch period subphases.

II. The Nominal Galileo Sequence of Events

As previously mentioned, the Galileo spacecraft is currently scheduled for a 1985 launch. The overall launch period begins

with the transporting of the spacecraft to the Kennedy Space Center (KSC), extends through liftoff, and terminates with a successful DSN initial acquisition. Typical prelaunch (prior to liftoff) and postlaunch (following liftoff) Galileo sequences of events are described in detail below.

A. Nominal Galileo Prelaunch Sequence of Events

The Galileo spacecraft is transported from the Jet Propulsion Laboratory (JPL) to the Kennedy Space Center (KSC) by a United States Air Force C-5A aircraft. After being unpacked in the Spacecraft Assembly and Encapsulation Facility (SAEF), the spacecraft is inspected to determine if any damage was sustained during transit. Baseline tests, including the use of S- and X-band radio frequency (RF) subsystems, are evaluated. At the conclusion of these tests, Radioisotope Thermoelectric Generators (RTGs) are installed on the spacecraft and tests are rerun using the RTGs as the source of electrical power. When these tests are completed, all power is removed from the spacecraft to ensure that the installation of pyrotechnic devices and the loading of consumables in the propulsion system will be carried out in a safe environment. At the conclusion of the propulsion loading and pyrotechnics installation, the spacecraft will again be activated for further testing and will be prepared for transfer to the Vertical Processing Facility (VPF). At the VPF the spacecraft will be mated to the Spacecraft Injection Module and the two Inertial Upper Stage (IUS) stages. End-to-end tests via the Merritt Island Launch Area DSN Facility (MIL 71), IUS, and Space Shuttle communications routes will be performed. A series of operational tests will also be conducted. Upon satisfactory completion of these tests, the spacecraft will be placed into the storage mode to await shipment to the launch pad. About ten days before launch, the spacecraft will be removed from storage and transported to the launch pad, preceding the Shuttle arrival by two days. After cargo preparation procedures are completed the spacecraft will be installed in the Shuttle Orbiter Bay. The RTGs are then reinstalled in the spacecraft. Final end-to-end communications tests using MIL 71 (hardline only) and Space Transportation System (STS) Tracking Data Relay Satellite (TDRS) communications links will be conducted. These tests should last about four days. At their conclusion the launch countdown will commence.

Figure 1 presents the nominal Galileo prelaunch sequence of events for an early May 1985 launch.

B. Nominal Galileo Postlaunch Sequence of Events

About one hour after liftoff, the project will commence checking out the condition of the spacecraft to see if the vibration and acceleration forces encountered during the

powered portion of the Shuttle flight have affected the observed prelaunch condition of the spacecraft. The telemetry checkout will be via the STS-TDRS communication link. If required, updated commands may be sent to the spacecraft via this same communications link. A decision to continue with the planned flight to Jupiter must be made by L plus 7.5 hours.

If a "go" decision is made, a Spacecraft-Shuttle separation should occur during the sixth Shuttle orbit, about 9.5 hr after launch. During the more favorable periods of the launch window, the separation could be delayed until the seventh or eighth orbit without jeopardizing the objectives of the mission. After separation, an IUS S-band link can be used to route data between the spacecraft and the shuttle. The maximum useful range of this link is 20 kilometers. At longer ranges it will require the IUS communications network to provide IUS performance and Galileo telemetry data.

About 45 minutes after separation, the IUS first-stage engine will burn until the propellant is exhausted; during the engine burn the Galileo transmitter will be turned on. Separation of the first stage will occur about 10 seconds after fuel depletion. Thirty-five seconds later the IUS second stage will commence a one minute and 40-second burn. Fifty seconds after burnout, the second stage will separate. The spacecraft transponder will now be the only means of exchanging data between the flight project and the spacecraft. Fifteen seconds later the injection module engine will start, and shortly after engine start the spacecraft boom will be deployed. After an 85-second burn, the engine will stop for approximately 10 seconds, and then restart for a 50-second burn. During this second burn, the spacecraft will be rotated to achieve spin stability. After fuel depletion, the injection module will separate from the spacecraft. The injection module will then perform a maneuver to avoid the same trajectory path as the Galileo spacecraft.

Figure 2 presents the nominal Galileo postlaunch sequence of events.

III. Major Launch Subphases During a Shuttle Deep Space Payload Launch

The overall launch period of a deep space payload launch is conveniently subdivided into three major categories, as follows:

(1) Prelaunch phase.

(2) Shuttle attached phase.

(3) Shuttle detached phase.

These are described as follows

A. Prelaunch Phase

This phase starts with spacecraft assembly in the Spacecraft Assembly and Encapsulation Facility, and includes that time the spacecraft spends in test at the Vertical Processing Facility, and finally the move to the launch pad. This phase terminates with liftoff from the pad. Subphases during the prelaunch period are identified as:

(1) Spacecraft Assembly and Encapsulation Facility.

(2) Vertical Processing Facility.

(3) Launch pad.

B. Shuttle Attached Phase

This phase starts at the moment of liftoff, and proceeds through the Shuttle ascent and the Shuttle on-orbit operations. This phase is terminated when the IUS-spacecraft is placed outside the Shuttle Orbiter. Subphases during the shuttle attached phase are identified as:

(1) Shuttle ascent.

(2) Shuttle on-orbit.

C. Shuttle Detached Phase

This phase starts when the IUS-spacecraft is placed outside the Shuttle vehicle, and proceeds through the IUS burn and spacecraft injection module burn. This phase terminates with a successful DSN initial acquisition. Subphases during the Shuttle detached phase are identified as:

(1) IUS burn.

(2) Spacecraft injection module burn.

IV. Galileo Telemetry and Command Data Flow Configurations During the Launch Period

There are 4 major telemetry and 2 major command data flow configurations during the launch period, and these are detailed as follows:

A. Galileo Spacecraft—DSN

This path exists for both telemetry and command data. The link between the spacecraft and the Merritt Island Launch Area (MILA) DSN facility (MIL 71) is both radio frequency (RF) and hardline. Communications from MIL 71 to the JPL Mission Control and Computing Center (MCCC) is via the JPL Ground Communications Facility (GCF).

B. Galileo Spacecraft—TDRSS

This path is for telemetry data only. The link between the spacecraft and the MILA Ground Spacecraft Tracking and Data Network (GSTDN) station is RF. From the GSTDN station an RF uplink is established to the Tracking and Data Relay Satellite (TDRS). Alternately, an RF link can be established directly from the spacecraft to TDRS. From TDRS, an RF downlink is established to the White Sands Ground Station (WSGS). From there, the data is transmitted via domestic satellite (DOMSAT) to the Goddard Space Flight Center (GSFC) NASA Communications (NASCOM) switching center, and thence through DOMSAT to JPL MCCC.

C. Galileo Spacecraft—IUS

This path is for telemetry data only. Galileo telemetry is embedded in IUS telemetry. An RF link is established from the IUS to the Air Force Space Ground Link System (SGLS). The Galileo/IUS data is transmitted to the Air Force Satellite Control Facility (AFSCF), where Galileo telemetry is stripped out and transmitted via GSFC NASCOM switching to JPL MCCC.

D. Galileo Spacecraft—Shuttle

This path is for both telemetry and command. For telemetry, the link can be direct from the Galileo spacecraft to the Shuttle Orbiter, or embedded in IUS telemetry data to the Shuttle Orbiter. From Shuttle the link is RF to TDRS to WSGS. From WSGS, the data is transmitted via GSFC NASCOM to Johnson Space Center (JSC) Mission Control Center (MCC). From JSC MCC, Galileo telemetry is stripped out and transmitted via GSFC NASCOM to JPL MCCC. Alternately, IUS/Galileo telemetry is transmitted via GSFC NASCOM to AFSCF, where Galileo telemetry is stripped out and provided through GSFC NASCOM to JPL MCCC.

For command, the link begins with the IUS Control Center, in conjunction with voice communication from JPL MCCC. From the IUS Control Center, the link is to WSGS through GSFC NASCOM, and then RF to TDRS to the Shuttle Orbiter. From the Orbiter, the link is either hardline or RF (IUS-Shuttle Orbiter distance <20 km) to IUS, and finally, to the Galileo spacecraft. For this mode, only a series of eight commands ("discrete commands") is possible.

Figures 3 through 12 illustrate the above data flow paths for the various launch subphases.

Reference

1. Khatib, A. R., Berman, A. L., and Wackley, J. A., "Space Shuttle Launch Era Spacecraft Injection Errors and DSN Initial Acquisition", in *The TDA Progress Report 42-64*, Jet Propulsion Laboratory, Pasadena, California, pp. 80-82, August 15, 1981.

| SPACECRAFT EVENT	FACILITY	MIL 71	ACTIVITY	1985 FEBRUARY	MARCH	APRIL	MAY
• INSPECTION, BASELINE TEST	S/C ASSEMBLY AND ENCAPSULATION	✓	S-AND X-BAND T/R	▭			
• INSTALL/TEST RTG	S/C ASSEMBLY AND ENCAPSULATION	✓	S-AND X-BAND T/R COAX (TLM)	▫			
• PROPULSION PREPARATIONS	S/C ASSEMBLY AND ENCAPSULATION	−	NU POWER FOR SPACECRAFT		▭		
• BASELINE TEST	S/C ASSEMBLY AND ENCAPSULATION AND VERTICAL PROCESSING	✓	S-BAND RECEIVE COAX (TLM), TDRS, IUS		▫		
• INJECTION MODULE/IUS MATE AND TEST	VERTICAL PROCESSING	✓	S-BAND R. 'EIVE COAX (TLM)			▭	
• STORAGE AND BASELINE TEST	VERTICAL PROCESSING	✓	S-BAND RECEIVE COAX (TLM)			▭	
• INSTALL IN SHUTTLE CARGO BAY	LAUNCH PAD	−	NO POWER FOR SPACECRAFT				▫
• RTG FINAL INSTALL AND TEST	LAUNCH PAD	✓	S-BAND RECEIVE COAX (TLM)				▫
• END-TO-END TEST	LAUNCH PAD	✓	S-BAND RECEIVE COAX (TLM), IUS, TDRS VIA SHUTTLE				▫
• LAUNCH	LAUNCH PAD	✓	TRACKING AND TELEMETRY VIA IUS AND TDRSS NETWORK				▽

Fig. 1. Nominal Galileo prelaunch sequence of events for an early May 1985 launch

25

LAUNCH + HOURS

- SPACECRAFT CHECKOUT
- JPL GO-NO GO
- SHUTTLE - IUS SEPARATION *
- SPACECRAFT TRANSMITTER ON *
- IUS No. 1 BURN - SEPARATION *
- IUS No. 2 BURN - SEPARATION *
- INJECTION MODULE *
- SPACECRAFT SPIN-UP *
- DEPLOY HIGH-GAIN ANTENNA *
- SUN AND DSN ACQUISITION *

9 HOURS, 15 MINUTES

9 HOURS, 17 MINUTES

BURN 9 HOURS, 22 MINUTES

SEPARATION AND FLY AWAY 9 HOURS, 50 MINUTES

STS ORBIT

*THIS SEQUENCE COULD BE DELAYED 1 OR 2 ORBITS

Fig. 2. Nominal Galileo postlaunch sequence of events for an early May 1985 launch

Fig. 3. Telemetry data flow in the prelaunch phase—Vertical Processing Facility subphase

① GALILEO TELEMETRY (600 AND 1200 B/S)
② GALILEO/IUS TELEMETRY

③ SHUTTLE ORBITER DATA TELEMETRY
④ GALILEO TELEMETRY (≤ 115.2 KB/S)

*ALSO HARDLINE

27

Fig. 4. Telemetry data flow in the prelaunch phase—launch pad subphase

① GALILEO TELEMETRY (600 AND 1200 B/S)
② GALILEO/IUS TELEMETRY
③ SHUTTLE ORBITER DATA TELEMETRY
④ GALILEO TELEMETRY (≤ 115.2 KB/S)
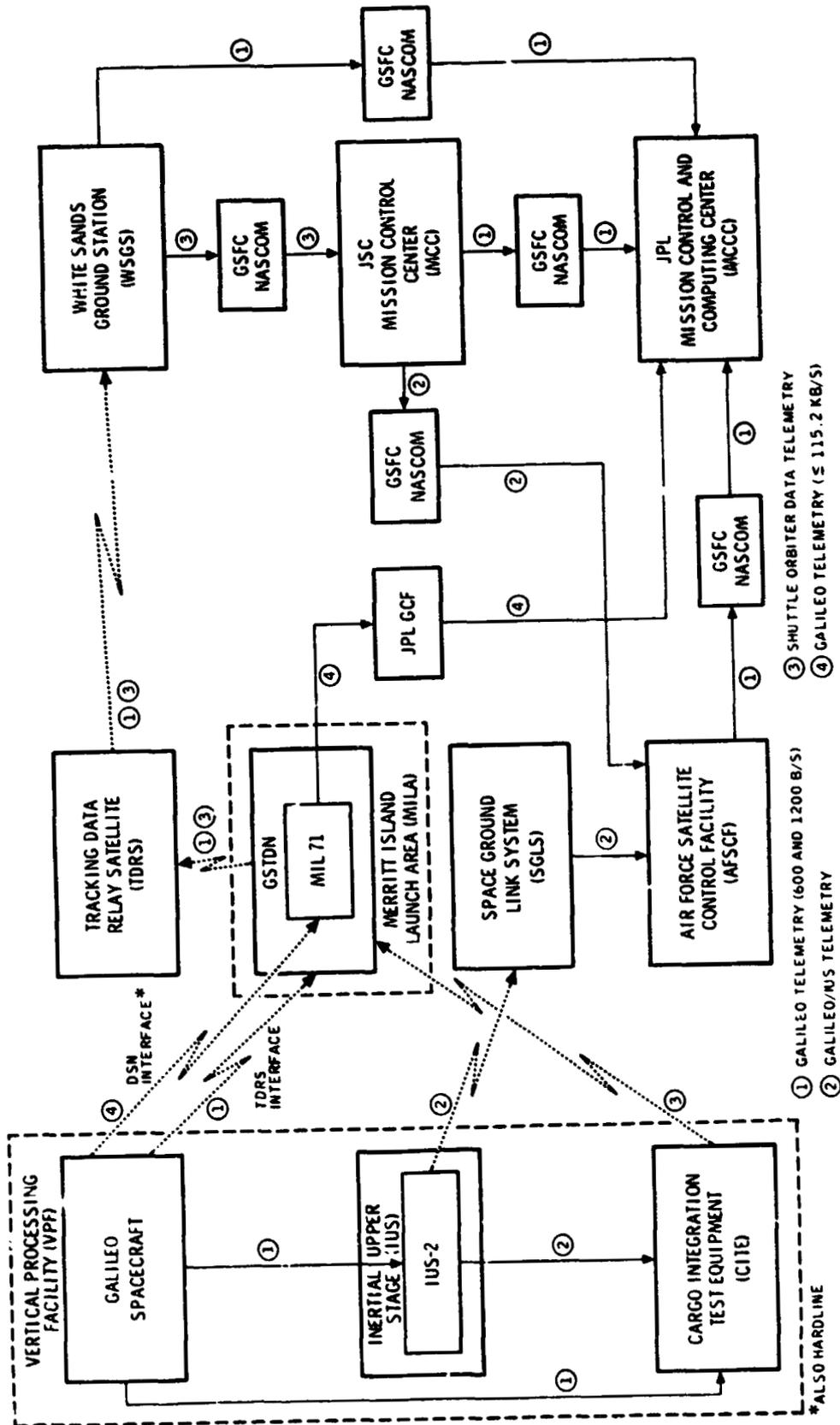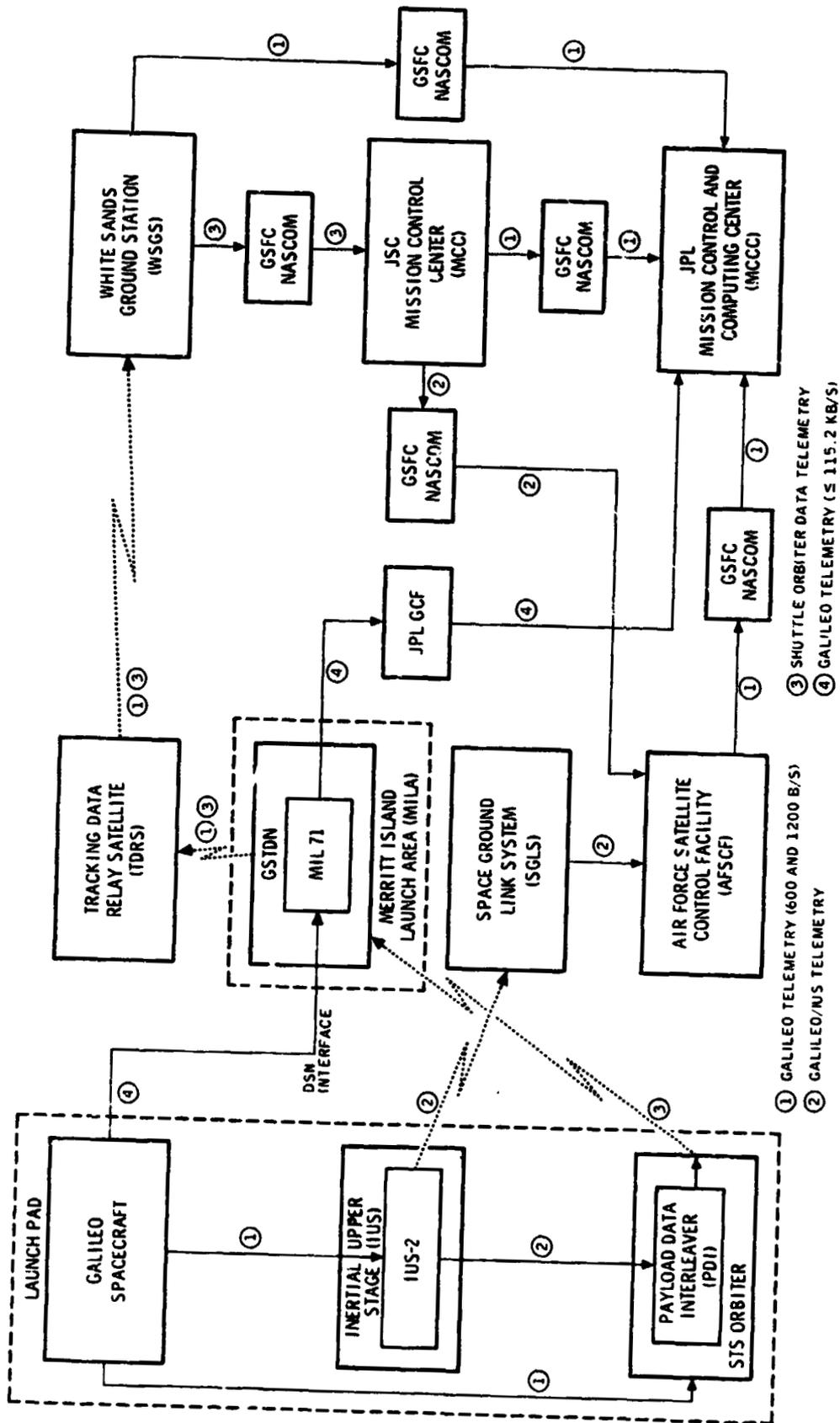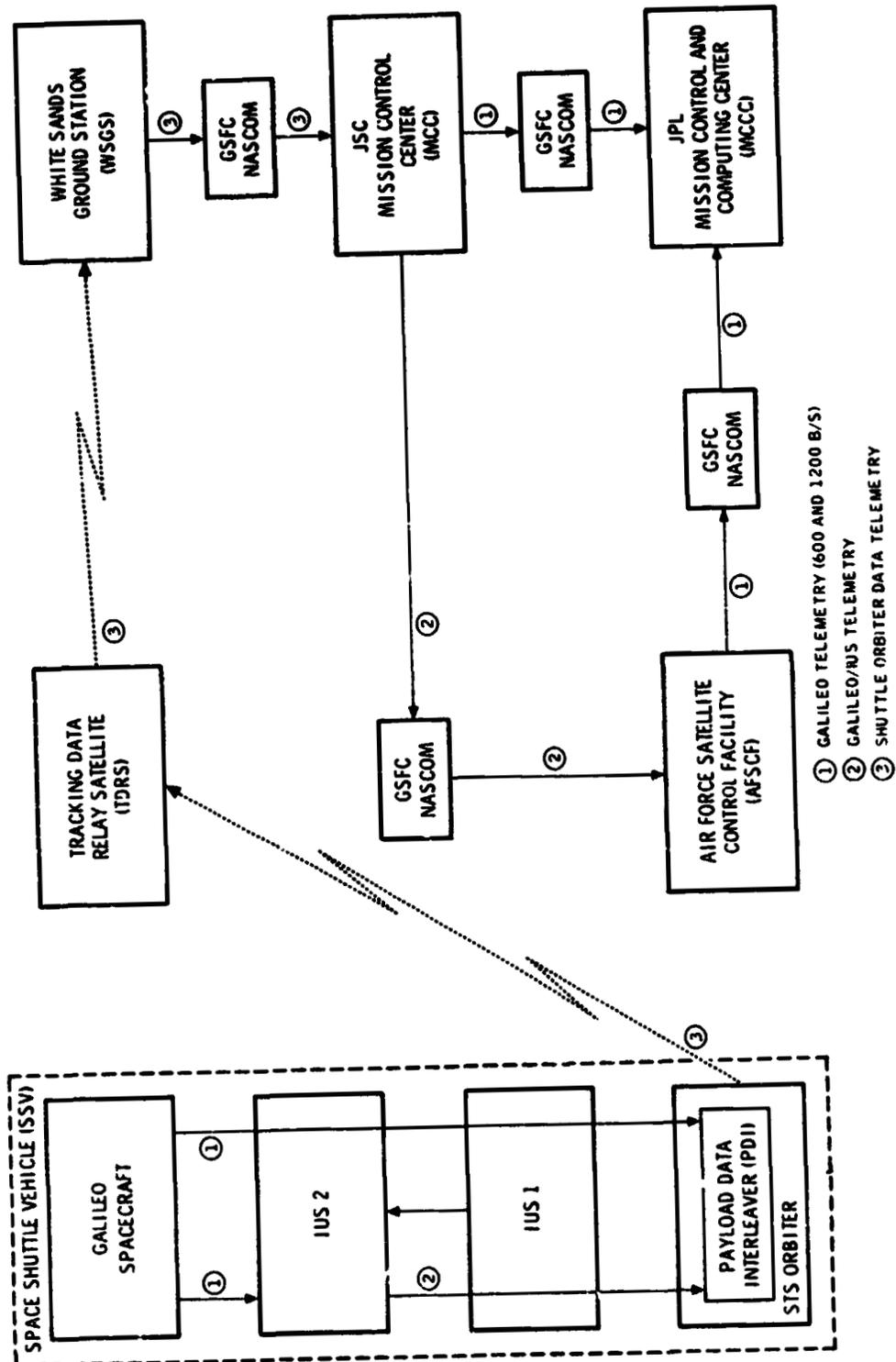
Fig. 5. Telemetry data flow in the Shuttle attached phase—ascent subphase
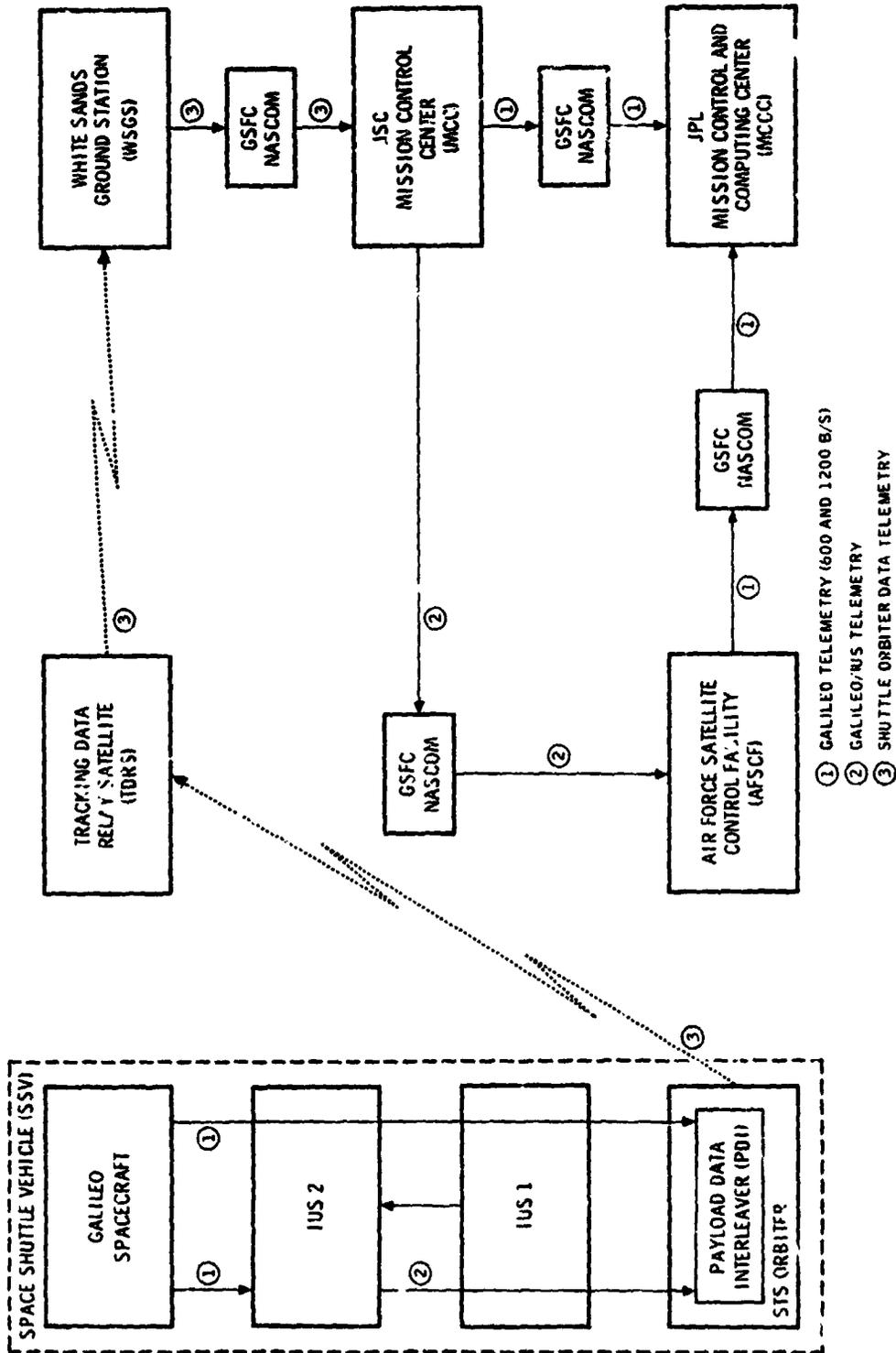
① GALILEO TELEMETRY (600 AND 1200 B/S)
② GALILEO/IUS TELEMETRY
③ SHUTTLE ORBITER DATA TELEMETRY

Fig. 6. Telemetry data flow in the Shuttle attached phase—Shuttle on-orbit phase
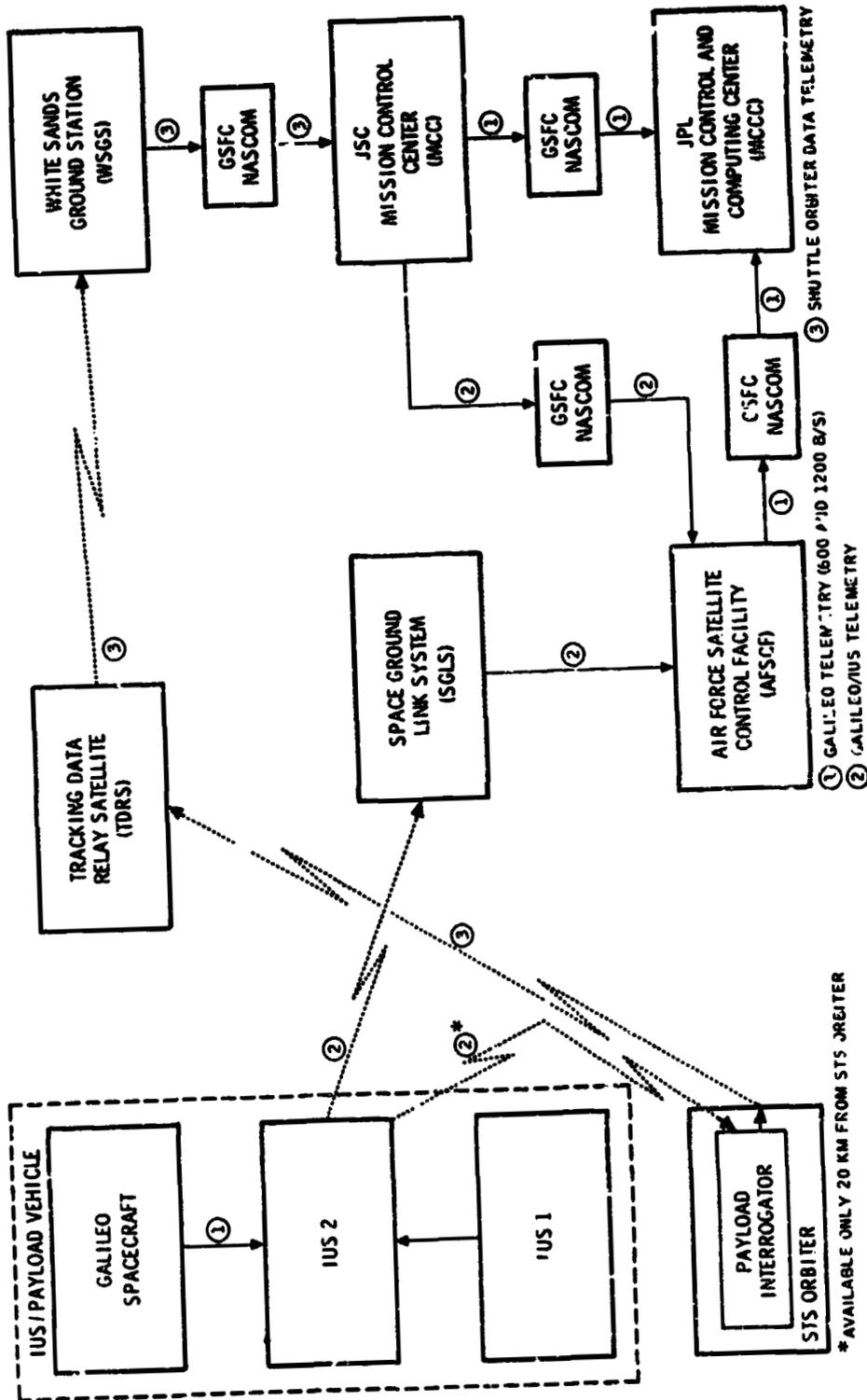
Fig. 7. Telemetry data flow in the Shuttle detached phase—IUS burn subphase

① GALILEO TELEMETRY (600 AND 1200 B/S)
② GALILEO/IUS TELEMETRY
③ SHUTTLE ORBITER DATA TELEMETRY

*AVAILABLE ONLY 20 KM FROM STS ORBITER
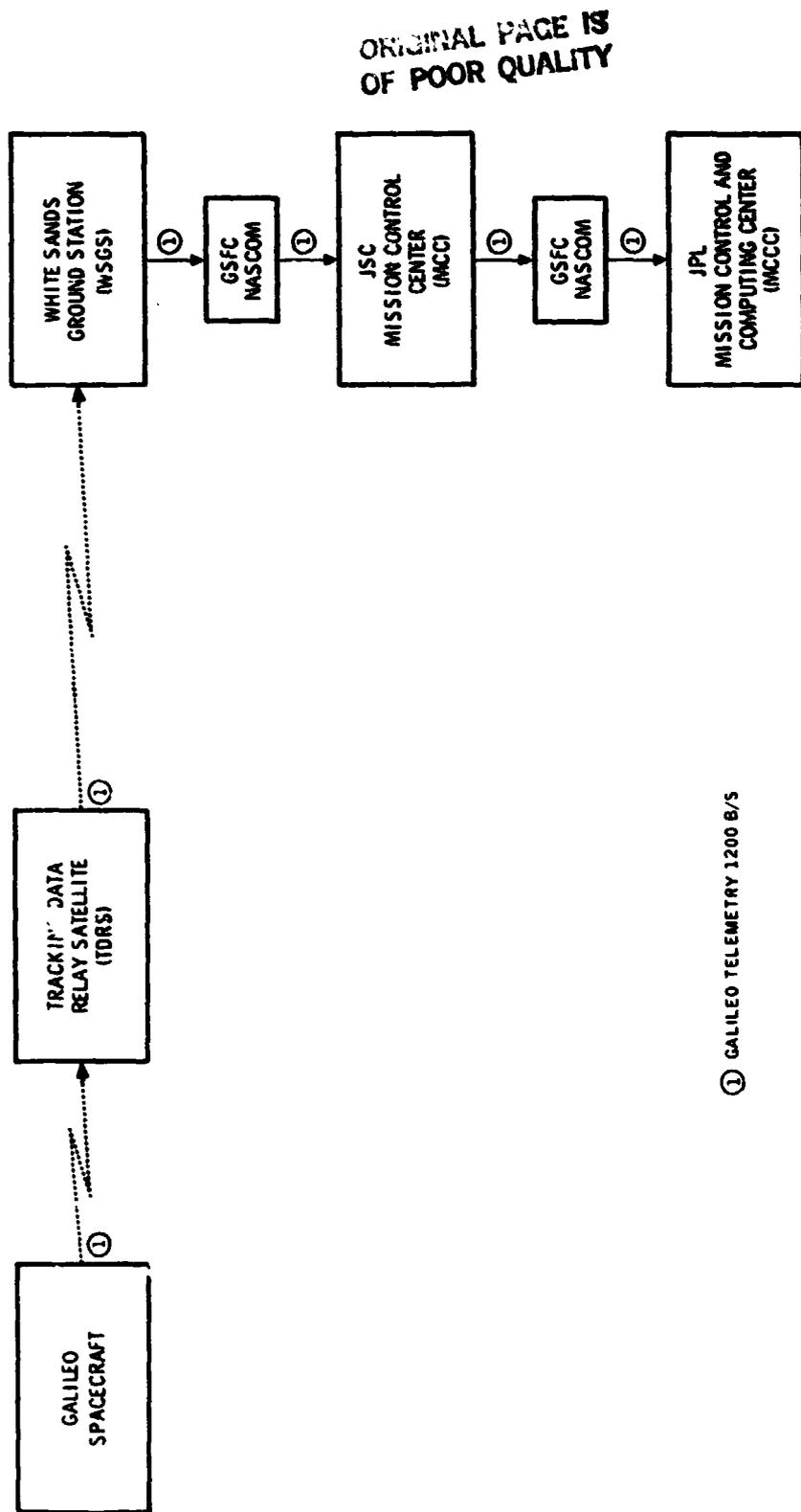
Fig. 8. Telemetry data flow in the Shuttle detached phase—spacecraft injection module burn subphase

① GALILEO TELEMETRY 1200 B/S

WHITE SANDS
GROUND STATION
(WSGS)

GSFC
NASCOM

JSC
MISSION CONTROL
CENTER
(MCC)

GSFC
NASCOM

JPL
MISSION CONTROL AND
COMPUTING CENTER
(MCCC)

TRACKING DATA
RELAY SATELLITE
(TDRS)

GALILEO
SPACECRAFT
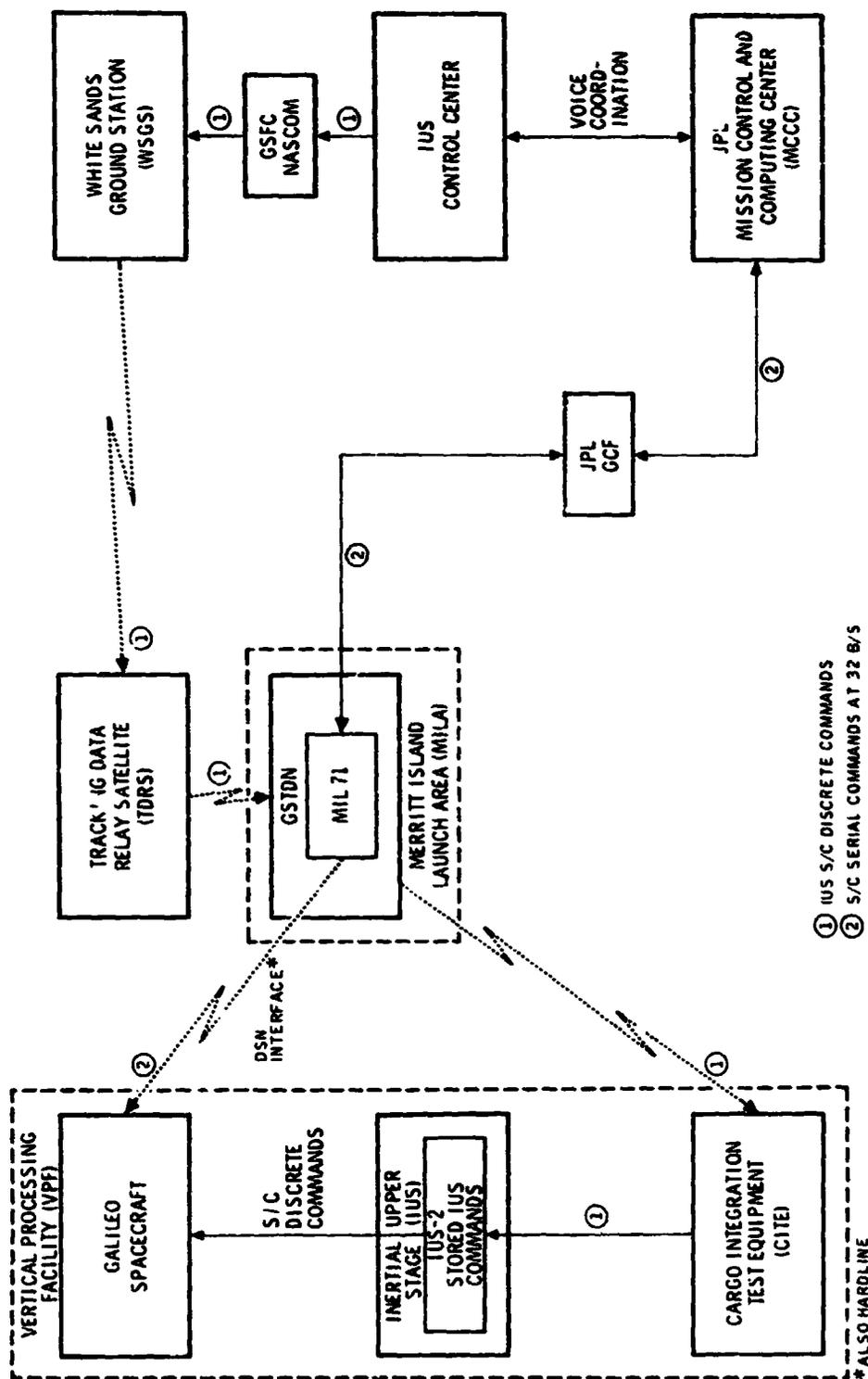
Fig. 9. Command data flow in the prelaunch phase—Vertical Processing Facility subphase
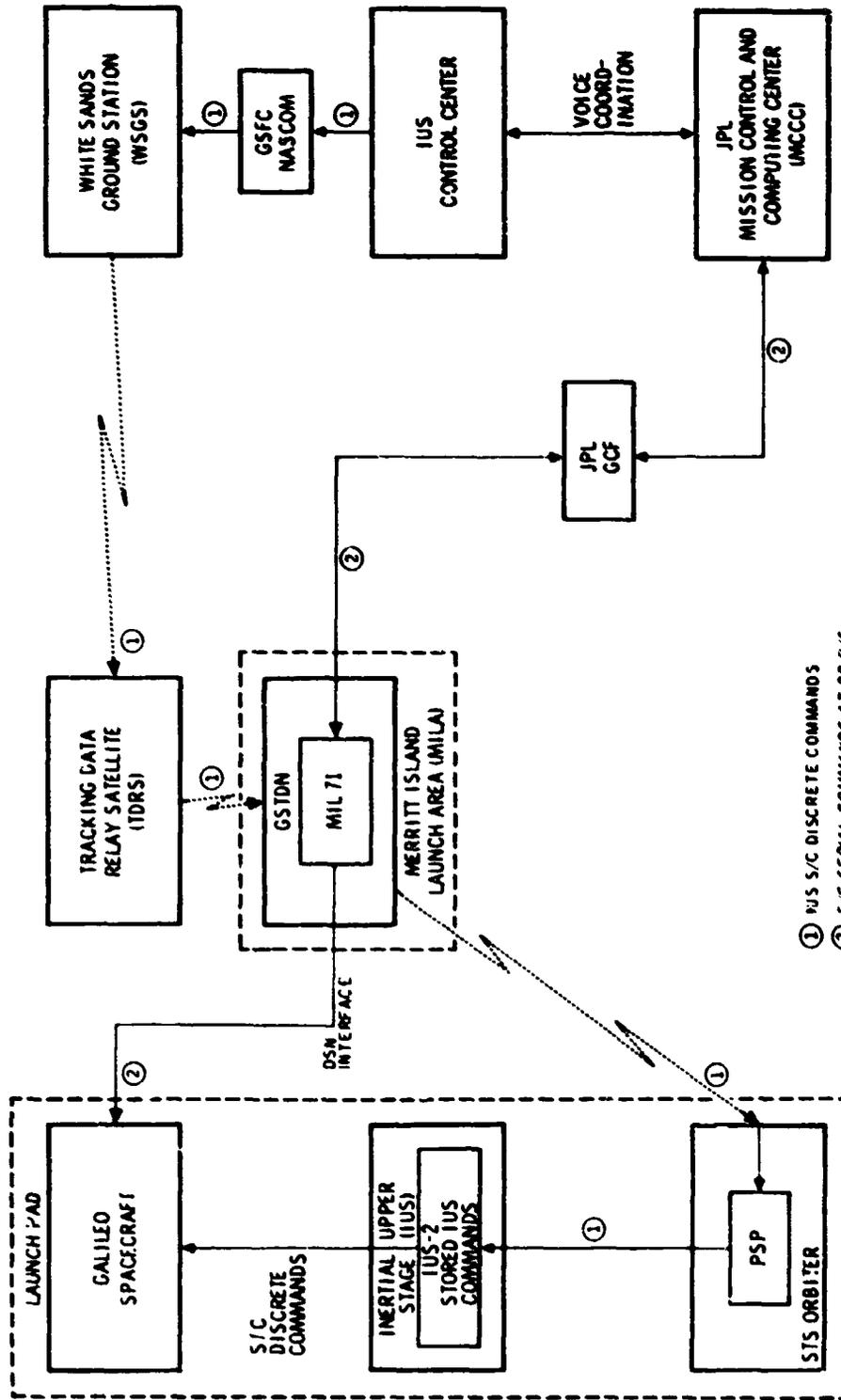
Fig. 10. Command data flow in the prelaunch phase—launch pad subphase

① IUS S/C DISCRETE COMMANDS

② S/C SERIAL COMMANDS AT 32 B/S
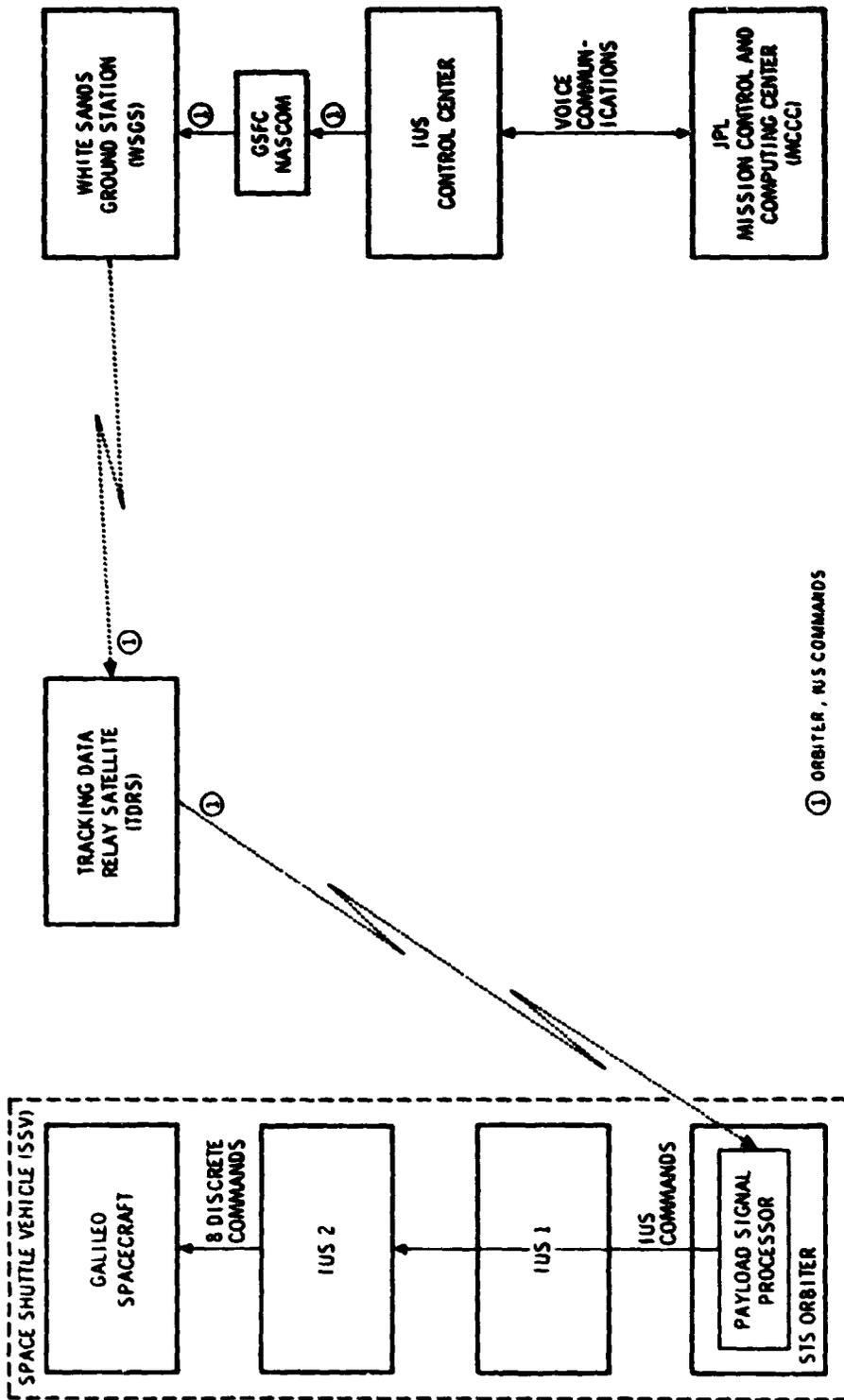
Fig. 11. Command data flow in the Shuttle attached phase—Shuttle on-orbit subphase

① ORBITER , IUS COMMANDS

Fig. 12. Command data flow in the Shuttle detached phase—IUS burn phase
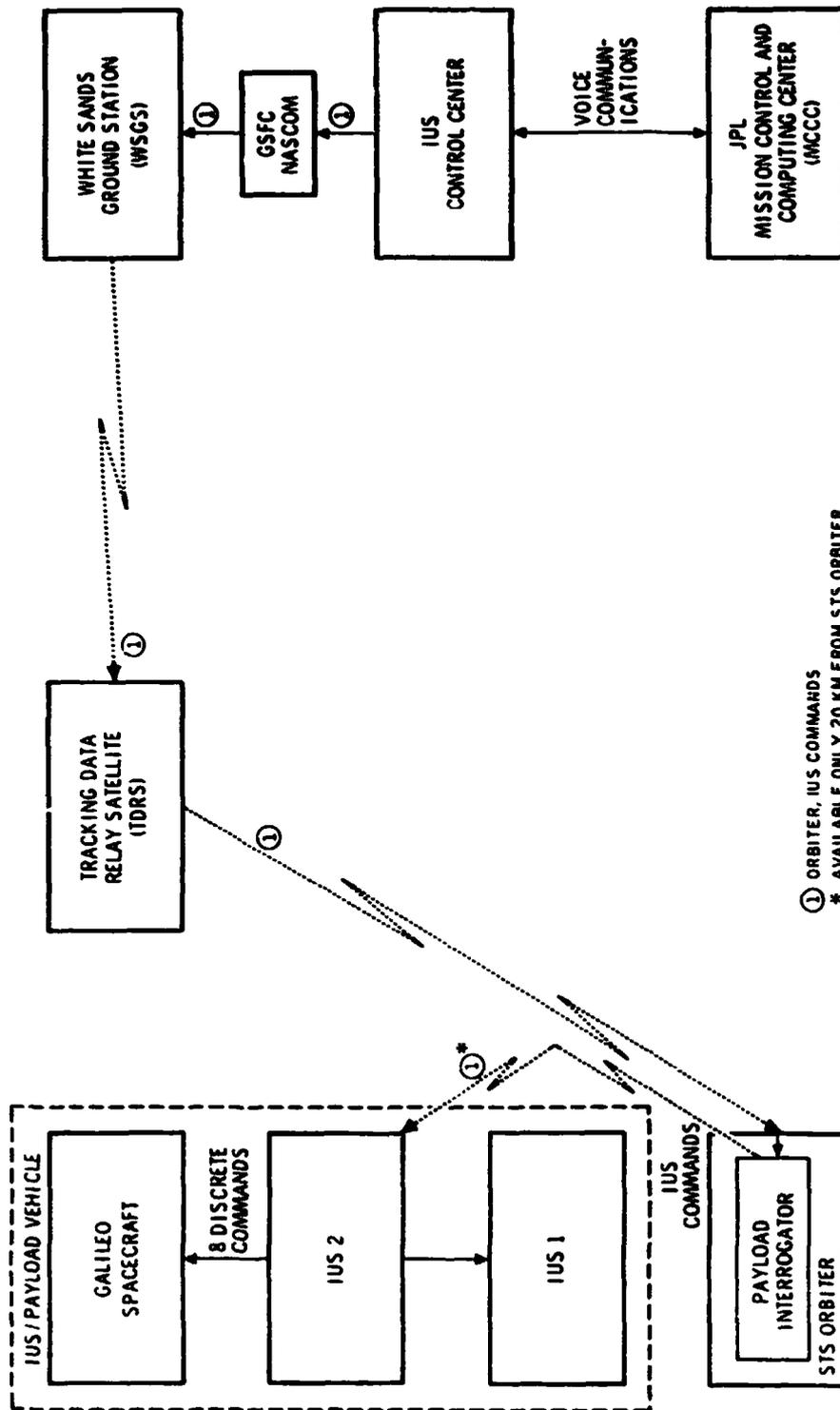
① ORBITER, IUS COMMANDS
* AVAILABLE ONLY 20 KM FROM STS ORBITER

A Structure Function Representation Theorem With Applications to Frequency Stability Estimation

C. A. Greenhall

Communications Systems Research Section

Random processes with stationary nth differences serve as models for oscillator phase noise. The theorem proved here allows one to obtain the structure function (covariances of the nth differences) of such a process in terms of the differences of a single function of one time variable. In turn, this function can easily be obtained from the spectral density of the process. The theorem is used for computing the variance of two estimators of frequency stability.

I. Introduction

Let the output of a precision oscillator be modelled by $\cos \{2\pi\nu_0 [t + x(t)]\}$, where $x(t)$ is a random process representing the "phase time" noise. The most widely used time-domain measure of oscillator stability is the Allan variance, defined by the ensemble average

$$\sigma_A^2(\tau) = \frac{1}{2\tau^2} E\left[\Delta_\tau^2 x(s)\right]^2, \qquad (1)$$

provided that the expectation exists and is independent of the time s. Here, Δ_τ^2 is the backward 2nd difference operator, given by

$$\Delta_\tau^2 f(t) = f(t) - 2f(t - \tau) + f(t - 2\tau).$$

The theorem given here arose from the desire to compute the performance of estimators of $\sigma_A^2(\tau)$. Suppose that $x(t)$ is given on an interval $0 \leqslant t \leqslant T$. Fix $\tau \leqslant T/2$, and set

$$\xi(t) = \frac{1}{\tau\sqrt{2}} \Delta_\tau^2 x(t).$$

A class of unbiased estimators of $\sigma_A^2(\tau)$ is given by

$$V = \int_{2\tau}^{T} \xi^2(t) w(t) dt, \qquad (2)$$

where $w(t)$ is a weighting function (or measure) whose total weight on $[2\tau, T]$ is 1. Two members of this class are treated in Section IV.

Assume further that $\xi(t)$ is a stationary Gaussian process with zero mean and autocovariance function $R_\xi(t)$. Then $\xi^2(t)$ is a stationary process with mean $\sigma_A^2(\tau)$ and autocovariance function $2R_\xi^2(t)$. It follows that

$$EV = E\xi^2(t) = \sigma_A^2(\tau),$$

$$\text{Var } V = \int_{2\tau}^{T} \int_{2\tau}^{T} 2R_\xi^2(s - t) w(s) w(t) ds dt. \qquad (3)$$

Evidently, to compute Var V, we need to obtain $R_\xi(t)$ from the model for $x(t)$. Let us suppose, temporarily, that $x(t)$ is stationary, with autocovariance function

$$R_x(t) = \text{Cov}\,[x(s+t), x(s)],$$

not depending on s. A straightforward computation yields $E\Delta_\tau^2 x(s) = 0$, and

$$E\Delta_\tau^2 x(s+t)\Delta_\tau^2 x(s) = \delta_\tau^4 R_x(t), \tag{4}$$

where $\delta_\tau^4 = \Delta_\tau^2\Delta_{-\tau}^2$, the central 4th difference operator, also given by

$$\delta_\tau^4 f(t) = f(t-2\tau) - 4f(t-\tau) + 6f(t) - 4f(t+\tau) + f(t+2\tau).$$

The left side of (4) is called the 2nd *structure function* of $x(t)$ (Refs. 1, 2). Letting $t = 0$ in (4), we have

$$\sigma_A^2(\tau) = \frac{1}{\tau^2}\,[3R_x(0) - 4R_x(\tau) + R_x(2\tau)], \tag{5}$$

as pointed out by Barnes (Ref. 3). Consequently, since $|R_x(t)| \leqslant R_x(0) = \text{Var}\,x$,

$$\sigma_A^2(\tau) \leqslant \frac{8\,\text{Var}\,x}{\tau^2},$$

$$\sigma_A^2(\tau) \sim \frac{3\,\text{Var}\,x}{\tau^2} \quad (\tau \to \infty), \tag{6}$$

provided $R_x(t) \to 0\ (t \to \infty)$.

This is for stationary $x(t)$. On the other hand, for actual oscillators a behavior like (6) is observed only for *small* τ, below 1 s for quartz crystal oscillators and 100 s for hydrogen masers. As τ increases, the measured $\sigma_A(\tau)$ decreases to a minimum, then stays constant or increases. Of course, since our measurement times are finite, this observation does not "prove" that $x(t)$ is nonstationary. A stationary process with a huge variance and a tiny bandwidth would explain what we see, for we would be looking only at a small piece of the process. If τ were to increase beyond the scale of our observations, then $\sigma_A(\tau)$ would ultimately behave like (6) again. However, if we want to describe the behavior of $x(t)$ on realistic time intervals, a nonobservable low-frequency cutoff only gets in the way, and eventually has to be driven to zero. It is mathematically easier to use a nonstationary model from the very start.

For modelling oscillator phase noise, it is usually sufficient to let $x(t)$ belong to the class of processes whose 2nd differences $\Delta_\tau^2 x(t)$ are stationary. This class includes the stationary processes and those with stationary 1st differences. Such a process has a two-sided "formal" spectral density $S_x(\omega)$, which can have a singularity at $\omega = 0$ that is strong enough to make

$$\int_{-1}^{1} S_x(\omega)\,d\omega = \infty.$$

Nevertheless, we always have

$$\int_{-\infty}^{\infty} \frac{\omega^4}{1+\omega^4} S_x(\omega)\,d\omega < \infty \tag{7}$$

for this class of processes. An example is the power-law spectrum

$$S_x(\omega) = \frac{K}{|\omega|^\beta},$$

where $1 < \beta < 5$.

A rigorous theory of these processes exists (Ref. 1); basically, it shows that one can plunge ahead with the formalism from stationary processes as long as the integrals converge. For example, the transfer function of the operator Δ_τ^2 is $(1 - e^{-i\omega\tau})^2$. Therefore,

$$E\Delta_\tau^2 x(t+s)\Delta_\tau^2 x(s) = \int_{-\infty}^{\infty} e^{i\omega\tau}\,|1 - e^{-i\omega\tau}|^4$$

$$S_x(\omega)\,\frac{d\omega}{2\pi} + c^2\tau^4. \tag{8}$$

The extra term $c^2\tau^4$ comes from a frequency drift component $ct^2/2$ in $x(t)$. Letting $t = 0$, we obtain the Allan variance (1). By this method, the theoretical Allan variance has been evaluated and tabulated (Ref. 4) for $S_x(\omega) = K/|\omega|^k$, k an integer, $0 \leqslant k \leqslant 4$. (For $k \leqslant 1$, a high-frequency cutoff is provided.) Allowing t to be nonzero appears to make (8) more difficult to evaluate. Yet, for our estimation problem, we do need the full covariance function of the process $\Delta_\tau^2 x(t)$. One longs for the simplicity of (4), with $R_x(t)$ given by the simple Fourier integral

$$R_x(t) = \int_{-\infty}^{\infty} e^{i\omega t} S_x(\omega)\,\frac{d\omega}{2\pi}. \tag{9}$$

This integral does not exist, however, unless $x(t)$ is stationary (or equal to a quadratic polynomial plus a stationary process).

The theorem to be proved here gives an easily computable replacement for $R_x(t)$, valid for all processes with stationary n^{th} differences. Equation (4) is replaced by

$$E\Delta_r^2 x(s+t)\Delta_r^2 x(s) = \delta_r^4 [2 \text{ Re } C(t)] + c^2 r^4, \quad (10)$$

where the (nonunique) function $C(t)$ can be computed by two different methods. Here is the second method: Choose an integer k such that $\omega^k S_x(\omega)$ is integrable near $\omega = 0$. Then let

$$C(z) = I^k \int_0^\infty e^{i\omega z}(i\omega)^k S_x(\omega)\frac{d\omega}{2\pi} \quad (11)$$

in the upper half-plane $\text{Im } z > 0$. The operator I^k instructs the user to integrate k times with respect to z. One may then allow z to be real.

Formally, all we are doing is differentiating (9) k times and integrating k times. If one does this correctly, one easily gets valid results for all the power-law oscillator noise models. Although Lindsey and Chie (Ref. 2) give a number of formulas that generalize (5), they have to assume that either the phase $x(t)$ or the frequency dx/dt is stationary. For flicker FM or random walk FM noise, these assumptions are false. Lindsey and Chie do hint at the need for distribution theory in this situation. Although our method has obvious connections to the analytic representation of distributions (Ref. 5), we use only the elementary theory of real and analytic functions to arrive at the main result.

To illustrate the theorem, let us consider the noise called *random walk frequency modulation*, defined by $S_x(\omega) = K/\omega^4$. For this noise, $\sigma_A^2(\tau)$ is proportional to τ (as we shall soon see); this kind of Allan variance behavior has been observed in hydrogen maser frequency standards for $\tau > 10^4$s (Ref. 6). Taking $k = 4$ in (11), we have

$$C(z) = I^4 \frac{Ki}{2\pi z}$$

$$= \frac{Ki}{2\pi}\frac{z^3 \ln z}{6} \quad (\text{Im } z > 0).$$

in which $\ln z$ is to be analytic in $\text{Im } z > 0$. The function $(z^3 \ln z)/6$ is just a particular solution to the equation $f^{(4)}(z) = 1/z$. Taking $\ln z = \ln |z| + i \text{ Arg } z$ in the upper half-plane, we get

$$2 \text{ Re } C(t) = \begin{cases} 0 & (t > 0) \\ -\frac{K}{6}t^3 & (t < 0) . \end{cases} \quad (12)$$

This function is nothing like the autocovariance function of a stationary process, yet applying the operator δ_r^4 to it gives the result

$$E\xi(s+t)\xi(s) = \frac{1}{2\tau^2}\delta_r^4(2 \text{ Re } C(t))$$

$$= \frac{K\tau}{12} \cdot \begin{cases} (2-r)^3 - 4(1-r)^3 & (0 \leqslant r \leqslant 1) \\ (2-r)^3 & (1 < r < 2) \\ 0 & (r > 2) \end{cases}$$

$$(13)$$

(where $r = |t|/\tau$), which is a perfectly good autocovariance function. In particular,

$$\sigma_A^2(\tau) = E\xi^2(s) = \frac{K\tau}{3} .$$

The result (13) can also be derived by expressing $\xi(t)$ as the output of a filter acting on white noise. The method given here is easier and applies to more general difference operators. Moreover, as Section III shows, all of the power-law spectral models become equally simple. Previously, the odd powers (the "flicker" models) were more difficult to handle than the even powers.

II. The Representation Theorem

A mean-continuous random process $x(t)$ is said to have *stationary n^{th} differences* if, for each real τ, the process $\Delta_r^n x(t)$ is stationary (in the wide sense). For such a process, define the *dc coefficient* c by

$$c = \lim_{T\to\infty} \frac{1}{T}\int_0^T \Delta_1^n x(t)dt$$

(in mean square).

For applications, we shall assume that the nth differences of $x(t)$ are ergodic, so that c is nonrandom. It can then be shown that

$$c\tau^n = E\Delta_r^n x(t). \quad (14)$$

The process $x_0(t) = x(t) - ct^n/n!$ equals a polynomial of degree $<n$ plus a mean-zero process. Associated with $x_0(t)$ is a (two-sided) nonnegative *formal spectral density* $S_x(\omega)$, from which all the covariances of the nth differences of $x(t)$ can be obtained. If $x(t)$ is real, then $S_x(-\omega) = S_x(\omega)$.

In connection with these processes, it is convenient to introduce a general real difference operator L that, when applied to a function $f(t)$, gives the result

$$Lf(t) = \sum_\nu a_\nu f(t + t_\nu) \qquad (15)$$

(a finite sum), where the a_ν are real. Its transfer function is the trigonometric polynomial

$$L(i\omega) = \sum_\nu a_\nu e^{i\omega t_\nu} .$$

For example,

$$\Delta_\tau^n f(t) = \sum_{\nu=0}^n \binom{n}{\nu}(-1)^\nu f(t - \nu\tau) ,$$

$$\Delta_\tau^n(i\omega) = \left(1 - e^{-i\omega\tau}\right)^n$$

A difference operator L is said to have *order n* if its transfer function satisfies

$$L^{(j)}(0) = \sum_\nu a_\nu t_\nu^j = 0 \quad (0 \leqslant j < n) ,$$

$$L^{(n)}(0) = \sum_\nu a_\nu t_\nu^n \neq 0 .$$

Such an operator annihilates all polynomials of degree $<n$, and reduces the degree of other polynomials by n, as shown by the computation

$$Lt^m = \sum_\nu a_\nu (t + t_\nu)^m = \sum_\nu a_\nu \sum_{j=0}^m \binom{m}{j} t_\nu^j t^{m-j}$$

$$\qquad (16)$$

$$= \sum_{j=0}^m \binom{m}{j} L^{(j)}(0) t^{m-j} = \sum_{j=n}^\infty \binom{m}{j} L^{(j)}(0) t^{m-j}$$

(Recall that $\binom{m}{j} = 0$ for $j > m$.) In particular, $Lt^n = L^{(n)}(0)$. If

L and M have orders l and m, and $\Lambda = LM$, then Λ has order $l + m$, and

$$\frac{\Lambda^{(l+m)}(0)}{(l+m)!} = \frac{L^{(l)}(0)}{l!} \frac{M^{(m)}(0)}{m!} \qquad (17)$$

If L is given by (15), then the operator L^* (of the same order) is defined by

$$L^*f(t) = \sum_\nu a_\nu f(t - t_\nu) .$$

As we mentioned, an example of order n is $L = \Delta_\tau^n$, for which $L^{(n)}(0) = n!\tau^n$. Another example, for $n = 2$, is the mixed difference operator $\Delta_\tau\Delta_{\tau'}$, which was used for estimating the relative drift rate of a pair of frequency standards (Ref. 7).

We are now set up to give the main result.

Theorem. Let $x(t)$ be a real process with stationary nth differences, nonrandom dc coefficient c, and spectral density $S_x(\omega)$. Let L and M be real difference operators of order n, and let Λ be the difference operator LM^* of order $2n$. Then, the mixed second moments of the processes $Lx(t)$ and $Mx(t)$ can be put into the form

$$E Lx(s + t) Mx(s) = \Lambda \left[(-1)^n \frac{c^2 t^{2n}}{(2n)!} + 2 \operatorname{Re} C(t) \right] ,$$

$$\qquad (18)$$

where the (nonunique) function $C(z)$ is analytic in $\operatorname{Im} z > 0$, continuous in $\operatorname{Im} z \geqslant 0$, and can be obtained by either of the following recipes:

Recipe 1. Choose an integer k between 0 and $2n$ such that

$$\int_0^1 \omega^k S_x(\omega) d\omega < \infty .$$

Let $C(z) = C_0(z)$, where

$$C_0(z) = \int_0^\infty \left[e^{i\omega z} - \frac{1}{1 + \omega^{2n}} \sum_{j=0}^{k-1} \frac{(i\omega z)^j}{j!} \right] S_x(\omega) \frac{d\omega}{2\pi}$$

$$(\operatorname{Im} z \geqslant 0). \qquad (19)$$

If $k = 0$, then omit the sum.

Recipe 2. Choose k as in Recipe 1. Define

$$B(z) = \int_0^\infty e^{i\omega z} (i\omega)^k S_x(\omega) \frac{d\omega}{2\pi} \quad (\text{Im } z > 0). \quad (20)$$

Let $C(z)$ be any kth integral of $B(z)$ on Im $z > 0$ (in other words, any function such that $C^{(k)}(z) = B(z)$). Extend $C(z)$ to the real line by continuity. (This is always possible.)

Remark. The first term in the brackets in (18), when acted upon by Λ, becomes

$$c^2 \frac{L^{(n)}(0)}{n!} \frac{M^{(n)}(0)}{n!} = E[Lx(t)] E[Mx(t)]. \quad (21)$$

Corollary. Let $x(t)$ have stationary nth differences. The structure function

$$D(t, \tau) = E\Delta_\tau^n x(s + t) \Delta_\tau^n x(s)$$

of the process $x(t)$ can be obtained from a function of one variable, namely 2 Re $C(t)$, by

$$D(t; \tau) = c^2 \tau^{2n} + (-1)^n \delta_\tau^{2n}(2 \text{ Re } C(t)). \quad (22)$$

where

$$\delta_\tau^{2n} = (\delta_\tau^2)^n = (-\Delta_\tau \Delta_{-\tau})^n,$$

the central difference operator of order $2n$.

These results show that 2 Re $C(t)$ contains the same information as $S_x(\omega)$ about the process $x(t)$. We can regard 2 Re $C(t)$ as an analog of the covariance function of a stationary process.

III. Examples for $n = 2$

Oscillator phase noise is often modelled by a linear combination of independent power-law noises with spectra

$$\frac{1}{\omega^4}, \frac{1}{|\omega|^3}, \frac{1}{\omega^2}, \frac{F(\omega/\omega_h)}{|\omega|}, F(\omega/\omega_h),$$

where $F(\omega/\omega_h)$ is an integrable low-pass power response function whose one-sided noise bandwidth is $\omega_h/(2\pi)$. Examples

are the *sharp* cutoff (1 for $|\omega| < \omega_h$, 0 elsewhere) and the *exponential* cutoff $\exp(-|\omega|/\omega_h)$. The above noises are called *random walk FM, flicker FM, white FM, flicker PM,* and *white PM.* (The term "white" will be used even if F is not exactly flat near zero frequency.)

The introduction has shown how to apply the representation theorem to random walk FM; let us proceed to the calculation of 2 Re $C(t)$ for the other noises. In order to express the results in the established notation of the frequency and timing literature (Ref. 4), we use the form

$$S_x(\omega) = K_\alpha |\omega|^{\alpha-2}, \quad K_\alpha = \frac{h_\alpha}{2(2\pi)^\alpha}, \quad (23)$$

where a high-frequency cutoff $F(\omega/\omega_h)$ is also applied when $\alpha \geq 1$ (PM noises).

First, let $\alpha = 0, -1$, or -2 (white, flicker, or random walk FM). Using Recipe 2 of the theorem with $k = 2 - \alpha$, we have

$$B(z) = K_\alpha \int_0^\infty e^{i\omega z} i^k \frac{d\omega}{2\pi} = \frac{K_\alpha i^{k+1}}{2\pi z} \quad (\text{Im } z > 0).$$

$$(24)$$

Since a kth integral of $1/z$ is $(z^{k-1} \ln z)/(k-1)!$, we get

$$C(z) = \frac{K_\alpha i^{k+1}}{2\pi} \frac{z^{k-1}}{(k-1)!} \ln z \quad (\text{Im } z \geq 0), \quad (25)$$

in which the branch of ln z must be analytic in the upper half-plane. We shall let ln t be real for $t > 0$. Let us examine the three cases separately.

White FM: $\alpha = 0, k = 2$.

$$C(z) = \frac{K_0}{2\pi i} z \ln z,$$

$$C(t) = \frac{K_0 t}{2\pi i} \cdot \begin{cases} \ln t & (t > 0) \\ \ln |t| + \pi i & (t < 0), \end{cases}$$

$$2 \text{ Re } C(t) = \frac{h_0}{2} \cdot \begin{cases} 0 & (t \geq 0) \\ t & (t < 0). \end{cases} \quad (26)$$

41

Flicker FM: $\alpha = -1, k = 3$.

$$C(z) = \frac{K_{-1}}{4\pi} z^2 \ln z ,$$

$$2 \, \text{Re} \, C(t) = \frac{h_{-1}}{2} t^2 \ln |t| . \tag{27}$$

Random walk FM. $\alpha = -2, k = 4$.

$$C(z) = \frac{K_{-2} i}{2\pi} \frac{z^3}{6} \ln z ,$$

$$2 \, \text{Re} \, C(t) = \frac{h_{-2} \pi^2}{3} \cdot \begin{cases} 0 & (t \geqslant 0) \\ -t^3 & (t \leqslant 0) . \end{cases} \tag{28}$$

In a sense, the flicker case is easier than the others because we don't have to keep track of the imaginary part of $\ln z$.

For handling the PM noises, the exponential cutoff is easier to use than the sharp cutoff, and may even be more realistic. Let $S_x(\omega) = K_\alpha |\omega|^{\alpha-2} \exp(-|\omega|/\omega_h)$, where $\alpha = 1$ or 2. Again letting $k = 2 - \alpha$, we get

$$B(z) = \frac{K_\alpha i^{k+1}}{2\pi} \frac{1}{z + i/\omega_h} . \tag{29}$$

Flicker PM: $\alpha = 1, k = 1$.

$$C(z) = -\frac{K_1}{2\pi} \ln (z + i/\omega_h) ,$$

$$2 \, \text{Re} \, C(t) = -\frac{h_1}{8\pi^2} \ln (t^2 + 1/\omega_h^2) . \tag{30}$$

For the sharp high-frequency cutoff, $2 \, \text{Re} \, C(t)$ turns out to be a cosine integral.

White PM. $\alpha = 2, k = 0$.

$$C(z) = B(z) = \frac{K_2 i}{2\pi} \frac{1}{z + i/\omega_h} ,$$

$$2 \, \text{Re} \, C(t) = \frac{h_2 \omega_h}{8\pi^3 (1 + \omega_h^2 t^2)} , \tag{31}$$

which is just $R_x(t)$, because now $x(t)$ is stationary.

Fractional noises. It is well known (Ref. 4) that if the spectrum satisfies a power law $S_x(\omega) = K_\alpha |\omega|^{\alpha-2}$, where $-3 < \alpha < 1$, then the Allan variance satifies another power law $\sigma_A^2(\tau) = \text{const} \cdot \tau^\mu$, where $\mu = -1 - \alpha$. When Allan variance measurements are made, fractional values of μ sometimes appear over a certain range of τ. Thus, we ought to show how to use the representation theorem for fractional values of α in the range $-3 < \alpha < 1$. We shall do this for $0 < \alpha < 1$, leaving the other cases as an exercise for the reader. Take $k = 1$. Then

$$B(z) = \frac{K_\alpha i}{2\pi} \int_0^\infty e^{i\omega z} \omega^{\alpha-1} \, d\omega$$

$$= \frac{K_\alpha i}{2\pi} \frac{\Gamma(\alpha)}{(-iz)^\alpha} ,$$

$$C(z) = i \int B(z) \, d(-iz) = -\frac{K_\alpha \Gamma(\alpha)}{2\pi (1 - \alpha)} (-iz)^{1-\alpha} ,$$

where the power functions are analytic in the right half-plane and positive on the positive real axis. Then

$$2 \, \text{Re} \, C(t) = -\frac{h_\alpha \Gamma(\alpha) \sin\left(\frac{1}{2} \pi \alpha\right)}{(2\pi)^{1+\alpha} (1 - \alpha)} |t|^{1-\alpha} , \tag{32}$$

As $\alpha \to 0$, this expression tends to $-h_0 |t|/4$, which is indeed a valid representative of $2 \, \text{Re} \, C(t)$ for $\alpha = 0$.

IV. The Variance of Two Allan Variance Estimators

Let us return to the estimation problem mentioned in the introduction. Recall that phase time $x(t)$ is given for $0 \leqslant t \leqslant T$. Therefore, $\xi(t) = \Delta_\tau^2 x(t)/(\tau\sqrt{2})$ is available for $2\tau \leqslant t \leqslant T$. Set $m = T/\tau$. Two unbiased estimators of the Allan variance (1) are

$$V_\tau = \frac{1}{m-1} \sum_{j=2}^m \xi^2(j\tau) \tag{33}$$

(m an integer ≥ 2), called the τ-overlap estimator, and

$$V_0 = \frac{1}{-(m-2)} \int_{2\tau}^{m\tau} \xi^2(t)\,dt \qquad (34)$$

(m real, >2), called the *continuous* estimator, which, although it cannot be achieved in practice, represents a limiting case for a sample time τ_0 much less than τ. The use of such an estimator was suggested by Hor e, Allan, and Barnes (Ref. 8).

Although the τ-overlap estimator has been used for many years, it is reasonable to ask whether the continuous estimator has a smaller variance. In other words, if $\tau_0 \ll \tau$, should we average all the available samples $\xi^2(j\tau_0)$, or should we use only the samples $\xi^2(j\tau)$? Since the data collection time T may be weeks or months, this question is more than academic.

The answer depends both on m and on the spectrum of the phase noise. Assume that $x(t)$ is a Gaussian process with stationary 2nd differences and zero dc component c. If we know $S_x(\omega)$, then we can compute $C(z)$. By the corollary, the autocovariance function of the mean-zero process $\xi(t)$ is

$$R_\xi(t) = \frac{1}{2\tau^2}\, \delta_\tau^4\, (2\,\mathrm{Re}\,C(t)). \qquad (35)$$

As we said in the introduction, the autocovariance function of the stationary process $\xi^2(t)$ is $2R_\xi^2(t)$. The means and variances of V_τ and V_0 are now computed straightforwardly:

$$E\,V_\tau \quad E\,V_0 = R_\xi(0) = \sigma_A^2(\tau), \qquad (36)$$

$$\mathrm{Var}\,V_\tau = \frac{2}{(m-1)^2} \sum_{j=-m+2}^{m-2} (m-1-|j|)\,R_\xi^2\,(j\tau), \qquad (37)$$

$$\mathrm{Var}\,V_0 = \frac{2}{T_2^2} \int_{-T_2}^{T_2} (T_2 - |t|)\,R_\xi^2\,(t)\,dt, \qquad (38)$$

where $T_2 = (m-2)\tau$. (Of course, we recognize that $R_\xi(t)$ is an even function.)

The computations have been carried out for white FM, flicker FM, and random walk FM (see Section III), all of which have been observed in actual oscillators. The flicker FM results

were computed numerically; the others are in closed form, which, however, we shall not give here. For white FM and random walk FM, the τ-overlap formulas agree with those of Lesage and Audoin (Ref. 9); for flicker FM, the τ-overlap numbers agree with Yoshimura's (Ref. 10).

The results are presented in terms of "degrees of freedom,' defined for a positive estimator V by

$$\mathrm{d.f.} = \frac{2\,(EV)^2}{\mathrm{Var}\,V}.$$

Given d.f., one sometimes uses the appropriate chi-squared distribution for constructing confidence intervals about the estimate (Ref. 8). Whether or not this is done, the d.f. remains a useful figure of merit.

In Fig. 1, for the above three noise types, we plot d.f./$(m-1)$ vs m for V_τ and V_0. For white FM, V_0 is always better than V_τ. For flicker FM, V_0 is better than V_τ except for $m \leq 3$. For random walk FM, V_τ is better than V_0 for $m \leq 18$. Of course, the smaller values of m are more critical, since d.f. is roughly proportional to $m-1$.

It may seem paradoxical for V_τ to be better than V_0, since V_0 uses all the available data. Both estimators are special cases of (2), however; if one looks for the optimal (minimal variance) estimator of the class (2) (for a given noise type), one will probably find that the optimal weighting function $w(t)$ is nonconstant and almost everywhere nonzero. In other words, one should use all the data, but in a nonuniform way.

V. Concluding Summary

Oscillator stability is usually characterized by the behavior of nth order differences of the phase. The theoretical evaluation, from the phase noise spectrum, of the variances and covariances of these differences involves messy trigonometric integrals, such as (9). The messiness is caused by a $(2n)$th order difference operator tangled up inside the integral. Our representation theorem breaks the integral evaluation into two easy steps. (1) evaluation of a much simpler integral depending only on the noise spectrum; (2) application of that same difference operator to the result of step (1).

In effect, the evaluation of these integrals is uncoupled into two independent operations. In Section III, we tabulated only the result of step (1) (the function called $2\,\mathrm{Re}\,C(t)$) for all the usual power-law oscillator noise models. This short "one-dimensional" table, plus another one-dimensional table of difference operators, can generate a two-dimensional table of results as found, for example, in Ref. 4.

As an application, we examined two Allan variance estimators. Because the representation theorem delivers such simple closed forms for the required autocovariance functions, the computations were quickly executed by a simple BASIC program. The theorem can also be used for evaluating the performance of frequency drift estimators. In general, we get an estimator of frequency drift rate (the dc coefficient c) by operating upon oscillator phase with a second-order difference operation L that need not be of form Δ_τ^2. Here, the full generality of the theorem is needed.

References

1. Yaglom, A. M., "Correlation theory of processes with random stationary nth increments," *Amer. Math. Soc. Translations*, Ser. 2, Vol 8, pp. 87-141, 1958.

2. Lindsey, W. C., and Chie, C. M., "Theory of oscillator instability based upon structure functions," *Proc. IEEE*, Vol. 64, pp. 1652-1666, 1976.

3. Barnes, J. A., "Atomic timekeeping and the statistics of precision signal generators," *Proc. IEEE*, Vol. 54, pp. 207-220, 1966.

4. Barnes, J. A., *et al.*, "Characterization of frequency stability," *IEFE Trans. Instrum. Meas.*, Vol. IM-20, pp. 105-120, 1971.

5. Bremermann, H., *Distributions, Complex Variables, and Fourier Transforms*, Addison-Wesley, Reading, Mass., 1965.

6. Kuhnie, P. F., "Hydrogen maser implementation in the Deep Space Network at the Jet Propulsion Laboratory," *Proc. 11th Ann. Precise Time and Time Interval Applications and Planning Meeting*, Goddard Space Flight Center, Greenbelt, Md., 1979.

7. Greenhall, C. A., "Removal of drift from frequency stability measurements," *TDA Progress Report 42-65*, pp. 127-132, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 15, 1981.

8. Howe, D. A., Allan, D. W., and Barnes, J. A., "Properties of signal sources and measurement methods," *Proc. 35th Ann. Symp. on Frequency Control*, USAERADCOM, Fort Monmouth, N.J., pp. A1-A47, 1981.

9. Lesage, P., and Audoin, C., "Estimation of the two-sample variance with limited number of data," *Proc. 31st Ann. Symp. on Frequency Control*, pp. 311-313, 1977.

10. Yoshimura, K., "Characterization of frequency stability: uncertainty due to the autocorrelation of the frequency fluctuations," *IEEE Trans. Instrum. Meas.*, Vol. IM-27, pp. 1-7, 1978.
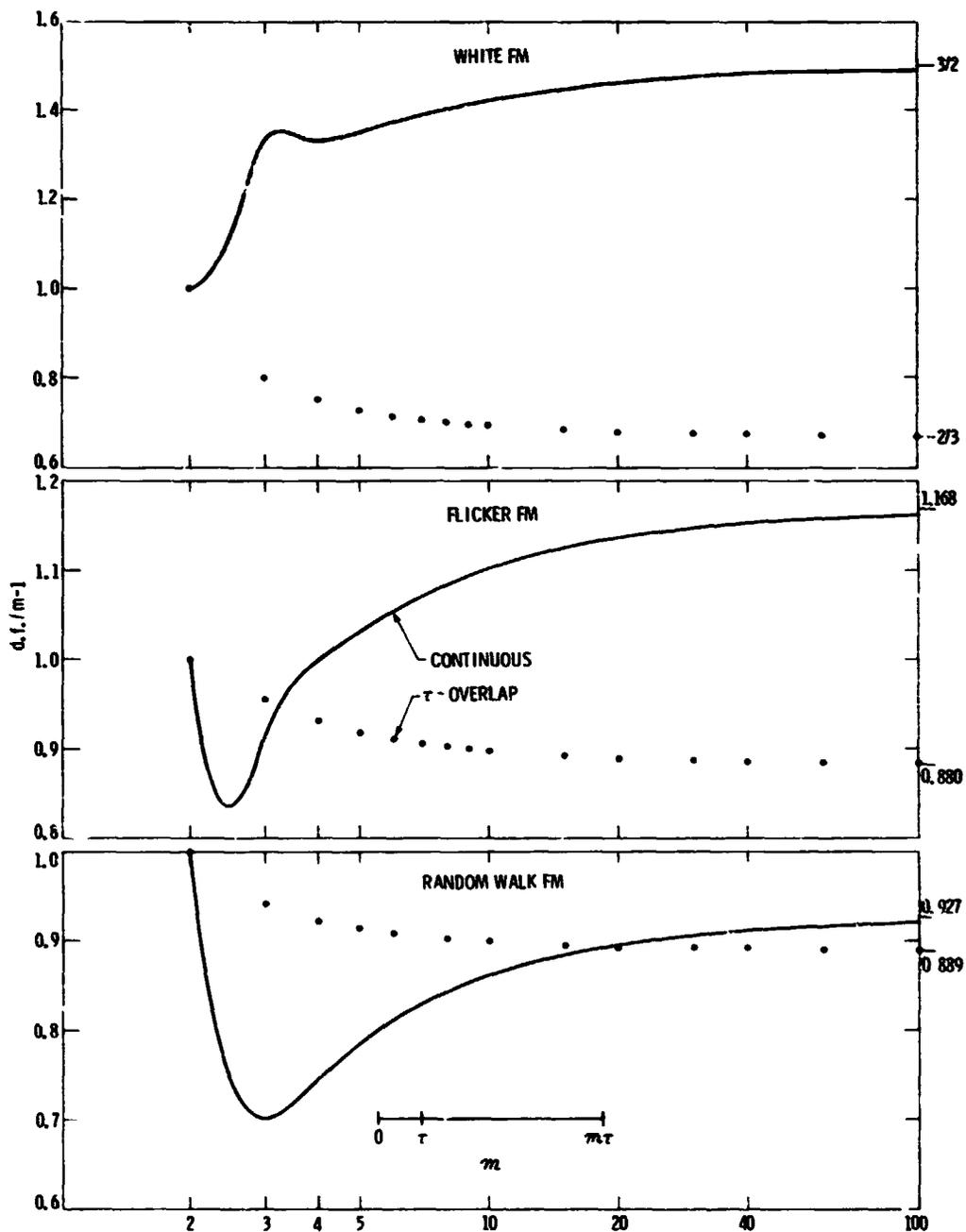
Fig. 1. Degrees of freedom of the τ-overlap and continuous Allan variance estimates for white, flicker, and random-walk frequency modulation. The numbers on the right give the horizontal asymptotes. The observation time is $m\tau$

Appendix

Proof of the Representation Theorem

For convenience, let us first set down an elementary estimate of the Taylor remainder for e^z: If $\mathrm{Re}\, z \leqslant 0, k \geqslant 1$, then

$$\left| e^z - \sum_{j=0}^{k-1} \frac{z^j}{j!} \right| \leqslant \frac{|z|^k}{k!}. \tag{A1}$$

This can be obtained from the integral form of the remainder.

In analogy with the usual notation for mixed structure functions $D(t; \tau_1, \tau_2)$ (Ref. 1), denote the left side of (18) by $D(t; L, M)$. Without loss of generality, we can assume that the dc coefficient c is zero. Begin with the spectral representation of $D(t; L, M)$ as given by Yaglom:

$$D(t; L, M) = \int_{-\infty}^{\infty} e^{i\omega t} L(i\omega) M(-i\omega) S_x(\omega) \frac{d\omega}{2\pi}$$

$$= \int_{-\infty}^{\infty} \Lambda e^{i\omega t} S_x(\omega) \frac{d\omega}{2\pi}$$

$$= 2 \,\mathrm{Re}\, A(t) \ . \tag{A2}$$

where

$$A(z) = \int_{0}^{\infty} \Lambda \, e^{i\omega z} S_x(\omega) \frac{d\omega}{2\pi} \tag{A3}$$

for $z = t + iu$ $(u \geqslant 0)$, and Λ operates on functions of t. The name of the game is pulling Λ outside the integral. If you do this brutally, the integral usually blows up.

Since the spectral density $S_x(\omega)$ always satisfies

$$\int_{0}^{\infty} \frac{\omega^{2n}}{1 + \omega^{2n}} S_x(\omega) d\omega < \infty$$

(Ref. 1), the integer k specified in Recipe 1 exists. We can always take $k = 2n$; it is often possible and desirable to use a smaller k. If $k = 0$ works, then $S_x(\omega)$ is integrable, Λ comes outside the integral in (A3), and we are done. In this case, $C(z) = A(z)$, and $x(t)$ is a polynominal plus a stationary process whose autocovariance function is $2\,\mathrm{Re}\, C(t)$.

Assume $k \geqslant 1$. Since, for all $u > 0$,

$$\int_{0}^{\infty} e^{-\omega u} \omega^k S_x(\omega) d\omega < \infty,$$

the function $B(z)$ of Recipe 2 exists and is analytic in $\mathrm{Im}\, z > 0$. Let us now look at $C_0(z)$ as given in Recipe 1. It will soon be shown that

(a) $C_0(z)$ is continuous on $\mathrm{Im}\, z \geqslant 0$ and analytic on $\mathrm{Im}\, z > 0$;

(b) $C_0^{(k)}(z) = B(z)$ for $\mathrm{Im}\, z > 0$;

(c) $\Lambda C_0(z) = A(z)$ for $\mathrm{Im}\, z \geqslant 0$.

Assuming (a) - (c), we see that $C_0(z)$ is a kth integral of $B(z)$ on $\mathrm{Im}\, z > 0$; any other k^{th} integral $C(z)$ differs from $C_0(z)$ by at most a polynomial of degree $< k$. Since Λ annihilates all polynomials of degree $< 2n$, properties (a) - (c) hold with $C_0(z)$ replaced by $C(z)$, and the theorem follows from (A2).

To prove (a) - (c) denoted the bracketed expression in (19) by $E(z, \omega)$, which splits into the two parts

$$E_1(z, \omega) = \frac{1}{1 + \omega^{2n}} \left[e^{i\omega z} - \sum_{j=0}^{k-1} \frac{(i\omega z)^j}{j!} \right] ,$$

$$E_2(z, \omega) = \frac{\omega^{2n}}{1 + \omega^{2n}} e^{i\omega z} \ .$$

Accordingly, $C_0(z)$ splits into two integrals $C_1(z)$ and $C_2(z)$. Since $C_2(z)$ is the Fourier transform of an integrable function, $C_2(z)$ is continuous on $\mathrm{Im}\, z \geqslant 0$; on $\mathrm{Im}\, z > 0$ it is analytic and can be differentiated repeatedly under the integral sign.

Differentiating E_1 repeatedly and applying (A1), we get the bound

$$\frac{\partial^p}{\partial z^p} E_1(z, \omega) \leqslant \frac{1}{1 + \omega^{2n}} \frac{a^{k-p} \omega^k}{(k - p)!} , \tag{A4}$$

valid for $0 \leqslant p \leqslant k$, $\mathrm{Im}\, z \geqslant 0$, $|z| \leqslant a$. The right side of (A4) times $S_x(\omega)$ is integrable. Therefore, the function $C_1(z)$, on the domain $\mathrm{Im}\, z \geqslant 0$, can be differentiated k times under the integral sign. In particular, $C_1(z)$ is continuous on this domain. This proves (a), and shows that $C_0(z)$ can be differentiated k times under the integral sign on $\mathrm{Im}\, z > 0$. Doing so gives (b). Property (c) is true because Λ annihilates polynomials of degree $k - 1$. The theorem is proved.

PPM Demodulation for Reed-Solomon Decoding
for the Optical Space Channel

D. Divsalar, R. M. Gagliardi, and J. H. Yuen

Communications Systems Research Section

Optical communications over space channels (satellite-to-satellite or deep-space-to-relay-satellite) are commonly designed as pulse-position-modulated (PPM) laser links. When coding is needed to improve the link performance, it is advantageous to use Reed-Solomon (RS) block codes over the PPM frames to obtain the largest degree of error correction. Since RS codes can correct both symbol errors and symbol erasures, a question arises as to the best way to demodulate the PPM laser fields in order to generate the input symbols for the RS decoder. The method selected for demodulating (converting the received laser field to digital symbols) will define the erasure and transmitted symbols of the laser link, and therefore will determine the word error probabilities of the system. In this paper, several possible demodulating schemes were considered, and the effect of each on RS decoding performance was computed. This computation was carried out for various optical receiver models, and required fairly lengthy numerical analysis to determine accurate word error probabilities when the RS code lengths are long. It is shown that simple threshold decisioning of pulse slots will produce performance that degrades as the background noise increases. This is caused by the generation of too many erasures for the RS decoder to handle. We propose a decision scheme, delta-max demodulation, which offers improvement over threshold decisioning by redefining the generation of an erasure.

I. Introduction

In this paper we study the M-ary optical pulse-position-modulation (PPM) communication system shown in Fig. 1. Source bits are encoded into channel symbols from an M-ary alphabet, which are used to generate a PPM laser pulse sequence. The optical pulse is transmitted to the optical receiver and photodetected. The photodetector produces random count variables for each slot corresponding to a PPM frame. The count variables are converted back to channel symbols for the Reed-Solomon (RS) decoder. The latter provides error correction capability for decoding the source bits.

A question arises as to how the observed photodetected counts should be converted to channel symbols so as to obtain the best RS decoding performance. This report addresses this question.

If no RS encoding is used (the source bits are directly blocked into PPM symbols), maximum likelihood decoding, using the counts as observables, requires a maximum count selection for each PPM frame, with a random choice among any count ties (more than one maximum count). If the resulting error probability is not low enough, coding must be used

to improve performance, with the source bits first encoded into channel words, then the words sent as PPM pulses. When coding is inserted, it is no longer obvious that the maximum likelihood frame decisioning is optimal, since it does not allow for channel symbol erasures. When background noise is negligible, it has been argued (Refs. 1, 2) that matched RS coding appears as a natural encoding scheme, since only channel erasures can occur, and RS decoding has maximal capability for correcting erasures. The RS code size is selected to match the PPM frame size (channel alphabet size) and maximum count demodulation is used, with all ties interpreted as erasures. In the noiseless case, an erasure can occur only if a PPM signaling slot produces no counts.

When background noise is present, conversion of counts to channel symbols will involve errors as well as erasures. The number of erasures that will occur will depend on how the conversion defines an erasure. Since RS decoding can correct more erasures than errors, a question then arises in determining the best way to allocate erasures and errors by proper selection of the conversion rule. In the following sections we examine several conversion algorithms and the resultant performance of each when operating with background noise and RS decoding. This performance will depend on the model of the photodetector used in the optical detection receiver. If a high gain, ideal photomultiplier tube is assumed for the photodetection, the count variables are Poisson distributed with mean values dependent on the received field during that slot. If a high-gain random photodetector is assumed, the counts are more nearly discrete-Gaussian distributed, centered around the mean multiplied count, with a variance dependent on the detector excess noise factor. In each case the postdetection thermal noise can be neglected.

II. Count-Symbol Conversion Rules

In this study we consider two different methods for converting the observed photodetected counts to channel symbols and erasures. The methods differ primarily in the way a symbol decision is made and the way in which an erasure is defined. The methods are labeled as threshold demodulation and delta-max (δ-max) demodulation. In threshold demodulation a threshold γ is set, and any count above γ is called a pulse and a count below γ is called a zero. A symbol decision is made *only if* a single pulse occurs in a PPM frame, selecting the symbol corresponding to the pulse location. All other situations are defined as an erasure. This sequence of frame decisions is then fed into the RS decoder. In δ-max demodulation, a symbol is selected only if no other count is within δ of the maximum count. Otherwise an erasure is declared. Note that both these methods have the advantage that the number of erasures can be controlled by adjustment of the parameters γ and δ.

III. Poisson Counting, δ-Max Demodulation

Consider a Poisson count model and δ-max demodulation for generating the RS symbols. In optical PPM communication, every $\log_2 M$ binary data bits are transmitted by placing an optical light pulse in one of the M designated pulse slots. M slots constitute a PPM frame (Ref. 3). Thus each pulse represents a symbol, depending on its pulse slot location. These $\log_2 M$ binary bits therefore correspond to a Reed-Solomon (RS) symbol. At the PPM optical receiver, a photodetector counts the number of photons in each slot. Let the M photon counts $\{n_i\}_{i=1}^{M}$ correspond to the M time slots. Let $\mathbf{n}$ be a vector with dimension M with elements n_i. Then the probability of receiving $\mathbf{n}$ given a pulse is sent in jth time slot is (note n_i's are independent Poisson distributed random variables)

$$p(\mathbf{n}|s_j) = \frac{(K_s + K_b)^{n_j}}{n_j!} e^{-(K_s + K_b)} \prod_{\substack{i=1 \\ i \neq j}}^{M} \frac{(K_b)^{n_i}}{n_i!} e^{-K_b} \quad (1)$$

where K_s is the average number of received photons per PPM frame and K_b is the average number of background noise photons per slot. We notice that the expected number of photons we receive in the signal slot is $K_s + K_b$ and the expected number of photons in other slots each is K_b. We set a level $\Delta \geqslant 1$ and we make a tentative decision for signal send in the jth slot if for some j

$$\frac{p(\mathbf{n}|s_j)}{p(\mathbf{n}|s_i)} > \Delta \quad \forall \, i \neq j \quad (2)$$

and make no tentative decision (erasure) otherwise. Equation (2) is equivalent to

$$\ln p(\mathbf{n}|s_j) > \ln\Delta + \ln p(\mathbf{n}|s_i) \quad \forall \, i \neq j \quad (3)$$

Redefine

$$\Delta = \left(\frac{K_s + K_b}{K_b}\right)^{\delta} \quad (4)$$

for some $\delta \geqslant 0$. Then using (1) in (3) we get equivalently

$$n_j > \delta + n_i \quad \forall \, i \neq j \quad (5)$$

Hence the maximum count test in (5) is equivalent to testing if the likelihood ratio in (2) is suitably large. The corresponding demodulator structure is shown in Fig. 2, with decision rule given in (5).

We wish to find expressions for the probability of correct detection of transmitted signal P_C, the probability of incorrect detection of transmitted signal P_S, and the probability of no tentative decision (erasure) P_E. We will correctly detect the true signal slot j corresponding to transmitted laser pulse s_j, if (5) is true. The probability of this occurring is

$$P_C = Pr \{n_j > n_1 + \delta, n_j > n_2 + \delta, \ldots, n_j > n_{j-1} + \delta,$$

$$n_j > n_{j+1} + \delta, \ldots, |s_j\}$$

$$= \sum_{k=\delta+1}^{\infty} Pr \{n_i < k - \delta \ \forall \ i \neq j | s_j\} \ Pr \{n_j = k | s_j\}$$

$$= \sum_{k=\delta+1}^{\infty} \left[\sum_{i=0}^{k-\delta-1} Pos(i, K_b) \right]^{M-1} Pos(k, K_s + K_b) \quad (6)$$

where

$$Pos(i, \lambda) = \frac{\lambda^i}{i!} e^{-\lambda} \quad (7)$$

By change of variable we get

$$P_C = \sum_{k=0}^{\infty} \left[\sum_{i=0}^{k} Pos(i, K_b) \right]^{M-1} Pos(k + \delta + 1, K_s + K_b) \quad (8)$$

On the other hand we make an incorrect decision if for a given transmitted pulse in jth time slot, for any $i \neq j$, we have

$$n_i > n_m + \delta \ \forall \ m \neq i \quad (9)$$

Then

$$P_S = Pr \{n_i > n_m + \delta \ \forall \ m \neq i, \text{any } i \neq j | s_j\}$$

$$= (M-1) \sum_{k=\delta+1}^{\infty} Pr \{n_m < k - \delta \ \forall \ m \neq i | s_j\} Pr \{n_i = k | s_j\}$$

$$= (M-1) \sum_{k=\delta+1}^{\infty} \left[\sum_{i=0}^{k-\delta-1} Pos(i, K_b) \right]^{M-2}$$

$$\cdot \left[\sum_{i=0}^{k-\delta-1} Pos(i, K_s + K_b) \right] \cdot Pos(k, K_b) \quad (10)$$

By change of variable we get

$$P_S = (M-1) \sum_{k=0}^{\infty} \left[\sum_{i=0}^{k} Pos(i, K_b) \right]^{M-2}$$

$$\left[\sum_{i=0}^{k} Pos(i, K_s + K_b) \right] \cdot Pos(k + \delta + 1, K_b) \quad (11)$$

Clearly the probability of no tentative decision (probability of erasure) is

$$P_E = 1 - P_C - P_S \quad (12)$$

A Reed-Solomon code of code block $N = M - 1$ and information block K can produce a correct code word if s the number of decoder input symbol errors and e the number of decoder input symbol erasures satisfy the following relation

$$2s + e < N - K + 1 \quad (13)$$

From this relation we note that the RS code can correct twice the number of erasures than the number of symbol errors. It is for this reason that we have tried to introduce some soft decisions at the demodulator in order to produce more erasures. Of course, if we expand the region of no hard decisions in the decision region by too large an amount, the number of erasures will increase in a given block code, and the RS decoder will not be able to correct them.

For the RS code three events may occur. The first event occurs if the number of error and erasure symbols satisfies (13), for which the decoder can correctly decode the code word, and therefore the information block. The second event occurs when (13) is not satisfied, and the combination of symbol errors and symbol erasures is such that the received code block resembles a code signal other than the transmitted one (i.e., the received code block is closer to some other code signal than the transmitted code signal).

In this second event the decoder errs, and gives an incorrect decoded code word. The third event is a complement of the two above events. In this third event, the decoder fails to decode and produces the undecoded channel symbols and randomly decides on erasures. For large M the probability of the second event, for the practical range of interest is usually very small and can be ignored. The probability that the incorrect code word is selected by the decoder, $P_w(RS)$, is (Ref. 4)

$$P_w(RS) = \sum_{s=0}^{N} \sum_{e=q}^{N-s} \binom{N}{s}\binom{N-s}{e} P_S^s P_E^e P_C^{N-s-e} \quad (14)$$

where

$$q = max \ (N - K + 1 - 2s, 0) \quad (15)$$

and the bit error probability $P_b(RS)$ is

$$P_b(RS) = \frac{M}{2(M-1)} \sum_{s=0}^{N} \sum_{e=q}^{N-s} \binom{N}{s}\binom{N-s}{e}\left(\frac{s+e}{N}\right)$$

$$\cdot P_S^s P_E^e P_C^{N-s-e} \quad (16)$$

Equation (16) has been numerically evaluated for the Poisson channel. We considered three classes of RS codes: the (255,127) code with code rate 1/2, the (255,191) code with code rate 3/4, and the (255,223) code with code rate 7/8. These codes are matched to a PPM frame with $M = 256$ slots. For each case we plotted $P_b(RS)$ in (16) versus K_s for various K_b and several values of δ. The results are shown in Figs. 3 through 8. We see that the performance degrades as the noise count K_b increases and as the correction capability of the RS code decreases. In addition, performance is uniformly improved as δ is decreased, with best performance occurring at $\delta = 0$. This corresponds to a maximum likelihood decision on each PPM frame with all maximum ties denoted as erasures. In other words, there appears to be no advantage in widening the erasure definition for these parameter values.

IV. Poisson Counting, Threshold Demodulation

PPM threshold demodulation with Reed-Solomon decoding has been studied for the case of extremely low background noise[1] and thermal noise (Ref. 4). Here we examine the high-gain photodetector case so that the Poisson Counting Process is a valid model. In threshold demodulation, we set a threshold γ and count the number of received photons in each slot. We then compare each number with γ: if it exceeds γ, we claim signal detection in that time slot. If it does not, we claim noise detection in that time slot. We can detect the transmitted signal correctly only if in one of the slots the number of photons exceeds γ, while in all other slots it does not. Then if P_{ds} denotes the probability of signal detection in a time slot, and P_{dn} denotes the probability of correct detection of noise

---

[1]Only dark current was assumed in Ref. 4 and can be treated as extremely low background noise.

in a time slot, the probability of correct PPM signal detection is

$$P_C = P_{ds} P_{dn}^{M-1} \quad (17)$$

The probability of incorrect detection is

$$P_S = (M - 1)(1 - P_{ds})(1 - P_{dn}) P_{dn}^{M-2} \quad (18)$$

and the probability of erasure is

$$P_E = 1 - P_C - P_S \quad (19)$$

For the Poisson channel

$$P_{ds} = \sum_{k=\gamma+1}^{\infty} \frac{(K_s + K_b)^k}{k!} e^{-(K_s + K_b)} \quad (20)$$

$$P_{dn} = \sum_{k=0}^{\gamma} \frac{(K_b)^k}{k!} e^{-K_b} \quad (21)$$

Equations (17)–(21) can again be used in (16) to evaluate performance. The numerical computation has been carried out for i. ⁻me code and count parameters as in the previous section, and the results superimposed in Figs. 3 to 8. The thresholds were set at $\gamma = 1$ and 2 counts, while $\gamma = 0$ corresponds to no threshold (any observed count was considered a pulse). We see that performance with threshold demodulation also degrades with noise count and decreasing code capability, but is much more sensitive to noise levels. In particular we note a severe degradation when no threshold is used and the noise increases from 10^{-4} to 10^{-3} counts. Note that in all cases the δ-max procedure, with $\delta \rightarrow 0$, is uniformly better than the threshold tests, although the two perform similarly if the noise count is low enough. Also note that in Figs. 3, 5 and 7 the optimum threshold γ changes with K_s.

V. Gaussian Counting, δ-Max Demodulation

When nonideal photodetectors are introduced, the count statistics no longer are Poisson. Although primary photoelectrons released from photoemissive surfaces are usually governed by Poisson statistics, secondary electrons generated via multianode secondary emissions, as in photomultiplier vacuum tubes or by avalanche photodetectors (APD), generally produce more symmetrical distributions. The later distributions can often be modeled by Gaussian-shaped distributions (Refs. 5, 6).

Let the PPM slot integrations generate the sequence of secondary count variables n_i, $i = 1, \ldots, M$, where the mean and variance of n_i's are as follows:

In signal slot j: $\quad E\{n_j\} \overset{\Delta}{=} m_1 = G(K_s + K_b)$

$$\sigma_{n_j}^2 \overset{\Delta}{=} \sigma_1^2 = G^2 F(K_s + K_b) \qquad (22)$$

In noise slot i: $\quad E\{n_i\} \overset{\Delta}{=} m_0 = GK_b$

$$\sigma_{n_i}^2 \overset{\Delta}{=} \sigma_0^2 = G^2 F K_b \qquad (23)$$

where G is the photomultiplier or APD gain and F denotes its excess noise factor.

Let n be a vector with dimension M with elements n_i. Then the probability of receiving n given that a pulse is sent in the jth time slot is

$$p(n|s_j) = \frac{c_1}{\sqrt{2\pi\sigma_1^2}} e^{\frac{-(n_j - m_1)^2}{2\sigma_1^2}} \prod_{\substack{k=1 \\ k \neq j}}^{M} \frac{c_0}{\sqrt{2\pi\sigma_0^2}} e^{\frac{-(n_k - m_0)^2}{2\sigma_0^2}} \qquad (24)$$

where c_0 and c_1 are normalization factors.

We again set a level $\Delta > 1$ and we make a tentative decision for signal sent in the jth slot if for some j

$$\frac{p(n|s_j)}{p(n|s_i)} > \Delta \quad \forall\, i \neq j \qquad (25)$$

and make no tentative decision (erasure) otherwise. Equivalently,

$$\ln p(n|s_j) > \ln\Delta + \ln p(n|s_i) \quad \forall\, i \neq j \qquad (26)$$

Redefine

$$\Delta = \exp\left[\delta(\sigma_1^2 - \sigma_0^2)/2\sigma_0^2\sigma_1^2\right] \qquad (27)$$

for some $\delta \geq 0$. Then using (24) in (26) we get equivalently

$$n_j > \sqrt{\delta + n_i^2} \qquad \forall\, i \neq j \qquad (28)$$

The corresponding demodulator structure is as shown in Fig. 2, with decision rule given in (28).

Then

$$P_C = Pr\{n_j > \sqrt{\delta + n_i^2} \quad \forall\, i \neq j | s_j\}$$

$$= \sum_{k > \sqrt{\delta}} Pr\{n_i < \sqrt{k^2 - \delta} \quad \forall\, i \neq j | s_j\} Pr\{n_j = k | s_j\}$$

$$= \sum_{k > \sqrt{\delta}} \left[\sum_{0 < i < \sqrt{k^2 - \delta}} P_0(i)\right]^{M-1} P_1(k) \qquad (29)$$

where

$$P_i(k) = \frac{c_i}{\sqrt{2\pi\sigma_i^2}} e^{\frac{-(k - m_i)^2}{2\sigma_i^2}} \quad ; \quad i = 0, 1 \qquad (30)$$

and c_i is a normalization factor. such that

$$\sum_{k=0}^{\infty} P_i(k) = 1 \qquad (31)$$

Similarly

$$P_S = Pr\{n_i > \sqrt{n_m^2 + \delta} \ \forall\, m \neq i, \text{any } i \neq j | s_j\}$$

$$= (M-1) \sum_{k > \sqrt{\delta}} \left[\sum_{0 < i < \sqrt{k^2 - \delta}} P_0(i)\right]^{M-2}$$

$$\cdot \left[\sum_{0 < i < \sqrt{k^2 - \delta}} P_1(i)\right] P_0(k) \qquad (32)$$

and finally

$$P_E = 1 - P_C - P_S \qquad (33)$$

Since G is very large for numerical computations we can approximate summations in (29) and (32) by integrations. Then we get

$$P_C = \frac{c_0^{M-1} c_1}{\sqrt{2\pi}} \int_{\frac{\sqrt{\delta}-m_1}{\sigma_1}}^{\infty} \left[1 - Q\left(\frac{m_0}{\sigma_0}\right) \right.$$

$$\left. - Q\left(\frac{\sqrt{(\sigma_1 y + m_1)^2 - \delta} - m_0}{\sigma_0}\right) \right]^{M-1} e^{-y^2/2} \, dy \quad (34)$$

and

$$P_S = \frac{(M-1) c_0^{M-1} c_1}{\sqrt{2\pi}} \int_{\frac{\sqrt{\delta}-m_0}{\sigma_0}}^{\infty} \left[1 - Q\left(\frac{m_0}{\sigma_0}\right) \right.$$

$$- Q\left(\frac{\sqrt{(\sigma_0 z + m_0)^2 - \delta} - m_0}{\sigma_0}\right) \right]^{M-2} \left[1 - Q\left(\frac{m_1}{\sigma_1}\right) \right.$$

$$\left. - Q\left(\frac{\sqrt{(\sigma_0 z + m_0)^2 - \delta} - m_1}{\sigma_1}\right) \right] \cdot e^{-z^2/2} \, dz \quad (35)$$

where

$$Q(x) \overset{\Delta}{=} \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} \, dt$$

$$= 0.5 \, \mathrm{Erfc}(x/\sqrt{2}) \quad (36)$$

Equation (16), with (33), (34), and (35) inserted, gives the performance for the photomultiplier or APD case. Results of the computation are shown in Figs. 9 through 11 for δ-max demodulation. Each δ-m.: : curve has been optimized at each value of K_s by adjusting δ for minimal P_b (RS).

VI. Gaussian Counting, Threshold Demodulation

Here demodulator concept is the same as discussed in Section IV. We can use results of Section IV, but replacing P_{ds} and P_{dn} with

$$P_{ds} = \sum_{k=\gamma+1}^{\infty} \frac{c_1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(k-m_1)^2}{2\sigma_1^2}} \quad (37)$$

$$P_{dn} = \sum_{k=0}^{\gamma} \frac{c_0}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(k-m_0)^2}{2\sigma_0^2}} \quad (38)$$

Again approximating summations by integrations we get

$$P_{ds} = Q\left(\frac{\gamma - m_1}{\sigma_1}\right) c_1 \quad (39)$$

and

$$P_{dn} = \left[1 - Q\left(\frac{m_0}{\sigma_0}\right) - Q\left(\frac{\gamma - m_0}{\sigma_0}\right) \right] c_0 \quad (40)$$

Numerical results using (37) – (40) are included in Figs. 9 through 11. It again follows that uniformly better performance occurs with δ-max demodulation over threshold demodulation.

VII. Conclusion

This paper proposes a delta-max demodulator for Reed-Solomon coded M-ary PPM modulation over an optical communication channel. This delta-max demodulator is compared with the threshold demodulator which is currently in use. Both of these demodulators have identical performance in the absence of background noise. As the intensity of background noise increases, the delta-max demodulator outperforms the threshold demodulator. Also, the higher the code rate, the more advantage the delta-max demodulator has.

References

1. McEliece, R. J., "Practical Codes for Photon Communication," *IEEE Trans. Infor. Theory*, Vol. IT-27, No. 4, pp. 393-398, July 1981.

2. Massey, J. L., "Capacity, Cutoff Rate, and Coding for a Direct Detection Optical Channel," *IEEE Trans. Comm.*, Vol. COM-26, Nov. 1981.

3. Gagliardi, R., and Karp. S., *Optical Communications*, Wiley, 1976.

4. Lesh, J. R., Katz, J., Tan, H. H., and Zwillinger, D., "2.5-Bit/Detected Photon Demonstration Program: Analysis and Phase I Results," *TDA Progress Report 42-66*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 115-132, Dec. 1981.

5. Webb, P. P., McIntyre, R. J., and Conradi, J.," Properties of Avalanche Photodiodes," *RCA Review*, Vol. 35, June 1974, pp. 234-278.

6. Sorensen, N., and Gagliardi, R., "Performance of Optical Receivers with Avalanche Photodetection," *IEEE Trans. Comm.*, Vol. COM-27, Sept. 1979.
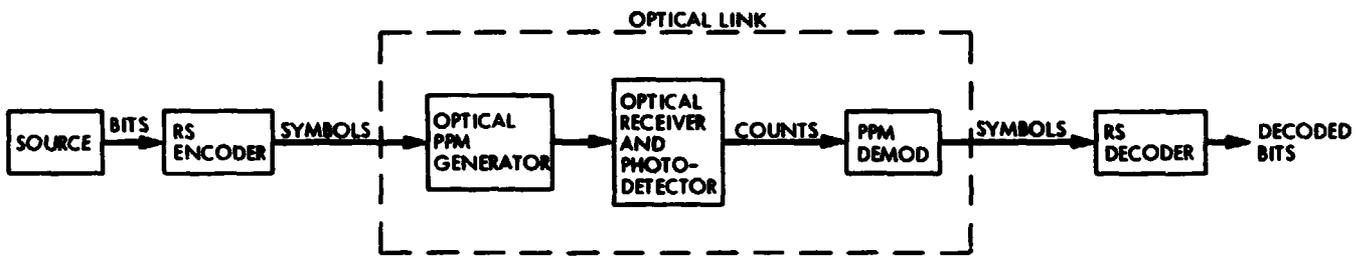
Fig. 1. PPM communication system, block diagram
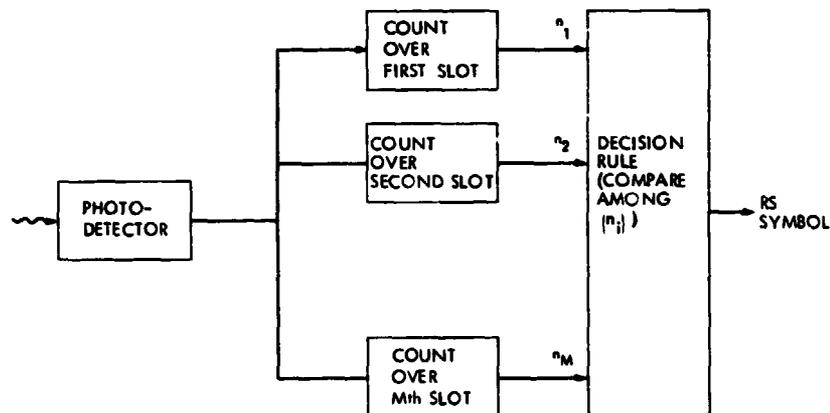
Fig. 2. Demodulator structure
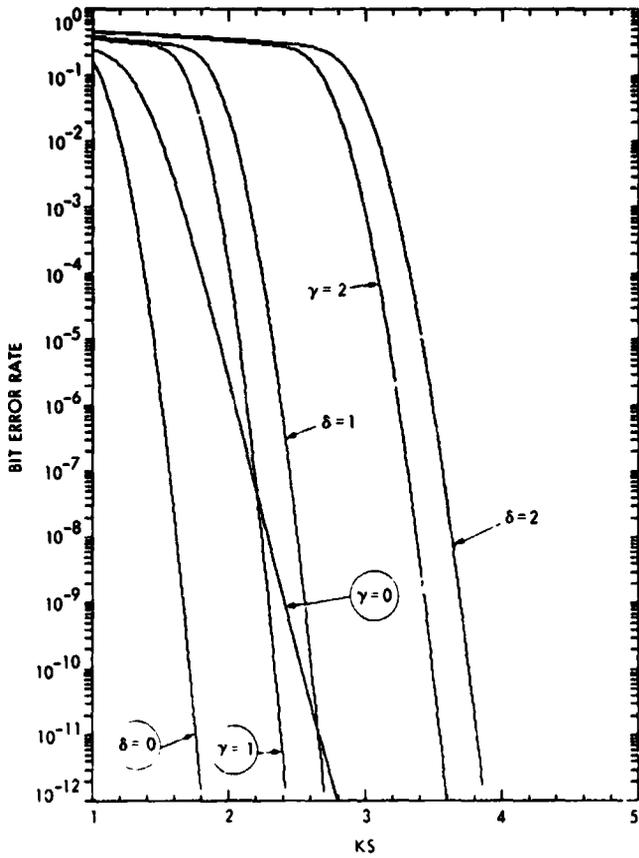
Fig. 3. Bit error rate vs K_s for $K_b = 10^{-3}$ for RS (255, 127)
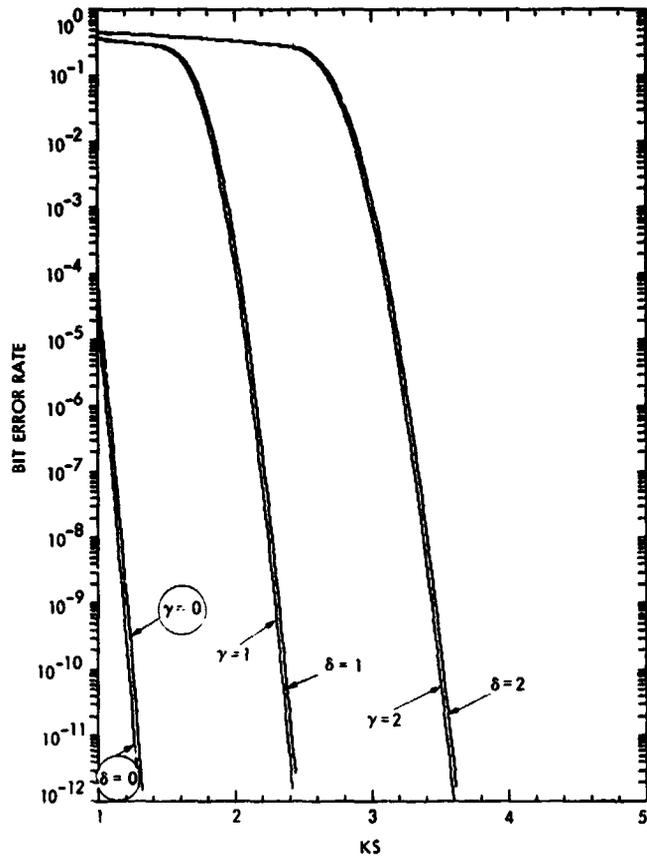
Fig. 4. Bit error rate vs K_s for $K_b = 10^{-4}$ for RS (255, 127)
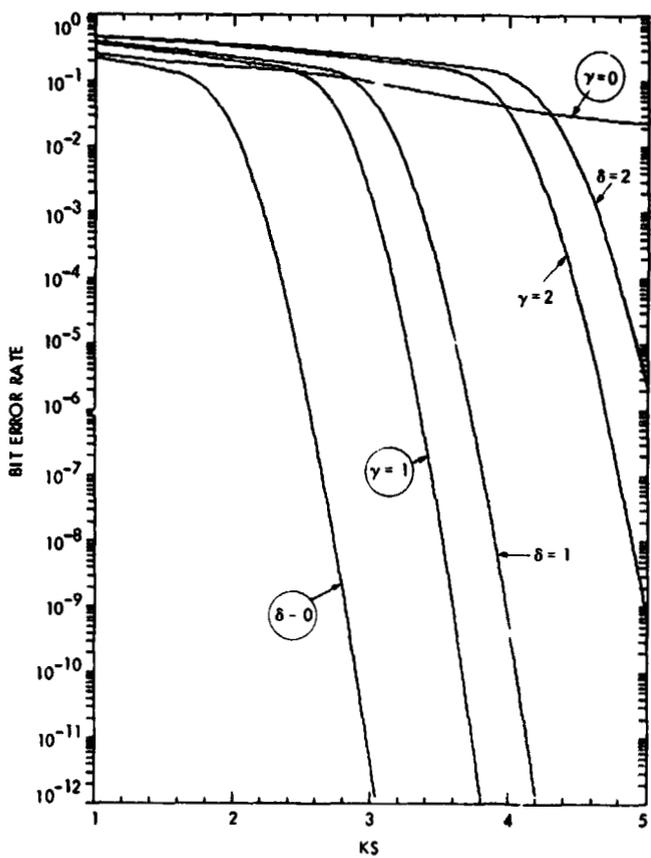
Fig. 5. Bit error rate vs K_s for $K_b = 10^{-3}$ for RS (255, 191)
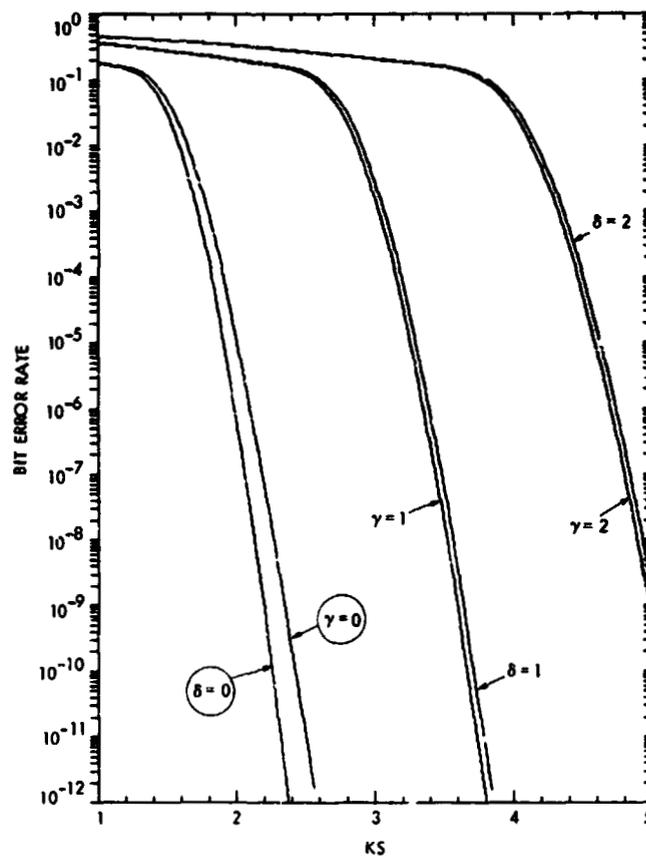
Fig. 6. Bit error rate vs K_s for $K_b = 10^{-4}$ for RS (255, 191)
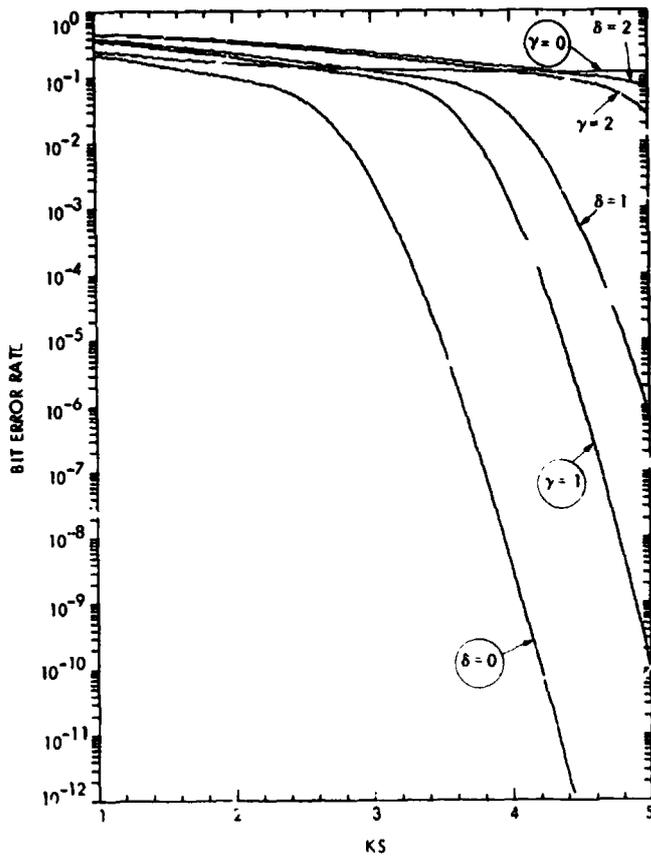
Fig. 7. Bit error rate vs K_s for $K_b = 10^{-3}$ for RS (255, 223)
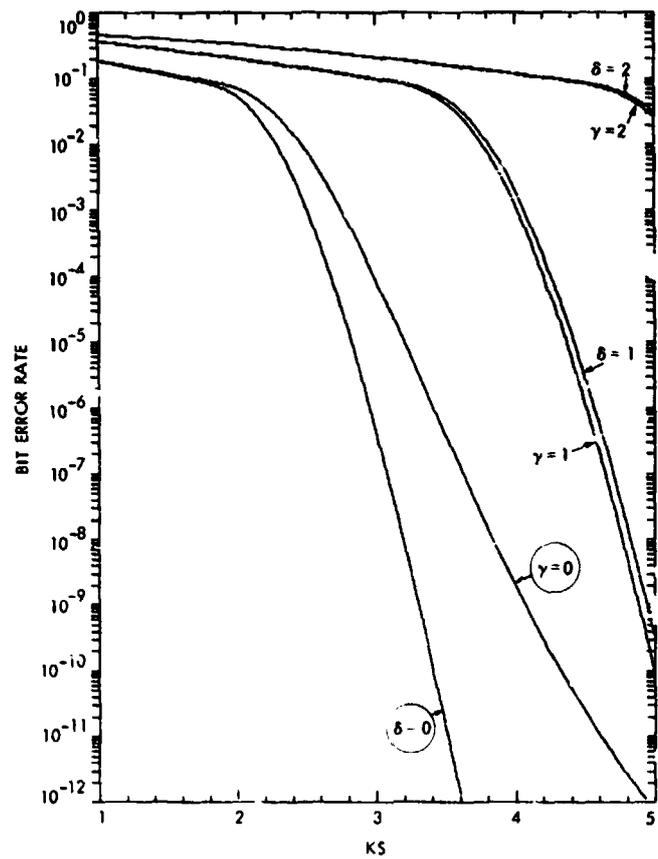

Fig. 8. Bit error rate vs K_s for $K_b = 10^{-4}$ for RS (255, 223)
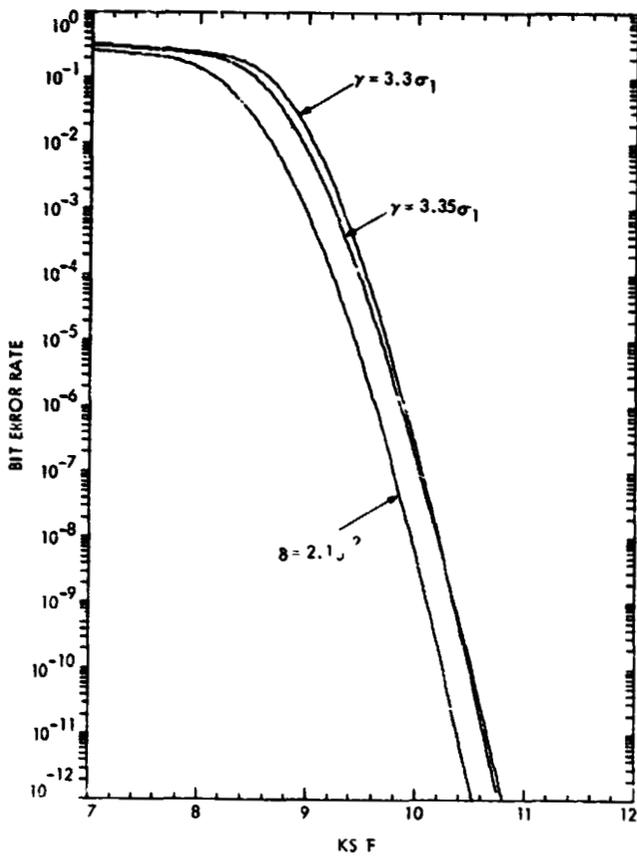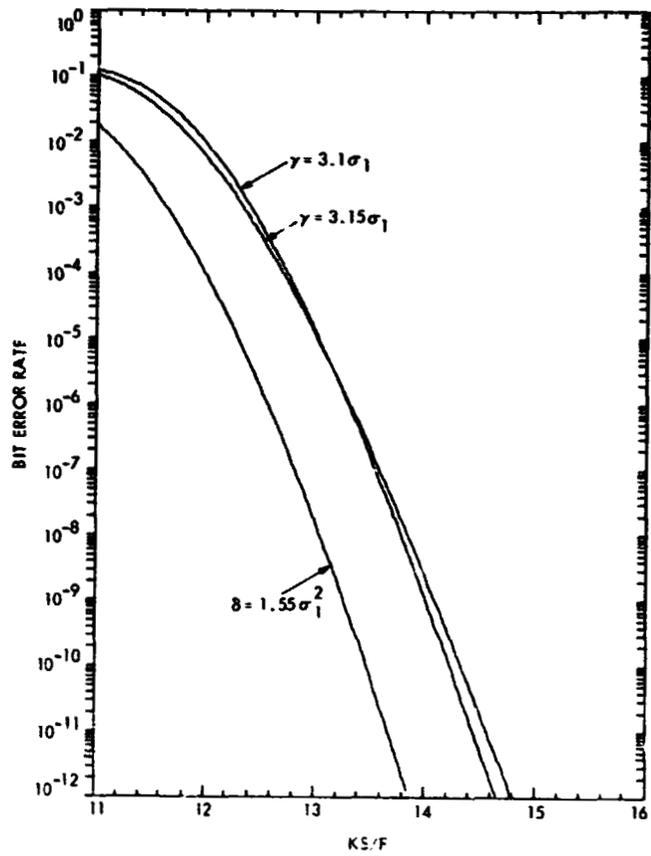
Fig. 9. Bit error rate vs K_s/F for $K_b/F = 5$ for RS (255, 127)

Fig. 10. Bit error rate vs K_s/F for $K_b/F = 5$ for RS (255, 191)
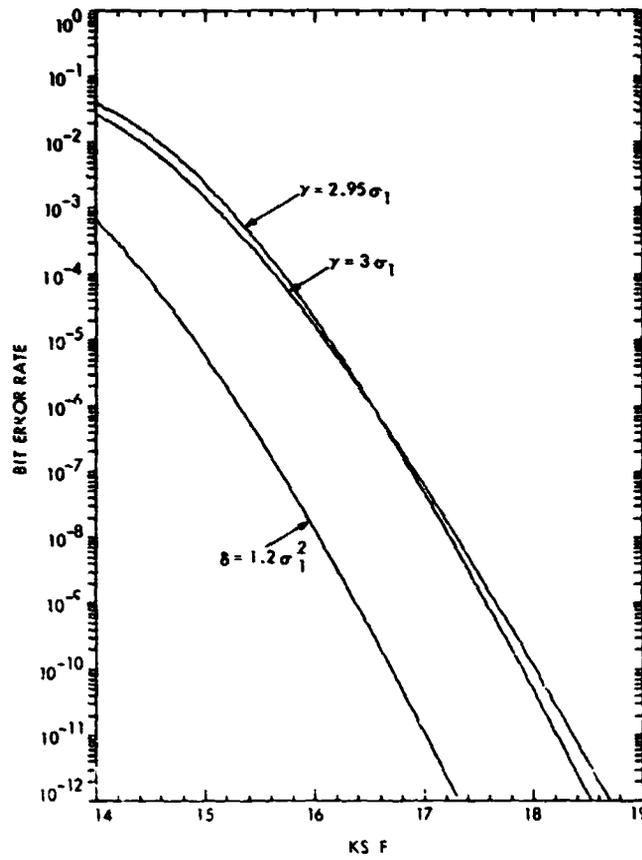
Fig. 11. Bit error rate vs K_s/F for $K_b/F = 5$ for RS (255, 223)

A Parallel VLSI Architecture for a Digital Filter of Arbitrary Length Using Fermat Number Transforms[1]

T. K. Truong
Communications Systems Research Section

I. S. Reed, C. S. Yeh, and H. M. Shao
University of Southern California

In this paper a parallel architecture is developed to compute the linear convolution of two sequences of arbitrary lengths using the Fermat number transform (FNT). In particular a pipeline structure is designed to compute a 128-point FNT. In this FNT, only additions and bit rotations are required. A standard barrel shifter circuit is modified so tha. it performs the required bit rotation operation.

The overlap-save method is generalized for the FNT to compute a linear convolution of arbitrary length. A parallel :-.nitecture is developed to realize this type of overlap-save method using one FNT and several inverse FNTs of 128 points. The generalized overlap-save method alleviates the usual dynamic range limitation in FNTs of l, transform lengths. Its architecture is regular, simple, and expandable, and therefore illy suitable for VLSI implementation.

I. Introduction

Fermat number transforms (FNTs) were developed to compute cyclic convolutions (Refs 1-3). A cyclic convolution of two sequences can be obtained by taking the inverse FNT of the product of the FNTs of these two sequences.

FNTs over certain transform lengths have the advantage over most number-theoretic transforms in that no multiplications are required. McClelland (Ref. 4) designed a hardware system to realize a 64-point 17-bit FNT that used commercially available ECL IC chips. For this purpose he developed a new binary number representation and the binary arithmetic operations modulo a Fermat number (Refs. 4, 5). The Fermat number transform can be applied to digital filtering (Refs. 2, 3), image processing (Refs. 6, 7), X-ray reconstruction (Ref. 8), and to the encoding and decoding of certain Reed-Solomon codes (Refs. 9, 10).

In this paper, a parallel architecture is designed to realize a digital filter of arbitrary length using the FNT. In Section II, a pipeline structure is used to compute a 128-point FNT. Only additions and bit rotations are required in this structure. The bit rotation operations are implemented by a modification of a standard barrel shifter circuit (Ref. 11). In Section III, the overlap-save method is generalized to compute the linear convolution of a digital filtering system. Then a parallel architecture is designed to realize the generalized overlap-save

method using one FNT and several inverse FNTs of 128 points. The circuit design of an FNT butterfly is given in the Appendix.

II. A Parallel Structure for Computing a 128-Point FNT

Let $F_t = 2^{2^t} + 1$ be the tth Fermat number where $t \geqslant 0$. F_t is a prime number for $0 \leqslant t \leqslant 4$. Let $\{x_n\}$ be a N-point sequence of integer numbers, where $0 \leqslant x_n \leqslant F_t - 1, 0 \leqslant n \leqslant N - 1$, and N is a power of 2. The Fermat number transform $\{X_k\}$ of $\{x_n\}$ over F_t is defined as follows:

$$X_k \equiv \sum_{n=0}^{N-1} x_n \, \alpha^{nk} \pmod{F_t} \, , \quad k = 0, 1, \ldots, N-1 \quad (1)$$

where $0 \leqslant X_k \leqslant F_t - 1$ and α is an Nth root of unity. That is, N is the least positive integer such that $\alpha^N \equiv 1 \pmod{F_t}$. The corresponding inverse FNT is the following:

$$x_n \equiv \left(\frac{1}{N}\right) \sum_{k=0}^{N-1} X_k \, \alpha^{nk} \pmod{F_t} \, , \quad n = 0, 1, \ldots, N-1$$

$$(2)$$

In order that a cyclic convolution can be computed by the FNT pair in Eqs. (1) and (2), N depends on the F_t and α chosen (Refs. 2, 3). More details of an FNT can be found in (Refs. 2 and 3).

In this paper F_t, α, and N are selected specifically to be $F_5 = 2^{32} + 1$, $\sqrt{2}$, and 128 respectively. That is, the data of this FNT are integers between 0 and 2^{32}. Hence 33 bits are required to represent a number. The transform length of this FNT is 128. In an FNT over F_t, the quantity $\sqrt{2}$ represents the integer $2^{2^{t-2}} (2^{2^{t-1}}-1)$ (Refs 2, 3). For $t = 5$, since $2^{32} \equiv -1 \pmod{F_5}$, $\sqrt{2} = 2^{24} - 2^8 = 2^{24} + 2^{40}$. A conservative value of the dynamic range (Ref. 12) is $\sqrt{2^{32}/(2^8)} \cong 2^{12}$. This value is sufficiently large for a number of applications.

Since the FNT has a mathematical algorithm similar to the FFT, an FFT-type structure can be applied to perform a fast FNT. Figure 1 shows a pipeline structure (Ref. 13) for computing a 128-point FNT over F_5. The radix-2 decimation-in-time (DIT) technique is used in this structure. The structure for performing an inverse FNT is the mirror image of the circuit shown in Fig. 1 if the radix-2 decimation-in-frequency (DIF) technique is used.

In Fig. 1 z^{-j} denotes a j-step delay element, which can be realized by a set of j first-in-first-out (FIFO) registers. The

symbolic diagram and operations of a DIT FNT butterfly are shown in Fig. 2. The design of a DIT FNT butterfly is given in the Appendix. A similar DIF FNT butterfly was designed in Ref. 4.

In Fig. 1, SW_i is a shuffle-exchange switch controlled by the control signal S_i for $1 \leqslant i \leqslant 6$. The operations of the SW_i are shown in Fig. 3. The S_i's can be implemented simply by a 6-stage up-counter if no buffer registers are used in the FNT butterflies (Ref. 13). With the buffer registers in the butterflies, delay elements are needed at the outputs of the counter, as shown in Fig. 4, for the purpose of synchronization.

In the next section the overlap-save method (Ref. 13) is generalized to implement a digital filter of arbitrary length using one FNT and several inverse FNTs of 128 points over F_5. Then a parallel VLSI architecture is designed to realize this overlap-save method using the FNT structure designed above.

III. A Digital Filter Architecture of Arbitrary Length Using the FNT

In the previous section F_t, α, and N are chosen to be F_5, $\sqrt{2}$, and 128 respectively. $N = 128$ is the maximum transform length over F_5 (Refs. 2, 3), and 2^{12} is the dynamic range. One could increase the transform length by choosing F_t for $t \geqslant 6$. In so doing, however, at least $2^6 + 1 = 65$ bits are required to represent a number. Alternatively, one could use a specific α, where α is not a power of $\sqrt{2}$, over F_3 or F_4 to increase the transform length. In such a case a complete multiplication is required. In addition, the dynamic range is used up readily. To remedy this difficulty, the overlap-save method is generalized to compute the linear convolution of a digital filter of arbitrary input data and filter lengths. A parallel architecture is developed to realize this generalized overlap-save method using the 128-point FNT structure designed in the previous section.

Let $\{x_n\}$ and $\{h_m\}$ be the input and filter sequences of a digital filter, respectively, where $0 \leqslant n \leqslant N - 1$ and $0 \leqslant m \leqslant M - 1$. The output sequence $\{y_k\}$ of the filter is the linear convolution of $\{x_n\}$ and $\{h_m\}$, where $0 \leqslant k \leqslant N + M - 1$ (Ref. 13). It is shown (Ref. 13) that such a linear convolution can be obtained by computing a cyclic convolution. For purposes of exposition it is assumed that $N = 1024$ and $M = 256$ in the following argument.

In order to use 128-point FNTs to compute $\{y_k\}$, four 128-point subfilters $\{h_m^1\}$, $\{h_m^2\}$, $\{h_m^3\}$ and $\{h_m^4\}$ are formed by partitioning $\{h_m\}$ as follows:

$$h_m^i = \begin{cases} h_{m+64(i-1)} & \text{for } 0 \leqslant m \leqslant 63 \\ 0 & \text{for } 64 \leqslant m \leqslant 127 \end{cases} \quad (3)$$

for $1 \leq i \leq 4$. Next the overlap-save method (Ref. 13) is used to compute the linear convolution $\{y_k^i\}$ of $\{x_n\}$ and $\{h_m^i\}$ by using the cyclic convolution technique, where $1 \leq i \leq 4$ and $0 \leq k \leq 1087$. To accomplish this $\{x_n\}$ is sectioned into 128-point subsequences with 64 points of $\{x_n\}$ overlapped between two consecutive subsequences. That is $\{x_n\}$ is sectioned into $\{x_m^1\} = \{x_m\}$, $\{x_m^2\} = \{x_{m+64}\}$, ..., $\{x_m^{15}\} = \{x_{m+896}\}$, where $0 \leq n \leq 1023$ and $0 \leq m \leq 127$. Then $\{y_k^i\}$, for $1 \leq i \leq 4$, is computed by overlapping the cyclic convolution of $\{h_m^i\}$ and $\{x_m^j\}$ for $1 \leq j \leq 15$ using 128-point FNTS. Finally the output sequence $\{y_k\}$, for $0 \leq k \leq 1024 + 256 - 1 = 1279$, results evidently from $\{y_k^i\}$ for $1 \leq i \leq 4$ by the following equation:

$$ y_k = y_k^1 + y_k^2 z^{-64} + y_k^3 z^{-128} + y_k^4 z^{-192} $$

$$ = \left(y_k^1 + y_k^2 z^{-64} \right) + \left(y_k^3 + y_k^4 z^{-64} \right) z^{-128} \qquad (4) $$

The relationship between $\{y_k\}$ and $\{y_k^i\}$ for $1 \leq i \leq 4$ is illustrated in Fig. 5. Other cases of the generalized overlap-save method are constructed in a similar manner.

In Fig. 6 is shown the block diagram of an architecture for the generalized overlap-save method of a digital filter using one FNT and four inverse FNTs of 128 points. In this system the DIT and DIF techniques are used for the FNT and inverse FNTs, respectively. In the generalized overlap-save method, one of the two outputs of the inverse FNT butterfly in the last stage is not needed. Hence, the inverse FNT butterfly in the last stage is a degenerative butterfly circuit, and the delay elements associated with this butterfly circuit are not needed. The H_k^i's in Fig. 6 are the FNTs of $\{h_m^i\}$. The $(1/N)$ factor in Eq. (2) is incorporated into the H_k^i's. These H_k^i's can be precomputed and stored in the system. The adders in Fig. 6 perform normal binary additions, not additions modulo F_t.

The advantage of the generalized overlap-save method for implementing a digital filter using FNT transforms are the following: (1) It requires no multiplications. Only additions a. d bit rotations are needed. (2) It alleviates the usual dynamic range limitation for long sequence FNTs. (3) It utilizes the FNT and inverse FNT circuits 100% of the time. (4) The lengths of the input data and filter sequences can be arbitrary and different.

IV. Conclusion

A pipeline structure is developed to compute a 128-point Fermat number transform. In this 128-point FNT, only additions and bit rotations are required. A barrel shifter circuit is modified to perform the multiplication of an integer by a power of 2 modulo a Fermat number. The overlap-save method is generalized to compute the linear convolution of a digital filter with arbitrary input data and filter lengths. An architecture is developed to realize this generalized overlap-save method by a simple combination of one 128-point FNT and several inverse FNT structures. This realization alleviates the dynamic range limitations of the FNT with a long transform length. The architecture is simple and regular, and hence suitable for VLSI implementation.

References

1. Rader, C.M., "Discrete Convolutions Via Mersenne Transforms," *IEEE Trans. Computers*, Vol. C-21, No. 12, pp. 1269-1273, Dec. 1972.

2. Agarwal, R. C., and Burrus, C. S., "Fast Convolution Using Fermat Number Transforms with Applications to Digital Filtering," *IEEE Trans. Acoustic, Speed, and Signal Processing*, Vol. ASSP-22, No. 2, pp. 87-97, April 1974.

3. Agarwal, R. C., and Burrus, C. S., "Number Theoretical Transforms to Implement Fast Digital Convolution," *Proc. IEEE*, Vol. 63, No. 4, pp. 550-560, April 1975.

4. McClellan, J. H., "Hardware Realization of A Fermat Number Transform," *IEEE Trans. Acoustic, Speed, and Signal Processing*, Vol. ASSP-24, No. 3, pp. 216-225, June 1976.

5. Leibowitz, L. M., "A Simplified Binary Arithmetic For The Fermat Number Transform," *IEEE Trans. Acoustic, Speed, and Signal Processing*, Vol. ASSP-24, No. 5, pp. 356-359, Oct. 1976.

6. Reed, I. S., Truong, T. K., Kwoh, Y. S., and Hall, E. L., "Image Processing by Transforms Over A Finite Field," *IEEE Trans. Computers*, Vol. C-26, No. 9, pp. 874-881, Sep. 1977.

7. Rader, C. M., "On the Application of the Number Theoretic Methods of High Speed Convolution to Two-Dimensional Filtering," *IEEE Trans. Circuits and Systems*, Vol. CAS-22, No 6, pp. 575, June 1975.

8. Reed, I. S., Kwoh, Y. S., Truong, T. K., and Hall, E. L., "X-Ray Reconstruction by Finite Field Transforms," *IEEE Trans. on Nuclear Science*, Vol. NS-24, No. 1, Feb. 1977.

9. Reed, I. S., Truong, T. K., and Welch, L. R., "The Fast Decoding of Reed-Solomon Codes Using Fermat Number Transforms," *IEEE Trans. Information Theory*, Vol. IT-24, No. 4, pp. 497-499, July 1978.

10. Roots, H. F. A., and Best, M. R., "Concatenated Coding on a Spacecraft-to-Ground Telemetry Channel Performance," Processing ICC 81, Denver, CO. 1981.

11. Mead, C. A., and Conway, L. A., *Introduction to VLSI Systems*, Addison-Welsey, Reading, Mass., 1980.

12. Reed, I. S., and Truong, T. K., "The Use of Finite Field to Compute Convolutions," *IEEE Trans. Information Theory*, Vol. IT-21, No. 2, pp. 208-213, March 1975.

13. Rabiner, L. R., and Gold, B., Theory and Application of Digital Signal Processing, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.
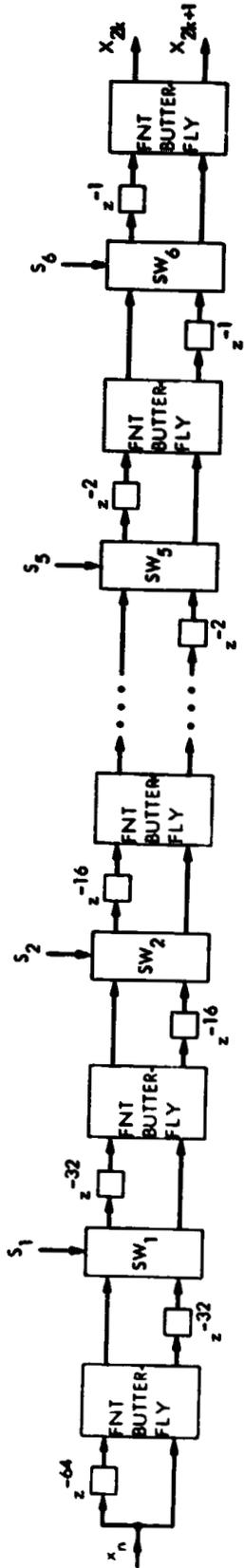
Fig. 1. A pipelined structure for computing a 128-point Fermat number transform
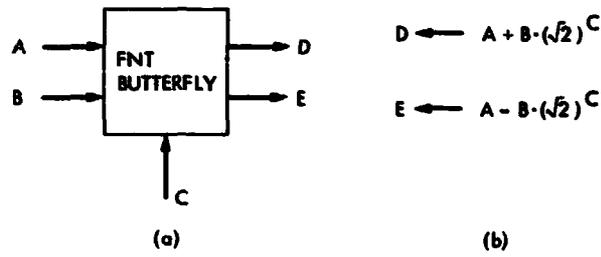
Fig. 2. (a) The symbolic diagram and (b) operations of a DIT FNT butterfly
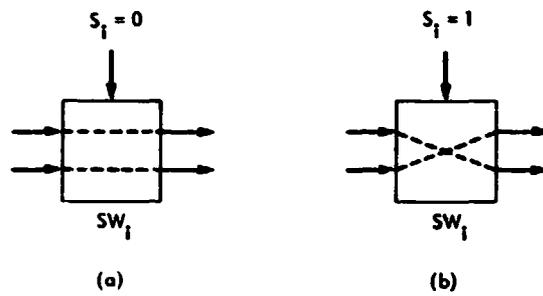
Fig. 3. A shuffle-exchange switch. (a) Direct connection.
(b) Crossed connection
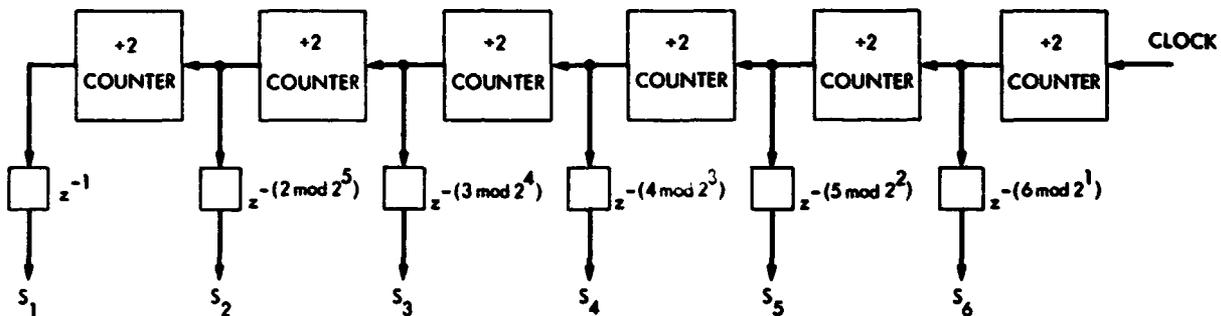
Fig. 4. A 6-stage up-counter used to generate the control signals S_i's in Fig. 1
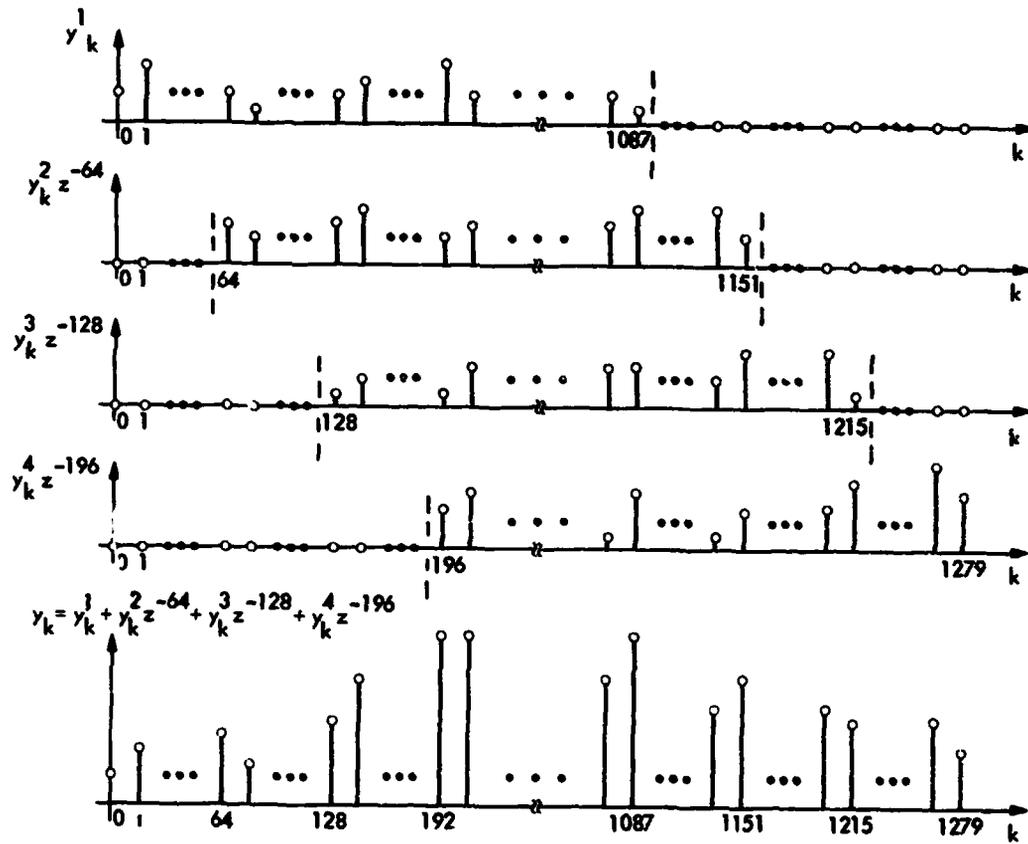
Fig. 5. The example of the generalized overlap-save method in Eq. (4)
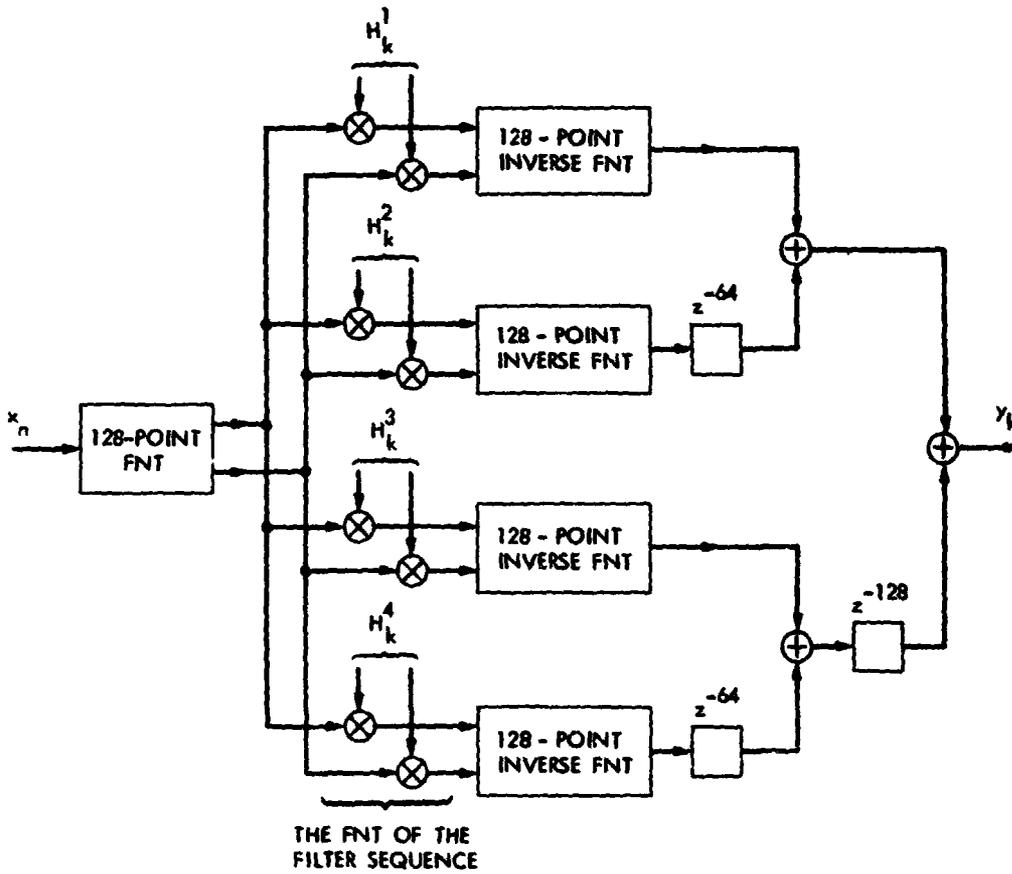
Fig. 6. A realization of a digital filter with a filter sequence of 256 points by using the generalized overlap-save method and the FNT technique

Appendix

In this appendix a circuit is designed to implement a DIT FNT butterfly shown in Fig. 2. A similar DIF FNT butterfly was designed in Ref. 4. To efficiently perform the FNT, number representations have been proposed (Refs. 4, 5) for binary arithmetic operations modulo F_t. The diminished-1 representation proposed by Liebowitz (Ref. 5) is used in the following design. Let A be represented by $[a_{32} \ a_{31} \ldots a_1 \ a_0]$, where $0 \leqslant A \leqslant 2^{32}$ and a_i is the ith bit of A. Table A-1 shows the correspondence between decimal numbers in a normal binary representation and their values in the diminished-1 representation. The most significant bit (MSB) a_{32} can be viewed as the zero-detection bit in the diminished-1 representation.

Two basic binary arithmetic operations modulo F_t with $\alpha = \sqrt{2}$ are addition and multiplication by a power of 2. Other operations can be expressed in terms of these two operations. In the following, some details of these operations are described briefly. More specifics can be found in Ref. 5.

(1) Addition: Let $S = A + B$. If $A = 0$, then $S = B$. If $B = 0$, then $S = A$. If neither A nor B equals 0, add $[a_{31} \ a_{30} \ldots a_1 \ a_0]$ and $[b_{31} \ b_{30} \ldots b_1 \ b_0]$. Then complement the carry and add it to the previous sum. This yields S.

(2) Multiplication by a power of 2: Let $B = A \cdot 2^C$. If $A = 0$, then $B = 0$. If $A \neq 0$, left rotate $[a_{31} \ a_{30} \ldots a_1 \ a_0]$ C bit positions, but complement the value of bit 31 when it is rotated to bit position 0, and set $b_{32} = 0$.

(3) Negation: Since $2^{32} \equiv -1 \pmod{F_5}$, $-A = A \cdot 2^{32}$. Hence if $A \neq 0$, $-A = [a_{32} \ \bar{a}_{31} \ \bar{a}_{30} \ldots \bar{a}_1 \ \bar{a}_0]$ where $\bar{a}_i$ denotes the complement of a_i. If $A = 0$, then $-A = 0$.

(4) Multiplication by $\sqrt{2}$: Since $\sqrt{2} = 2^{24} + 2^{40}$, $A \cdot \sqrt{2} = A \cdot 2^{24} + A \cdot 2^{40}$.

(5) Multiplication by a power of $\sqrt{2}$: Let $B = A \cdot (\sqrt{2})^C$. If C is even, then $B = A \cdot (2)^{C/2}$. If C is odd, then $B = (A \ \sqrt{2}) \cdot 2^{(C-1)/2}$.

In Fig. A-1 is shown a block diagram of an FNT butterfly shown in Fig. 2. In this design, A, B, D, and E are 33-bit data, and C is the 7-bit exponent nk in Eq. (1). Two realizations of an FNT adder can be found in Ref. 4. Figure A-2 shows a pass-transistor full-adder, which requires less silicon area. The multiplier in Fig. A-1 is used to multiply a number by a power of 2 modulo F_5. Figure A-3 shows a block diagram of this multiplier. The shifter in Fig. A-3 is a modification of a barrel shifter (Ref. 11) for performing bit rotation operations.

For purposes of illustration, consider the simple FNT over $F_0 = 2 + 1$. In such an FNT butterfly the functional table and circuit of a modified barrel shifter are shown in Fig. A-4, where the inputs are $[b_1 \ b_0]$ and $[s_3 \ s_2 \ s_1 \ s_0]$, and the outputs are $[b_1^* \ b_0^*]$.

Table A-1. The correspondence among decimal numbers, their values in the normal binary representation, and in the diminished-1 representation

Decimal number	Normal binary representation							Diminished-1 representation						
	a_{32}	a_{31}	a_{30}	$\cdots$	a_2	a_1	a_0	a_{32}	a_{31}	a_{30}	$\cdots$	a_2	a_1	a_0
0	0	0	0	...	0	0	0	1	0	0	...	0	0	0
1	0	0	0	...	0	0	1	0	0	0	...	0	0	0
2	0	0	0	...	0	1	0	0	0	0	...	0	0	1
.				.							.			
.				.							.			
.				.							.			
2^{32}-2	0	1	1	...	1	1	0	0	1	1	...	1	0	1
2^{32}-1	0	1	1	...	1	1	1	0	1	1	...	1	1	0
2^{32}	1	0	0	...	0	0	0	0	1	1	...	1	1	1
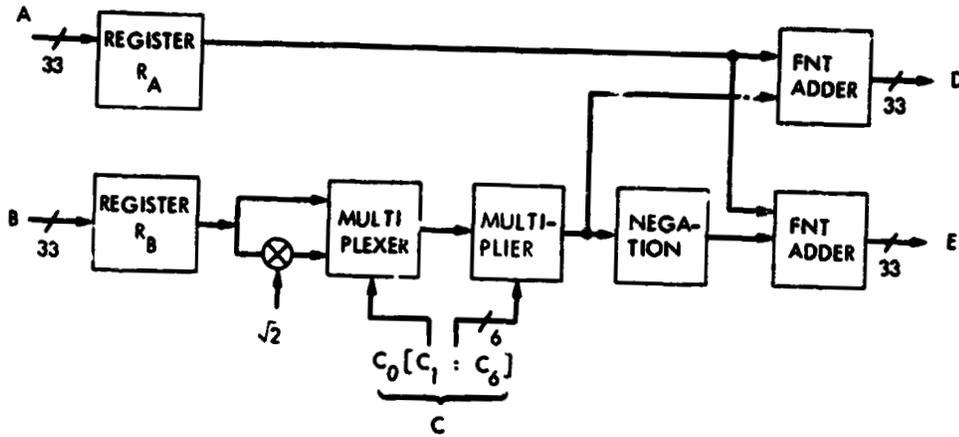
Fig. A-1. A block diagram of a DIT FNT butterfly
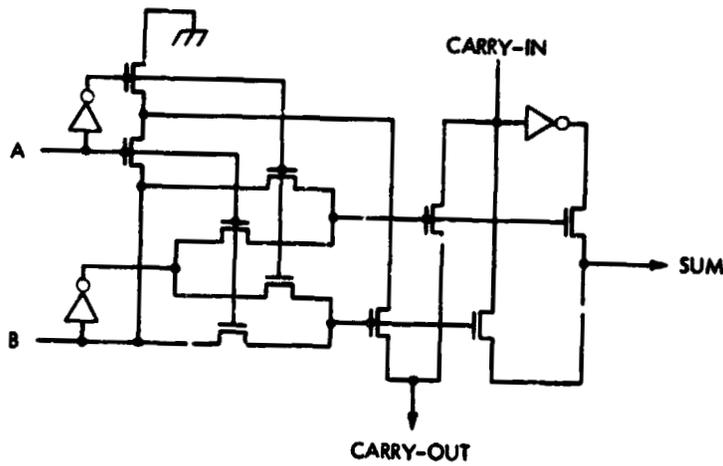
Fig. A-2. A pass-transistor full-adder
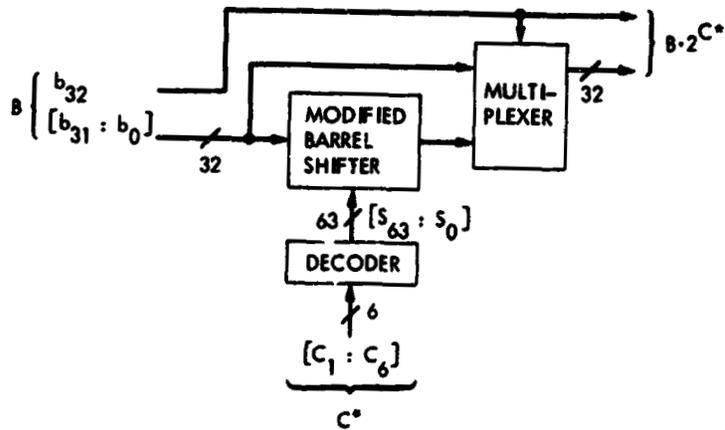
Fig. A-3. A circuit to perform $B \cdot 2^{C^*}$

INPUTS				OUTPUTS	
S_3	S_2	S_1	S_0	b_1^*	b_0^*
0	0	0	1	b_1	b_0
0	0	1	0	b_0	$\overline{b_1}$
0	1	0	0	$\overline{b_1}$	$\overline{b_0}$
1	0	0	0	$\overline{b_0}$	b_1
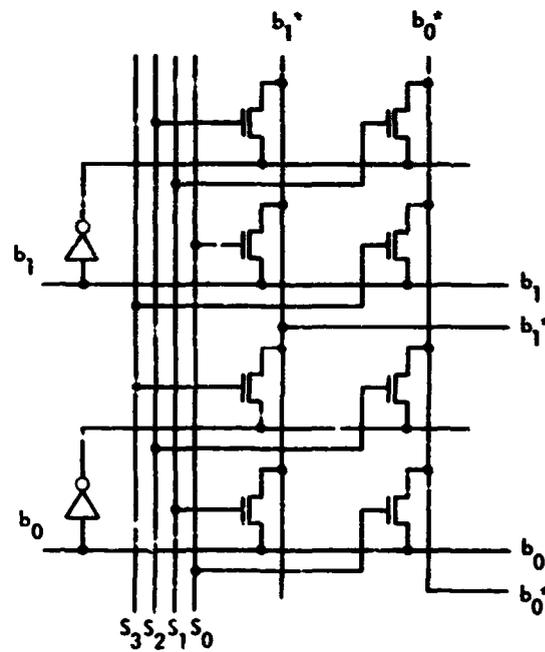
(a)

(b)

Fig. A-4. (a) The functions' table and (b) circuit of a modified barrel shifter

The VLSI Design of a Reed-Solomon Encoder Using Berlekamp's Bit-Serial Multiplier Algorithm

T. K. Truong and L. J. Deutsch
Communications Systems Research Section

I. S. Reed, I.-S. Hsu, K. Wang, and C.-S. Yeh
University of Southern California

E. R. Berlekamp has developed for the Jet Propulsion Laboratory a bit-serial multiplication algorithm for the encoding of Reed-Solomon (RS) codes, using a dual basis over a Galois field. The conventional RS-encoder for long codes often requires look-up tables to perform the multiplication of two field elements. Berlekamp's algorithm requires only shifting and exclusive-OR operations. It is shown in this paper that the new dual-basis (255, 223) RS-encoder can be realized readily on a single VLSI chip with NMOS technology.

I. Introduction

A concatenated Reed-Solomon/Viterbi channel encoding system has been suggested both by the European Space Agency (ESA) (Ref. 1) and JPL (Refs. 2, 3) for the deep-space downlink. The standard RS-encoder design developed by JPL assumes the following codes and parameters.

Let $GF(2^m)$ be a finite field. Then an RS code is a sequence of the symbols in $GF(2^m)$. This sequence of symbols can be considered to be the coefficients of a polynomial. The code polynomial of such a code is

$$C(x) = \sum_{i=0}^{..} c_i x^i \qquad (1)$$

where $c_i \in GF(2^m)$.

The parameters of an RS code are summarized as follows:

m = number of bits per symbol

$n = 2^m - 1$ = the length of a codeword in symbols

t = maximum number of error symbols that can be corrected

$d = 2t + 1$ = design distance

$2t$ = number of check symbols

$k = n - 2t$ = number of information symbols

In the JPL design, $m = 8$, $n = 255$, $t = 16$, $d = 33$, $2t = 32$, and $k = 223$. This code is the (255, 223) RS code.

The generator polynomial of an RS code is defined by

$$g(x) = \sum_{j=b}^{b+2t-1} (x - \gamma^j) = \sum_{i=0}^{2t} g_i x^i \qquad (2)$$

where b is a nonnegative integer, usually chosen to be 1, and γ is a primitive element in $GF(2^m)$. In order to reduce the complexity of the encoder it is desirable to make the coefficients of $g(x)$ symmetric so that $g(x) = x^{-d-1} g(1/x)$. To accomplish this b must be chosen to satisfy $2b + d - 2 = 2^m - 1$. Thus for the JPL code $b = 112$.

Let $I(x) = c_{2t}x^{2t} + c_{2t+1}x^{2t+1} + \cdots + c_{n-1}x^{n-1}$ and $P(x) = c_0 + c_1 x + \cdots + c_{2t-1}x^{2t-1}$ be the information polynomial and the check polynomial, respectively. Then the encoded RS code polynomial is represented by

$$C(x) = I(x) + P(x) \qquad (3)$$

•

To be an RS code $C(x)$ must be also a multiple of $g(x)$. That is,

$$C(x) = q(x)g(x) \qquad (4)$$

To find $P(x)$ in Eq. (3) such that Eq. (4) is true, divide $I(x)$ by $g(x)$. The division algorithm yields

$$I(x) = q(x)g(x) + r(x) \qquad (5)$$

Also let $r(x) = -P(x)$, then by Eq. (5)

$$q(x)g(x) = I(x) - r(x) = I(x) + P(x) = C(x) \qquad (6)$$

Figure 1 shows the structure of a t-error correcting RS encoder over $GF(2^m)$. In Fig. 1 R_i for $0 \le i \le 2t - 1$ and Q are m-bit registers. Initially all registers are set to zero, and both switches (controlled by a control signal SL) are set to position A.

The information symbols $c_{n-1}, \cdots, c_{2t}$ are fed into the division circuit of the encoder and also transmitted out of the encoder one by one. The quotient coefficients are generated and loaded into Q register sequentially. The remainder coefficients are computed successively. Immediately after c_{2t} is fed to the circuit, both switches are set to position B. At the very same moment c_{2t-1} is computed and transmitted. Simultaneously, c_i is being computed and loaded into register R_i, f . $0 \le i \le 2t - 2$. Next $c_{2t-2}, \cdots, c_0$ are transmitted out of the encoder one by one. $c_{2t-2}, \cdots, c_0$ retain their values because the content of Q is set to zero when the upper switch is at position B.

The complexity of the design of an RS encoder results from the computation of products zg_i for $0 \le i \le 2t - 2$. These computations can be performed in several ways (Ref. 3). Unfortunately none of them is suited to the pipeline processing structures usually seen in VLSI design. Recently, Berlekamp (Ref. 4) developed a bit-serial multiplier algorithm that has the features needed to solve this problem. Perlman and Lee (Ref. 5) show in detail the mathematical basis for this algorithm. In this paper Berlekamp's method is applied to the VLSI design of a (255, 22?) RS-encoder, which can be implemented on a single VLSI chip.

II. Berlekamp's Bit-Serial Multiplier Algorithm Over $GF(2^m)$

In order to understand Berlekamp's multiplier algorithm some mathematical preliminaries are needed. Toward this end the mathematical concepts of the trace and a complementary (or dual) basis are introduced. For more details and proofs see Refs. 3, 4 and 5.

Definition 1: The trace of an element β belonging to $GF(p^m)$, the Galois field of P^m elements, is defined as follows:

$$Tr(\beta) = \sum_{k=0}^{m-1} \beta^{p^k}$$

In particular, for $p = 2$,

$$Tr(\beta) = \sum_{k=0}^{m-1} (\beta)^{2^k}$$

The trace has the following properties

(1) $[Tr(\beta)]^p = \beta + \beta^p + \cdots + \beta^{p^{m-1}} = Tr(\beta)$, where $\beta \in GF(p^m)$. This implies that $Tr(\beta) \in GF(p)$; i.e., the trace is on the ground field $GF(p)$

(2) $Tr(\beta + r) = Tr(\beta) + Tr(r)$, where $\beta, r \in GF(p^m)$

(3) $Tr(c\beta) = cTr(\beta)$, where $c \in GF(p)$.

(4) $Tr(1) \equiv m \pmod p$.

Definition 2: A basis $\{\mu_j\}$ in $GF(p^m)$ is a set of m linearly independent elements in $GF(p^m)$.

Definition 3: Two bases $\{\mu_j\}$ and $\{\lambda_k\}$ are said to be complementary or the dual of one another if

$$Tr(\mu_j \lambda_k) = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}$$

The basis $\{\mu_j\}$ is called the original basis, and the basis $\{\lambda_k\}$ is called the dual basis.

Lemma: If α is a root of an irreducible polynomial of degree m in $GF(p^m)$, then $\{\alpha^k\}$ for $0 \leqslant k \leqslant m - 1$ is a basis of $GF(p^m)$. The basis $\{\alpha^k\}$ for $0 \leqslant k \leqslant m - 1$ is called the normal or natural basis of $GF(p^m)$.

Theorem 1 (Theorem 19 in Ref. 4): Every basis has a complementary basis.

Corollary 1: Supposes the bases $\{\mu_j\}$ and $\{\lambda_k\}$ are complementary. Then a field element z can be expressed in the dual basis $\{\lambda_k\}$ by the expansion

$$z = \sum_{k=0}^{m-1} z_k \lambda_k = \sum_{k=0}^{m-1} Tr(z\mu_k)\lambda_k$$

where $z_k = Tr(z\mu_k)$ is the kth coefficient of the dual basis.

Proof: Let $z = z_0\lambda_0 + z_1\lambda_1 + \cdots + z_{m-1}\lambda_{m-1}$. Multiply both sides by α^k and take the trace. Then by Def. 3 and the properties of the trace,

$$Tr(z\alpha^k) = Tr\left(\sum_{i=0}^{m-1} z_i(\lambda_i\mu_k)\right) = z_k \qquad \text{Q.E.D.}$$

The following corollary is an immediate consequence of Corollary 1.

Corollary 2: The product $w = zy$ of two field elements in $GF(p^m)$ can be expressed in the dual basis by the expansion

$$w = \sum_{k=0}^{m-1} Tr(zy\mu_k)\lambda_k$$

where $Tr(zy\mu_k)$ is the kth coefficient of the dual basis for the product of two field elements.

These two corollaries provide a theoretical basis for the new RS-encoder algorithm.

III. A Simple Example of Berlekamp's Algorithm Applied to an RS-Encoder

This section follows the treatment in Ref. 3. It is included here for two purposes. First, Ref. 3 is not readily available for most readers. Second, this example is included to illustrate

how Berlekamp's new bit-serial multiplier algorithm can be used to realize the RS-encoder structure presented in Fig. 1.

Consider a $(15, 11)$ RS code over $GF(2^4)$. For this code, $m = 4$, $n = 15$, $t = 2$, $d = 2t + 1 = 5$, and $n - 2t = 11$ information symbols. Let α be a root of the primitive irreducible polynomial $f(x) = x^4 + x + 1$ over $GF(2)$. α satisfies $\alpha^{15} = 1$. An element z in $GF(2^4)$ is representable by 0 or α^j for some j, $0 \leqslant j \leqslant 14$. z can be represented also by a polynomial in α over $GF(2)$. This is the representation of $GF(2^4)$ in the normal basis $\{\alpha^k\}$ for $0 \leqslant k \leqslant 3$. That is, $z = u_0 + u_1\alpha + u_2\alpha^2 + u_3\alpha^3$, where $u_k \in GF(2)$ for $0 \leqslant k \leqslant 3$.

In Table 1, the first column is the index or logarithm of an element in base α. The logarithm of the zero element is denoted by an asterisk. Columns 2 to 5 show the 4-tuples of the coefficients of the elements expressed as polynomials.

The trace of an element z in $GF(2^4)$ is found by Def. 1 and the properties of the trace to be

$$TR(z) = u_0 Tr(1) + u_1 Tr(\alpha) + u_2 Tr(\alpha^2) + u_3 Tr(\alpha^3)$$

where $Tr(1) \equiv 4 \pmod 2 = 0$, $Tr(\alpha) = Tr(\alpha^2) = \alpha + \alpha^2 + \alpha^4 + \alpha^8 = 0$ and $Tr(\alpha^3) = \alpha^3 + \alpha^6 + \alpha^9 + \alpha^{12} = 1$. Thus $Tr(z) = u_3$. The trace element α^k in $GF(2^4)$ is listed in column 3 of Table 1.

By Def. 2 any set of four linearly independent elements can be used as a basis for the field $GF(2^4)$. To find the dual basis of the normal basis $\{\alpha^j\}$ in $GF(2^4)$ let a field element z be expressed in dual basis $\{\lambda_0, \lambda_1, \lambda_2, \lambda_3\}$. From Corollary 1 the coefficients of z are $z_k = Tr(z\alpha^k)$ for $0 \leqslant k \leqslant 3$. Thus $z_0 = Tr(z)$, $z_1 = Tr(z\alpha)$, $z_2 = Tr(z\alpha^2)$ and $z_3 = Tr(z\alpha^3)$. Let $z = \alpha^i$ for some i, $0 \leqslant i \leqslant 14$. Thus a coefficient z_k, for $0 \leqslant k \leqslant 3$, of an element z in the dual space can be obtained by cyclically shifting the trace column in Table 1 upward by k positions where the first row is excluded. These appropriately shifted columns of coefficients are shown in Table 1 as the last four columns. In Table 1 the elements of the dual basis, $\lambda_0, \lambda_1, \lambda_2, \lambda_3$, are underlined. Evidently $\lambda_0 = \alpha^{14}$, $\lambda_1 = \alpha^2$, $\lambda_2 = \alpha$ and $\lambda_3 = 1$ are the four elements of the dual basis.

In order to make the generator polynomial $g(x)$ symmetric b must satisfy the equation $2b + d - 2 = 2^m - 1$. Thus $b = 6$ for this code. The γ in Eq. (2) can be any primitive element in $GF(2^4)$. It will be shown in Section IV that γ can be chosen to simplify the binary mapping matrix. In this example let $\gamma = \alpha$. Thus the generator polynomial is given by

$$g(x) = \prod_{j=6}^{9} (x - \alpha^j) = \sum_{i=0}^{4} g_i x^i \tag{7}$$

where $g_0 = g_4 = 1, g_1 = g_3 = \alpha^3$ and $g_2 = \alpha$.

Let g_i be expressed in the normal basis $\{1, \alpha, \alpha^2, \alpha^3\}$. Let z, a field element, be expressed in the dual basis; i.e., $z = z_0\lambda_0 + z_1\lambda_1 + z_2\lambda_2 + z_3\lambda_3$. In Fig. 1 the products zg_i for $0 \leq i \leq 3$ needs to be computed.

Since $g_3 = g_1$, it is necessary to compute only zg_0, zg_1 and zg_2. Let the products zg_i for $0 \leq i \leq 2$ be represented in the dual basis. By Corollary 2 zg_i can be expressed in the dual basis as

$$ z \begin{bmatrix} g_0 \\ g_1 \\ g_2 \end{bmatrix} = \sum_{k=0}^{3} \begin{bmatrix} T_0^{(k)}(z) \\ T_1^{(k)}(z) \\ T_2^{(k)}(z) \end{bmatrix} \lambda_k \qquad (8) $$

where $T_i^{(k)}(z) = Tr(zg_i\alpha^k)$ is the kth coefficient (or kth bit) of zg_i for $0 \leq i \leq 2$ and $0 \leq k \leq 3$.

The present problem is to express $T_i^{(k)}(z)$ recursively in terms of $T_i^{(k-1)}(z)$ for $1 \leq k \leq 3$. Initially for $k = 0$,

$$ \begin{bmatrix} T_0^{(0)}(z) \\ T_1^{(0)}(z) \\ T_2^{(0)}(z) \end{bmatrix} = \begin{bmatrix} Tr(zg_0) \\ Tr(zg_1) \\ Tr(zg_2) \end{bmatrix} = \begin{bmatrix} Tr(z\alpha^0) \\ Tr(z\alpha^3) \\ Tr(z\alpha) \end{bmatrix} = \begin{bmatrix} z_0 \\ z_3 \\ z_1 \end{bmatrix} $$

(9)

where $TR(z\alpha^j) = Tr((z_0\lambda_0 + z_1\lambda_1 + z_2\lambda_2 + z_3\lambda_3)\alpha^j) = z_j$ for $0 \leq j \leq 3$. Equation (9) can be expressed in a matrix form as follows:

$$ \begin{bmatrix} T_0^{(0)}(z) \\ T_1^{(0)}(z) \\ T_2^{(0)}(z) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \\ z_2 \\ z_3 \end{bmatrix} \qquad (10) $$

The above matrix is the 3 × 4 binary mapping matrix of the problem.

To compute $T_i^{(k)}(z)$ for $k > 0$, observe that $T_i^{(k)}(z) = Tr((\alpha z)g_i\alpha^{k-1}) = T^{(k-1)}(\alpha z)$. Hence $T_i^{(k)}(z)$ is obtained from $T_i^{(k-1)}(z)$ by replacing z by $y = \alpha z$. Let $\alpha z = y = y_0\lambda_0 + y_1\lambda_1 + y_2\lambda_2 + y_3\lambda_3$, where $y_m = Tr(y\alpha^m) = Tr(z\alpha^{m+1})$ for $0 \leq m \leq 3$. Then $T_i^{(k)}$ is obtained from $T_i^{(k-1)}$ by replacing z_0 by $y_0 = Tr(z\alpha) = z_1$, z_1 by $y_1 = Tr(z\alpha^2) = z_2$, z_2 by $y_2 = Tr(z\alpha^3) = z_3$ and z_3 by $y_3 = Tr(z\alpha^4) = Tr(z(\alpha + 1)) = z_0 + z_1$.

To recapitulate $zg_i = T_i^{(0)}\lambda_0 + T_i^{(1)}\lambda_1 + T_i^{(2)}\lambda_2 + T_i^{(3)}\lambda_3$, where $0 \leq i \leq 3$ and $z = z_0\lambda_0 + z_1\lambda_1 + z_2\lambda_2 + z_3\lambda_3$, can be computed by Berlekamp's bit-serial multiplier algorithm as follows:

(1) Initially for $k = 0$, compute $T_0^{(0)}(z)$, $T_1^{(0)}(z)$ and $T_2^{(0)}(z)$ by Eq. (10). Also $T_3^{(0)}(z) = T_1^{(0)}(z)$.

(2) For $k = 1, 2, 3$, compute $T_i^{(k)}(z)$ by

$$ T_i^{(k)}(z) = T_i^{(k-1)}(y) $$

where $0 \leq i \leq 3$ and $y = \alpha z = y_0\lambda_0 + y_1\lambda_1 + y_2\lambda_2 + y_3\lambda_3$ with $y_0 = z_1$, $y_1 = z_2$, $y_2 = z_3$ and $y_3 = z_0 + z_1 = T_f$, where $T_f = z_0 + z_1$ is the feedback term of the algorithm.

The above example illustrates Berlekamp's bit-serial multiplier algorithm. This algorithm developed in Refs. 4 and 5 requires shifting and XOR operations only. Berlekamp's dual basis RS-encoder is well-suited to a pipeline structure which can be implemented in VLSI design. The same procedure extends similarly to the design of a (255, 223) RS-encoder over $GF(2^8)$.

IV. A VLSI Architecture of a (255, 223) RS-Encoder with Dual-Basis Multiplier

In this section an architecture is designed to implement (255, 223) RS-encoder using Berlekamp's multiplier algorithm. The circuit is a direct mapping from an encoder using Berlekamp's bit-serial algorithm as developed in the previous sections to an architectural design. This architecture can be realized quite readily on a single NMOS VLSI chip.

Let $GF(2^8)$ be generated by α, where α is a root of a primitive irreducible polynomial $f(x) = x^8 + x^7 + x^2 + x + 1$ over $GF(2)$. The normal basis of this field is $\{1, \alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7\}$. The representations of this field in both the normal basis and its dual basis are tabulated in Appendix A. From Corollary 1 the coefficients of a field element α^j can be obtained from $z_k = Tr(\alpha^{j+k})$ for $0 \leq k \leq 7$, where $\alpha^j = z_0\lambda_0 + \cdots + z_7\lambda_7$. From Table A-1 in Appendix A, the dual basis

$\{\lambda_0, \lambda_1, \cdots, \lambda_7\}$ of the normal basis is the ordered set $\{\alpha^{99}, \alpha^{197}, \alpha^{203}, \alpha^{202}, \alpha^{201}, \alpha^{200}, \alpha^{199}, \alpha^{100}\}$.

It was mentioned previously that γ in Eq. (2) can be chosen to simplify the binary mapping matrix. Two binary matrices, one for the primitive element $\gamma = \alpha^{11}$ and the other for $\gamma = \alpha$, were computed. It was found that the binary mapping matrix for $\gamma = \alpha^{11}$ had a smaller number of 1's. Hence this binary mapping matrix was used in the design. For this case the generator polynomial $g(x)$ of the RS-encoder over $GF(2^8)$ was given by

$$g(x) = \prod_{j=112}^{143} (x - \alpha^{11j}) = \sum_{i=0}^{32} g_i x^i \qquad (11)$$

where $g_0 = g_{32} = 1, g_1 = g_{31} = \alpha^{249}, g_2 = g_{30} = \alpha^{59}, g_3 = g_{29} = \alpha^{66}, g_4 = g_{28} = \alpha^4, g_5 = g_{27} = \alpha^{43}, g_6 = g_{26} = \alpha^{126}, g_7 = g_{25} = \alpha^{251}, g_8 = g_{24} = \alpha^{97}, g_9 = g_{23} = \alpha^{30}, g_{10} = g_{22} = \alpha^3, g_{11} = g_{21} = \alpha^{?}, g_{12} = g_{20} = \alpha^{50}, g_{13} = g_{19} = \alpha^{66}, g_{14} = g_{18} = \alpha^{170}, g_{15} = g_{17} = \alpha^5$, and $g_{16} = \alpha^{24}$

The binary mapping matrix for the coefficients of the generator polynomial in Eq. (11) is computed and shown in Appendix B. The feedback term T_f in Berlekamp's algorithm is found in this case to be:

$$T_f = Tr(\alpha^8 z) = Tr((\alpha^7 + \alpha^2 + \alpha + 1)z) = z_0 + z_1 + z_2 + z_7$$

$$(12)$$

In the following a VLSI chip architecture is designed to realize a (255, 223) RS-encoder using the above parameters and Berlekamp's algorithms. An overall block diagram of this chip is shown in Fig. 2. In Fig. 2 VDD and GND are power pins. CLK is a clock signal, which in general is a periodic square wave. The information symbols are fed into the chip from the data-in pin DIN bit-by-bit. Similarly the encoded codeword is transmitted out of the chip from the data-out pin DOUT sequentially. The control signal LM (load mode) is set to 1 (logic 1) when the information symbols are loaded into the chip. Otherwise, LM is set to 0.

The input data and LM signals are synchronized by the CLK signal, while the operations of the circuit and output data signal are synchronized by two nonoverlapping clock signals. To save space, dynamic registers are used in this ... of a 1-bit dynamic register with reset is ... ing diagram of CLK, $\phi1$, $\phi2$, LM, DIN ... in Fig. 4. The delay of DOUT ... the input and output flip-flops.

Figure 5 shows the block diagram of a (255, 223) RS-encoder over $GF(2^8)$ using Berlekamp's bit-serial multiplier algorithm. The circuit is divided into five units. The circuits in each unit are discussed in the following:

(1) Product Unit: The Product Unit is used to compute T_f, $T_{31}, \cdots, T_0$. This circuit is realized by a Programmable Logic Array (PLA) circuit [6]. Since $T_0 = T_{31}$, $T_1 = T_{30}, \cdots, T_{15} = T_{17}$, only $T_f, T_{31}, \cdots, T_{17}$ and T_{16} are actually implemented in the PLA circuit. $T_0, \cdots, T_{15}$ are connected directly to $T_{31}, \cdots, T_{17}$, respectively. Over other circuits a PLA circuit has the advantage of being easy to reconfigure.

(2) Remainder Unit: The Remainder Unit is used to store the coefficients of the remainder during the division process. In Fig. 5, S_i for $0 \leqslant i \leqslant 30$ are 8-bit shift registers with reset. The addition in the circuit is a modulo 2 addition or Exclusive-OR operation. While c_{32} is being fed to the circuit, c_{31} is being computed and transmitted sequentially from the circuit. Simultaneously c_i is computed and then loaded into S_i for $0 \leqslant i \leqslant 30$. Then $c_{30}, \cdots, c_0$ are transmitted out of the encoder bit-by-bit.

(3) Quotient Unit: In Fig. 5, Q and R represent a 7-bit shift register with reset and an 8-bit shift register with reset and parallel load, respectively. R and Q store the currently operating coefficient and the next coefficient of the quotient polynomial, respectively. A logic diagram of register R is shown in Fig. 6. z_i is loaded into R every eight clock cycles, where $0 \leqslant i \leqslant 7$. Immediately after all 223 information symbols are fed into the circuit, the control signal SL changes to logic 0. Thenceforth the contents of Q and R are zero so that the check symbols in the Remainder Unit sustain their values.

(4) I/O Unit: This unit handles the input/output operations. In Fig. 5 both F_0 and F_1 are flip-flops. A pass transistor controlled by $\phi1$ is inserted before F_1 for the purpose of synchronization. Control signal SL selects whether a bit of an information symbol or a check symbol is to be transmitted.

(5) Control Unit: The Control Unit generates the necessary control signals. This unit is further divided into 3 portions, as shown in Fig. 7. The two-phase clock generator circuit in Ref. 6 is used to convert a clock signal into two nonoverlapping clock signals $\phi1$ and $\phi2$. In Fig. 8 is shown a logic diagram of the circuit for generating control signals START and SL. Control signal START resets all registers and the divide-by-8

counter before the encoding process begins. Control signal LD is simply generated by a divide-by-8 counter to load the z_i's into the R_i's in parallel.

since a codeword contains 255 symbols, the computation of a complete encoded codeword requires 255 "symbol cycles." A symbol cycle is the time interval for executing a complete cycle of Berlekamp's algorithm. Since a symbol has 8 bits, a symbol cycle contains 8 "bit cycles." A bit cycle is the time interval for executing one step in Berlekamp's algorithm. In this design a bit cycle requires a period of the clock cycle.

The layout design of this (255, 223) RS-encoder is shown in Fig. 9. Before the design of the layout each circuit was simulated on a general-purpose computer by using SPICE (a transistor-level circuit simulation program) (Ref. 7). The cir-

cuit requires about 3000 transistors, while a similar JPL design requires 30 CMOS IC chips (Ref. 5). This RS-encoder design will be fabricated and tested in the near future.

V. Concluding Remarks

A VLSI structure is developed for a Reed-Solomon encoder using Berlekamp's bit-serial multiplier algorithm. This structure is both regular and simple.

The circuit in Fig. 2 can be modified easily to encode an RS code with a different field representation and different parameters other than those used in Section IV. Table 2 shows the primary modifications needed in the circuit to change a given parameter.

Acknowledgment

References

1. Reefs, H. F. A., and Best, A. R., "Concatenated Coding on A Spacecraft-to-Ground Telemetry Channel Performance," Proc. ICC 81, Denver, 1981.

2. Odenwader, J., et al., "Hybrid Coding System Study Final Report," Linkabit Corp., NASA CR114, 486, September 1972.

3. MacWilliams, P. J., and Sloane, N. J. A., The Theory Of Error-Correcting Codes, North-Holand Publishing Co., v. 1978

4. Berlekamp, E. R., "Techn. Proposal for A Low-Power Reed-Solomon Encoder/Interleaver Using About 30 CMOS IC's," Cyclotomics, Inc., in response to RFP No. BP-6-9007.

5. Perlman, M. and Lee, J. J., "A Comparison of Conventional Reed-Solomon Encoders and Berlekamp's Architecture," New Technology Report NPO-15568, March 10, 1982, Case No. D-15568, NASA.

6. Mead, C., and Conway, L., Introduction To VLSI Systems, Addison-Wesley Publishing Company, Calif., 1980.

7. Negal, L. W., and Pederson, D. O., "SPICE - Simulation Program with Integrated Circuit Emphasis," Memorandum No. ERL-M362, Electronics Research Laboratory, University of California, Berkelev, April 12, 1973.

Table 1. Representations of elements over $GF(2^4)$
generated by $\alpha^4 = \alpha + 1$

Power j	Elements in normal base $\alpha^3 \alpha^2 \alpha^1 \alpha^0$	$Tr(\alpha^j)$	Elements in dual base $z_0 z_1 z_2 z_3$
*	0000	0	0000
0	0001	0	0001 λ_3
1	0010	0	0010 λ_2
2	0100	0	0100 λ_1
3	1000	1	1001
4	0011	0	0011
5	0110	0	0110
6	1100	1	1101
7	1011	1	1010
8	0101	0	0101
9	1010	1	1011
10	0111	0	0111
11	1110	1	1111
12	1111	1	1110
13	1101	1	1100
14	1001	1	1000 λ_0

Table 2. The primary modifications of the encoder circuit
in Fig. 2 needed to change a parameter

	Parameter to be changed	The value used for the circuit in Fig. 2	New value	The circuits of Fig. 2 that require modification
1	Generator polynomial	Eq. (8)	$g(x)$	The PLA of the Product Unit needs to be changed
2.	The finite field used	$GF(2^8)$	$GF(2^m)$	All registers are m-bit resistors, except Q is a $(m - 1)$-bit register. A divide-by-m counter is used. (The generator polynomial may not be changed.)
3	Error-correcting capability	16	t	$2t$-2 shift registers are required in the Remainder Unit. (The generator polynomial is also changed.)
4	Number of information symbols	223	k	None is changed, since k is implicitly contained in the control signal LM
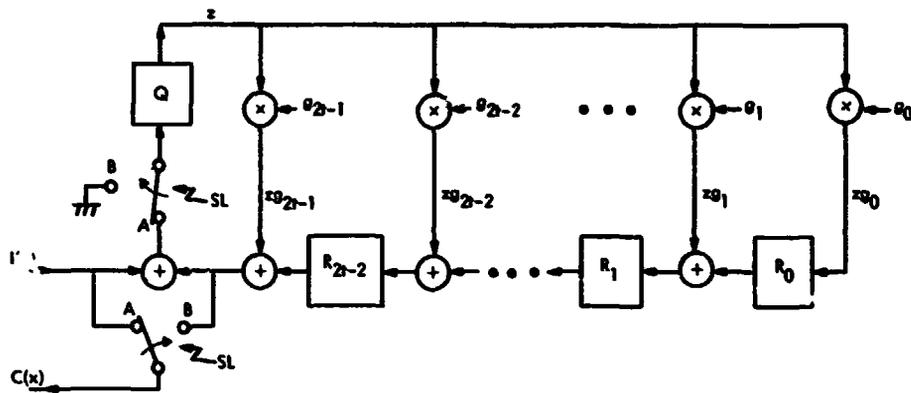
Fig. 1. A structure of a t-error correcting RS-encoder
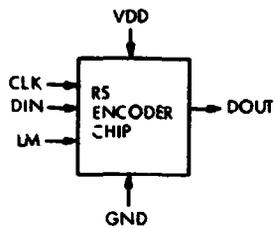
Fig. 2. Symbolic diagram of
a RS encoder chip
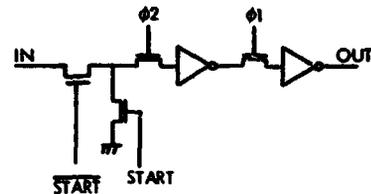
Fig. 3. Logic diagram of a 1-bit
dynamic register with reset
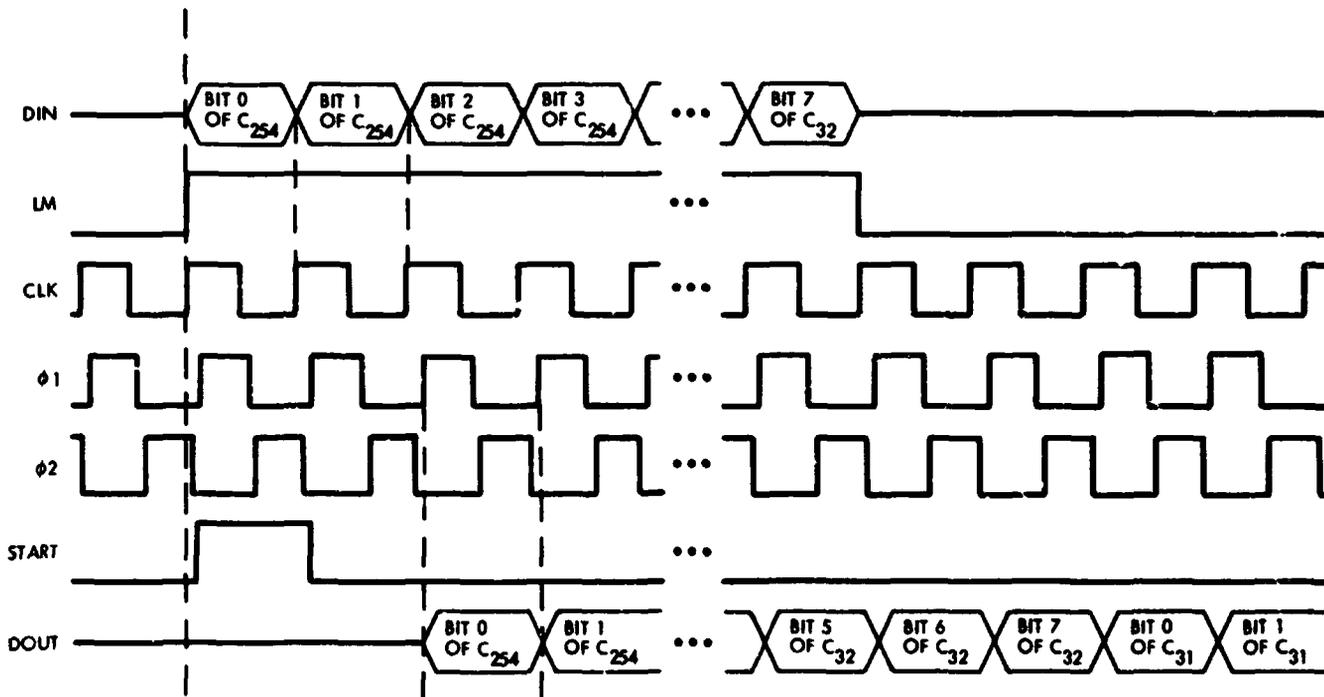
Fig. 4. The signals of DIN, LM, CLK, $\phi 1$, $\phi 2$, START, and DOUT
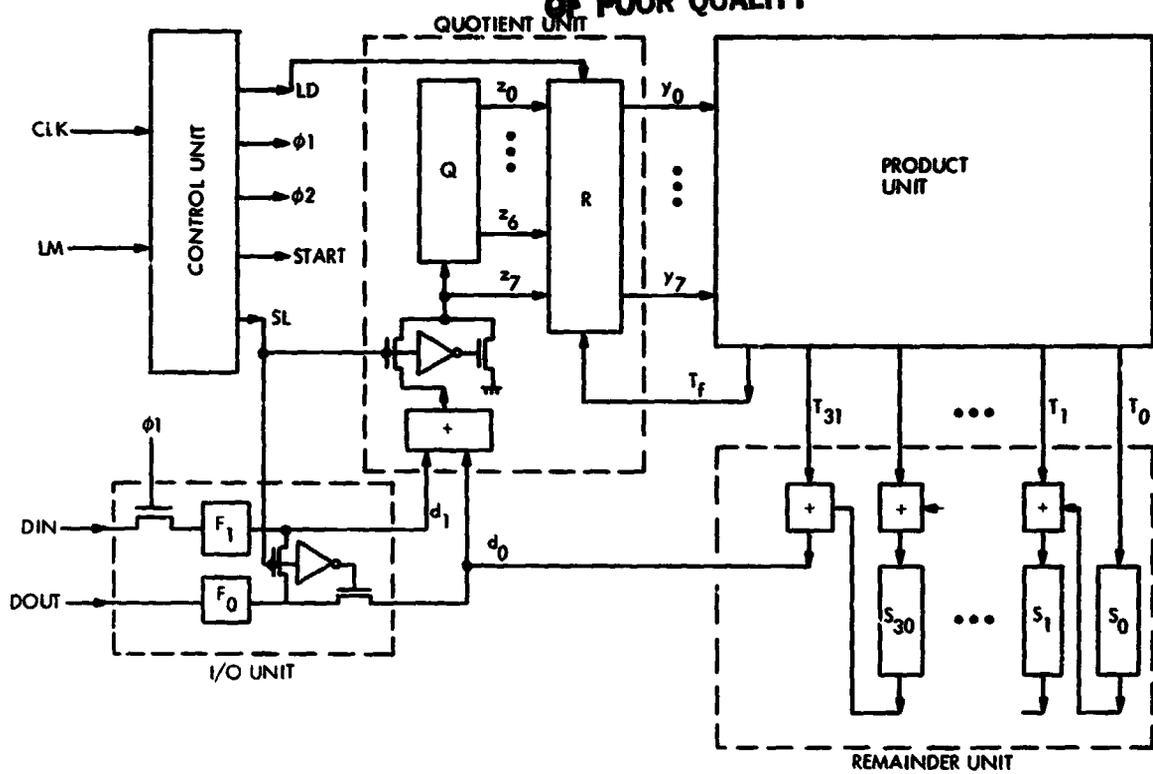
Fig. 5. Block diagram of the encoder
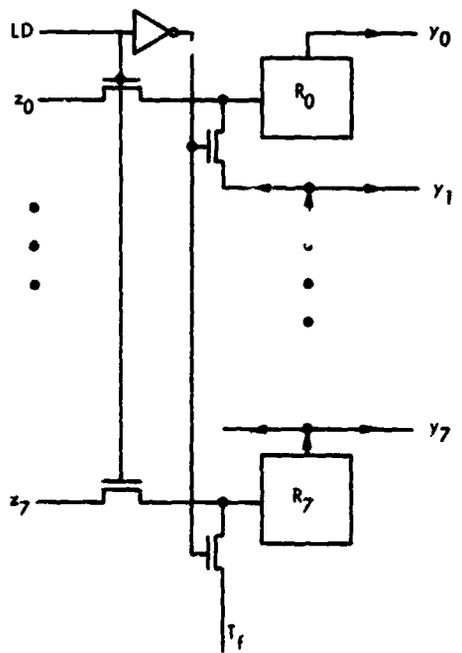
R$_i$: A 1-BIT REGISTER WITH RESET

Fig. 6. Block diagram of register R
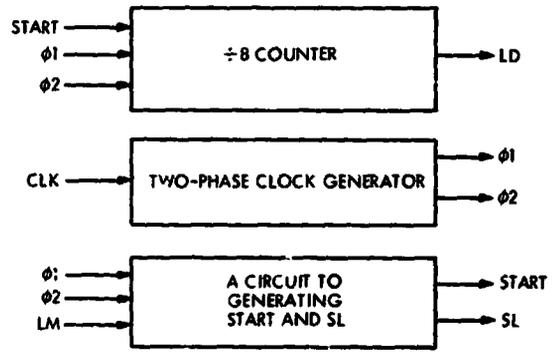
Fig. 7. Block diagram of the Control Unit
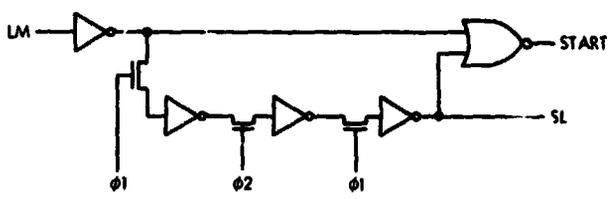
Fig. 8. Logic diagram of the circuit for generating control
signals START and SL
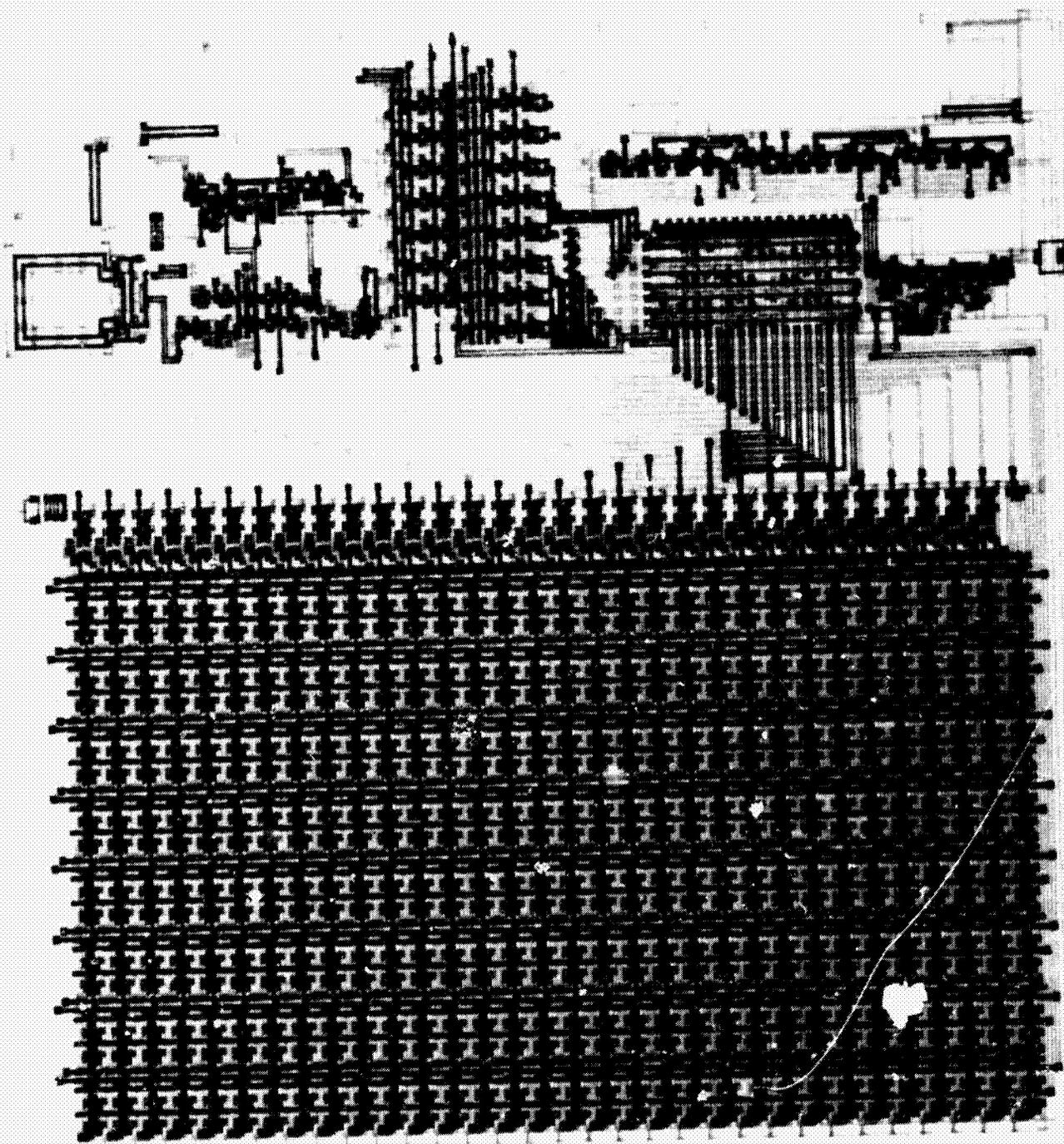
Fig. 9. Layout of the (255, 223) RS-encoder chip

Appendix A

In this appendix all 256 elements in $GF(2^8)$ are listed in Table A-1. These field elements are expressed in both the normal basis and its dual basis.

Table A-1. Representations of elements in $GF(2^8)$

| Power j	Elements in normal base	$Tr(\alpha^j)$	Elements in dual base	Power j	Elements in normal base	$Tr(\alpha^j)$	Elements in dual base
*	00000000	0	00000000	35	11101000	0	00110111
0	00000001	0	01111111	36	01010111	0	01101110
1	00000010	1	11111111	37	10101110	1	11011100
2	00000100	1	11111110	38	11011011	1	10111000
3	00001000	1	11111101	39	00110001	0	01110000
4	00010000	1	11111010	40	01100010	1	11100000
5	00100000	1	11110101	41	11000100	1	11000001
6	01000000	1	11101010	42	00001111	1	10000011
7	10000000	1	11010101	43	00011110	0	00000110
8	10000111	1	10101011	44	00111100	0	00001100
9	10001001	0	01010111	45	01111000	0	00011000
10	10010101	1	10101110	46	11110000	0	00110000
11	10101101	0	01011100	47	01100111	0	01100001
12	11011101	1	10111001	48	11001110	1	11000011
13	00111101	0	01110011	49	00011011	1	10000111
14	01111010	1	11100111	50	00110110	0	00001110
15	11110100	1	11001110	51	01101100	0	00011100
16	01101111	1	10011100	52	11011000	0	00111000
17	11011110	0	00111001	53	00110111	0	01110001
18	00111011	0	01110010	54	01101110	1	11100011
19	01110110	1	11100100	55	11011100	1	11000110
20	11101000	1	11001001	56	00111111	1	10001100
21	01011111	1	10010011	57	01111110	0	00011001
22	10111110	0	00100110	58	11111100	0	00110011
23	11111011	0	01001101	59	01111111	0	01100110
24	01110001	1	10011010	60	11111110	1	11001100
25	11100010	0	00110101	61	01111011	1	10011000
26	01000011	0	01101010	62	11110110	0	00110001
27	10000110	1	11010100	63	01101011	0	01100010
28	10001011	1	10101000	64	11010110	1	11000100
29	10010001	0	01010000	65	00101011	1	10001000
30	10100101	1	10100001	66	01010110	0	00010001
31	11001101	0	01000011	67	10101100	0	00100011
32	00011101	1	10000110	68	11011111	0	01000110
33	00111010	0	00001101	69	00101001	1	10001101
34	01110100	0	00011011	70	01110010	0	00011010

Table A-1 (contd)

Power j	Elements in normal base	$Tr(\alpha^j)$	Elements in dual base	Power j	Elements in normal base	$Tr(\alpha^j)$	Elements in dual base
71	11100100	0	00110100	112	01000111	1	10010100
72	0.001111	0	01101001	113	1^001110	0	00101C 1
73	10011110	1	11C10011	114	10011011	0	01010010
74	10111011	1	10100111	115	10110001	1	10100101
75	11110001	0	01001111	116	11100101	0	01001011
76	0110ui01	1	10011110	117	01001101	1	10010110
77	11001010	0	00111101	118	10011010	0	0010110i
78	0C010011	0	01111010	119	10110011		01011010
79	0010010.0	1	11110100	120	11100001	1	10110101
80	01001100	1	11101001	121	01000101	0	01101011
81	100110C0	1	11010010	122	10001010	1	..010111
82	10110111	1	10100100	123	10010011	1	10101111
83	11101001	0	01001000	124	10100001	0	01011111
84	01010101	1	10010001	125	11000101	1	10111110
85	10101010	0	00100010	126	00001101	0	01111100
86	11010011	0	01000101	127	000:1010	1	11111000
87	00100001	1	10001010	128	00110100	1	11110001
88	01000010	0	00010101	129	01101000	1	111 0010
89	10000100	0	00101011	130	11010000	1	1100010i
90	10001111	0	01010110	131	00100111	1	1000' '11
91	10011001	1	10101101	132	01001110	0	00010110
92	10110101	0	01011011	133	1001'100	0	00101100
93	11101101	1	10110110	134	10111111	0	01011001
94	01011101	0	01101100	135	11111001	1	.0110010
95	10111010	1	11011000	136	01110101	0	01100100
96	11110011	1	10110000	137	11101010	1	11001000
97	01100001	0	01100000	138	01010011	1	10010000
98	11000010	1	11000000	139	10100110	0	00100001
99	00000011	1	10000000 λ_0	140	11001011	0	01000010
100	00000110	0	00000001 λ_7	141	00010001	1	100021
101	00001100	0	00000011	142	00100010	0	00001010
102	00011000	0	000C0111	143	01000100	0	00010100
103	00110000	0	00001111	144	100 1000	0	00101000
104	01100000	0	00011111	145	10010111	0	01010001
105	11000000	0	00111111	146	10101001	1	10100010
106	00000111	0	01111110	147	11010101	0	01000100
107	00001110		11111100	148	00101101	1	10001001
108	00011100	1	11111001	149	01011010	0	00010010
109	00111000	1	11110010	150	10110100	0	00100100
110	01110000	1	11100101	151	11101111	0	01 010C
111	11100000	1	11001010	152	01011001	1	10010010

Table A-1 (contd)

Power j	Elements in normal base	$Tr(\alpha^j)$	Elements in dual base	Power j	Elements in normal base	$Tr(\alpha^j)$	Elements in dual base
153	10110010	0	00100101	194	01001001	0	01101000
154	11100011	0	01001010	195	10010010	1	11010000
155	01 ·001	1	10010101	196	10100011	1	10100000
1..	· ·010	0	00101010	197	11000001	0	01000000 λ_1
15.	10.0011	0	01010101	198	00000101	1	$\overline{10000001}$
15.	10000001	1	10101010	199	00001010	0	$\overline{00000010}\,\lambda_6$
159	10000101	0	01010100	200	00010100	0	$\overline{00000100}\,\lambda_5$
160	10001101	'	10101001	201	00101000	0	$\overline{00001000}\,\lambda_4$
161	10011101	0	01010011	202	01010000	0	$\overline{00010000}\,\lambda_3$
162	10111101	1	10100110	203	10100000	0	$\overline{00100000}\,\lambda_2$
163	11111101	0	01001100	204	11000111	0	$\overline{01000001}$
164	01111101	1	10011001	205	00001001	1	10000010
165	11111010	0	00110010	206	00010010	0	00000101
166	01110011	0	01100101	207	00100100	0	00001011
167	11100110	1	11001011	208	01001000	0	00010111
168	01001011	1	10010111	209	10010000	0	00101111
169	10010110	0	00101110	210	10100111	0	01011110
170	10101011	0	01011101	211	11001001	1	10111101
171	11010001	1	10111010	212	00010101	0	01111011
172	00100101	0	01110100	213	00101010	1	11110111
173	01001010	1	11101000	214	01010100	1	11101110
174	10010100	1	11010001	215	10101000	1	11011101
175	10101111	1	10100011	216	11010111	1	10111011
176	11011001	0	01000111	217	00101001	0	01110111
177	00110101	1	10001110	218	01010010	1	11101111
178	01101010	0	00011101	219	10100100	1	11011110
179	11010100	0	00111011	220	11001111	1	10111100
180	00101111	0	01110110	221	00011001	0	01111000
181	01011110	1	11101100	222	00110010	1	11110000
182	10111100	1	11011001	223	01100100	1	11100001
183	11111111	1	10110011	224	11001000	1	11000010
184	01111001	0	01100111	225	00010111	1	10000100
185	11110010	1	11001111	226	00101110	0	00001001
186	01100011	1	10011111	227	01011100	0	00010011
187	11000110	0	00111110	228	10111000	0	00100111
188	00001011	0	01111101	229	11110111	0	01001110
189	00010110	1	11111011	230	01101001	1	10011101
190	00101100	1	11110110	231	11010010	0	00111010
191	01011000	1	11101101	232	00100011	0	01110101
192	10110000	1	11011010	233	01000110	1	11101011
193	11100111	1	10110100	234	10001100	1	11010110

C-2

Power j	Elements in normal base	$Tr(\alpha^j)$	Elements in dual base	Power j	Elements in normal base	$Tr(\alpha^j)$	Elements in dual base
235	10011111	1	10101100	245	01111100	1	11100110
236	10111001	0	01011000	246	11111000	1	11001101
237	11110101	1	10110001	247	01110111	1	10011011
238	01101101	0	01100011	248	11101110	0	00110110
239	11011010	1	11000111	249	01011011	0	01101101
240	00110011	1	10001111	250	10110110	1	11011011
241	01100110	0	00011110	251	11101011	1	10110111
242	11001100	0	00111100	252	01010001	0	01101111
243	00011111	0	01111001	253	10100010	1	11011111
244	00111110	1	11110011	254	11000111	1	10111111

Appendix B

The binary mapping matrix for $\gamma = \alpha^{11}$ of the (255, 223) RS-encoder is given by

$$
\begin{bmatrix}
T_0 \\ T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \\ T_7 \\ T_8 \\ T_9 \\ T_{10} \\ T_{11} \\ T_{12} \\ T_{13} \\ T_{14} \\ T_{15} \\ T_{16}
\end{bmatrix}
=
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 1 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
Z_0 \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \\ Z_7
\end{bmatrix}
$$

Power Combining of Semiconductor Lasers: A Review

J. Katz

Communications Systems Research Section

The subject of power combining of semiconductor lasers is reviewed. Several methods of coherent power combining are described and compared. A comparison is also made between coherent and incoherent power combining, and important operational characteristics are considered. It is found that in communication links with demanding requirements coherent power combining is necessary.

I. Introduction

The growing interest in optical communications for free space applications (e.g., deep space, intersatellite links) (Refs. 1-3) has increased the need for appropriate light emitting sources. As discussed in an earlier report (Ref. 4), semiconductor injection lasers are excellent candidates for this application, particularly because of their very long lifetimes, high efficiency and small size and weight. Their drawback is that in high power (about 1 watt average) applications, a single device cannot radiate all the power needed in a stable radiation pattern and frequency. In the above-mentioned report (Ref. 4) and in subsequent ones (Refs. 5, 6), several aspects of solving this problem via mutal phase locking of several lasers through overlapping fields were analyzed.

It is the purpose of this report to review the general problem of power combining of semiconductor lasers by coherent and incoherent methods. Section II compares several methods of coherent power combining, namely mutual coupling (discussed before), injection locking and amplification. Regenerative amplification is also mentioned. The various methods are briefly described and then compared on the basis of several important operational characteristics, such as locking range,

power efficiency, thermal considerations, reliability, monolithic implementation, realization of two-dimensional configurations and the need for additional components. It is found that all the coherent methods are similar in their problems and performance, although coherent amplification might be somewhat better. Section III compares coherent and incoherent power combining. Again the two methods are briefly described and then compared on the basis of the spatial and spectral characteristics of their resulting radiation. It is found that although incoherent power combining is easier to implement, the significant advantages offered by coherent power combining seem to justify the additional efforts needed to realize devices based on these methods, especially in systems whose designs impose stringent requirements on beam directivity and optical background noise immunity.

II. Comparison Between Methods of Coherent Power Combining

In this section three methods for coherent power combining, namely mutual coupling, injection locking and amplification are discussed. The schematic configurations of these methods are shown in Fig. 1. The common feature of all these

methods is the establishment of some coherent interaction between all the elements of the array.

In the case of mutual coupling (Fig. 1a), no laser in the array has a privileged status. There is a certain amount of coupling among the lasers, which, under certain conditions, results in their synchronization. This method has been analyzed (Ref. 5) and demonstrated (Refs. 4, 6). In particular, it was found that phase-locking requires that the following inequality be approximately satisfied:

$$2 \left| \frac{\Delta \omega}{\omega_0} \right| \lesssim |\xi| \qquad (1)$$

$\Delta \omega$ is approximately equal to the rms deviation of the lasers' oscillation frequencies from the "center-of-mass" frequency ω_0 of the array and ξ is a dimensionless parameter describing the strength of the coupling interaction (Ref. 5). In the case of coupling due to field overlap of lasers which are in close proximity, ξ can assume in AlGaAs lasers the maximum value of (Ref. 5).

$$|\xi|_{max} \cong \frac{8 \cdot 10^{-4}}{d^2} \qquad (2)$$

where d is the distance between individual lasers of the array in micrometers. Coupling due to other mechanisms (e.g., diffraction) is also possible.

Injection locking of lasers (Fig. 1b) is obtained by similar physical mechanisms. In this case, however, there is a master laser oscillator. Portions of its emitted radiation are coupled simultaneously into all the other lasers in the array, forcing them to oscillate at its frequency. There is no coupling among the lasers in the array and no coupling from the array back into the master laser. Injection locking was analyzed in electrical oscillators (Ref. 7) and in lasers (Ref. 8), and has been experimentally demonstrated in lasers (Ref. 9). The condition for injection locking of two lasers can be expressed as (Ref. 7):

$$2 \left| \frac{\Delta \omega}{\omega_0} \right| < \frac{\delta E_m}{QE} \qquad (3)$$

where $\Delta \omega$ is the difference between the (radian) frequencies of the master laser and the array element laser, E_m is the electric field strength of the master laser oscillator, δ is the fraction of it that is coupled to the array element laser whose electric field strength is E, and the figure of merit of its cavity is Q.

Coherent amplification (Fig. 1c) is similar to injection locking: in both cases there is a master laser-oscillator. However, in the case of coherent amplification, the elements of the array are only gain-elements (i.e., amplifiers) without feedback. (This is accomplished by coating the laser mirrors with anti-reflection coating thus eliminating its feedback mechanism.) Light generated in the master laser-oscillator is split and fed simultaneously into all the gain elements, where a travelling-wave amplification is employed. The amplified outputs of the amplifiers are automatically phase-locked (provided, of course, that the output of the master oscillator is coherent over its near-field pattern). Amplification in semiconductors has been analyzed (Ref. 10), and the operation of coherently amplified GaAs homojunction devices has been demonstrated (Ref. 11).

An intermediate case between injection-locking and coherent amplification occurs when the gain elements in the array do have some amount of feedback, but it is insufficient to produce lasing; i.e., they operate as regenerative amplifiers (Ref. 12). Since the added complexity in implementing this method (the regenerative amplifier has to be biased very accurately just below the lasing threshold) does not yield improved performance over either regular amplification or injection locking, it will not be further considered here.

Schematic configurations of the above methods are shown in Fig. 1. In the following paragraphs they are compared from the aspects of locking range, power efficiency, thermal considerations, reliability, monolithic implementation, realization of two-dimensional configurations and the need of additional components.

A. Locking Range

We are considering the problem of the range of frequencies $\Delta \omega / \omega_0$ over which phase-locking can be maintained. In the case of mutual locking (Eqs. 1, 2), and for lasers that are spaced about $d \cong 5 \ \mu m$ apart, the result is $2 |\Delta \omega / \omega_0| \lesssim 3 \cdot 10^{-5}$. For injection locking (Eq. 3) with $E_m \cong E$, $Q \cong 10^4$ and $\delta = 0.1$, we obtain $2 |\Delta \omega / \omega_0| \lesssim 10^{-5}$. Since the actual requirement for phase-locking in the case of mutual coupling can actually be somewhat more stringent than the one expressed in Eq. (1) (Ref. 5), both methods have basically the same locking range. (This result applies also in the case of regenerative amplifiers.)

B. Power Efficiency

Under optimized conditions, all the coherent phase-locking methods basically have the same power efficiency. The reason is that the photon density distribution in semiconductor lasers that are optimized for power efficiency is very similar to the photon density distribution in travelling wave amplifiers (Ref. 13). Second-order differences between the methods

result from different coupling losses among the lasers or between the master-laser oscillator and the rest of the array.

C. Thermal Considerations

One of the problems in mutual coupling is that the lasers have to be put in close proximity (several micrometers) to one another so that sufficient coupling will be established among them (Ref. 5). This aggravates the problems of removing excess heat generated in the laser junctions and ohmic contacts. This problem can be mitigated by employing injection-locking or amplification, since in this case no mutual coupling has to be established among the elements of the array, and thus they can be placed further apart. However, doing that presents two new problems. First problem is that of efficient coupling from the master-laser to the array. Thus there is a tradeoff between thermal performance and the number of elements that can be locked, and the optimum configuration must be found in each case. The second problem is that as the array elements are further apart than in the mutual-coupling case, the increased separation causes the radiation pattern of the array to have more grating sidelobes (see next section).

D. Reliability

The mutual coupling approach is potentially more reliable than the other approaches since the performance of the entire array can, in principle, be designed in such a way that it is not critically affected by a failure of a single element. In the case of either injection-locking or amplification, failure of the master laser-oscillator means failure of the whole array. However, since the reliability of semiconductor laser devices is adversely affected at elevated temperatures, the actual advantage of the mutual coupling method can become insignificant because of its potentially inferior thermal characteristics.

E. Monolithic Implementation

Because of their simpler configurations, arrays based on mutual coupling are somewhat more amenable to monolithic integration than arrays which use injection locking or coherent amplification.

F. Realization of Two-Dimensional Configurations

This parameter is important for achieving reduction of the far-field pattern of the array in both directions (see next section). Generally, arrays based on injection locking or on coherent amplification can be more readily arranged in two-dimensional configurations (with a probable penalty of increased losses in the coupling from the master laser).

G. Additional Components

In all the coherent methods a phase-shifter in tandem with each array element is needed so that the individual phases (which are locked, but not necessarily at the desired values) can be modified to yield the desired radiation pattern. In addition, when employing injection locking or coherent amplification, there is also a need of optical isolators so that light that is generated by the array elements will not be coupled back into the master laser and thus interfere with the overall operation of the array. Such isolators can introduce some additional losses in the coupling from the master laser oscillator.

Before concluding this section it is important to note that the choice of the optimum method depends on the overall system parameters. Since there is no single coherent power combining method with decided advantages over the others, a detailed comparison between the coherent power combining method has to be carried out in any case of a particular system design. However, all other things being equal, it seems that the coherent amplification method is somewhat better than the other methods, delivering essentially the same performance without having to satisfy the additional and rather stringent requirement for synchronization of two (or more) oscillators.

III. Comparison Between Coherent and Incoherent Power Combining

In this section a comparison between coherent and incoherent methods of power combining of semiconductor lasers is carried out. In order to review the basic differences between the two approaches, a simplified one-dimensional analysis is first presented.

Assume a set of M identical lasers at locations $\{d_n\}$, $n = 1$, $2, \ldots M$. The near-field pattern of each laser (i.e., the field distribution at its output facet) is denoted by $\mathscr{E}(x)e^{i\phi}n$ where $\mathscr{E}$ and ϕ are the field amplitude and phase, respectively. The near-field of the whole array $\mathscr{E}_t$ is thus given by

$$\mathscr{E}_t(x) = \sum_{n=1}^{M} \mathscr{E}(x - d_n)e^{i\phi}n \qquad (4)$$

In the case of coherent power combining, the ϕ_n's in Eq. (4) are fixed numbers. The far-field intensity distribution of the array (i.e., its radiation pattern), I_{coh}, is approximately given by (Ref. 14, 15)

$$I_{coh}(\theta) = |\mathscr{F}\{\mathscr{E}(x)\} \cos\theta|^2 \cdot \left| \sum_{n=1}^{M} e^{i\left(\frac{2\pi}{\lambda} d_n \sin\theta + \phi_n\right)} \right|^2 \qquad (5)$$

where θ is the far-field angle and $\mathscr{F}\{\circ\}$ denotes a Fourier-transform operation.

In the case of incoherent power combining, the ϕ_n's in Eq. (4) are random variables. We can assume that over all the relevant time periods, the random fluctuations of the ϕ_n's are fast enough so that the cross terms that appear when calculating the intensity average to zero (for example, even wavelength separation of 1Å at $\lambda = 1$ μm corresponds to 30 GHz, which is much faster than typical detector bandwidths). The far-field intensity pattern in this case is

$$I_{inc}(\theta) = M \mid \mathscr{F}\{\mathscr{E}(x)\} \cos \theta \mid^2 \tag{6}$$

As expected, no cross-interference terms are present, and the far-field pattern of the incoherent array is an amplified version of the far-field pattern of its elements.

In the following paragraphs a comparison between coherent and incoherent power combining of semiconductor lasers will be made. In two important aspects, namely, improved radiation pattern and spectral distribution, coherent power combining has a significant advantage over incoherent power combining. Several advantages of incoherent power combining will also be presented.

A. Far-Field Pattern

From Eq. (5), which describes the case of coherent power combining, it is anticipated that by a judicious choice of the d_n's and adjustment of the ϕ_n's, the resulting beam pattern can become narrower, in a similar fashion to microwave phased arrays. The reduction of the angular extent of the beam pattern is an impoitant feature of coherent power combining, since narrower beams make the task of subsequent beam narrowing foi high-directivity free-space transmission much easier. (It should be emphasized that two-dimensional arrays are needed to obtain a reduction of the far-field beam pattern in both the horizontal and vertical planes.)

As a simple illustrative example, we describe the near-field profile of a single device by

$$\mathscr{E}(x) = \begin{cases} E_0 & |x| < a \\ 0 & |x| > a \end{cases} \tag{7}$$

The incoherent far-field intensity pattern is calculated from Eqs. (6) and (7) to be

$$I_{inc}(\theta) = 4_a^2 E_0^2 M \cos^2 \theta \operatorname{sinc}^2 \left(2\pi \frac{a}{\lambda} \sin \theta\right) \tag{8}$$

where $\operatorname{sinc}(Z) \equiv (\sin Z)/Z$. The far-field intensity of the coherent array is calculated in a similar fashion from Eqs. (5) and (7). For the case of $\phi_n \equiv 0$ and $d_n = n \cdot d$, the result is

$$I_{coh}(\theta) = I_{inc}(\theta) \cdot M \cdot \left[\frac{\sin \left(2\pi \frac{d \cdot M}{\lambda} \sin \theta\right)}{M \sin \left(2\pi \frac{d}{\lambda} \sin \theta\right)} \right]^2 \tag{9}$$

The distributions described by Eqs. (8) and (9) are shown in Fig. 2a for the following values of parameters: $\lambda = 0.9$ μm, $a = 2$ μm, $d = 9$ μm and $M = 10$.

The three important features of the far-field patterns, as deduced from Eqs. (8), (9) and shown in Fig. 2a, are:

(1) The intensity of the radiation in the forward direction ($\theta = 0$) is increased by a factor of M by using coherent instead of incoherent power combining.

(2) Under the same conditions, the angular extent of the forward direction radiation lobe is reduced by a factor of $(Md/2a)$.

(3) Coherent power combination is accompanied by the presence of grating lobes. Some of the problems of energy waste and pointing ambiguity associated with them can be mitigated by randomizing the locations of the array elements (Refs. 16-18). A calculated example is shown in Fig. 2b. All the parameters of the array are the same as before, but now the location of each element is randomly distributed within ±2 μm of its deterministic location. (In the case of mutual coupling, sufficient coupling should be maintained also in the new random locations.) It is clearly seen that the level of sidelobes is significantly reduced. The improvement increases with the number of elements of the array and with the amount of randomization allowed in their locations.

B. Spectral Characteristics

Semiconductor laser materials have wide gain linewidths, and thus they can support lasing modes over the range of many angstroms (Ref. 19). When we have an incoherent array of lasers, then even though each of them has an (almost) identical spatial beam pattern, the lasing wavelength will differ from one laser to another, due to minor differences in their lengths, currents, etc. In order for the receiver to collect all the spectral content (i.e., energy) of the received signal, a wide optical filter has to be used, with the unavoidable consequences of admitting more background radiation noise into the system. Systems employing phase-locked arrays, on the other hand, can use much narrower optical filters at the receiver – provided, of course, that the array elements and the array itself oscillate in a single longitudinal mode (i.e., a single spectral line). Single longitudinal mode operation has been demonstrated in many types of laser diodes (Ref. 20) and in laser

diodes placed in external cavities (Ref. 21), and it is conceivable that when these diodes are used as elements in the array, it will oscillate in a single longitudinal mode. The narrower optical filter banawidths which can be used in conjunction with coherent arrays can result in a significant reduction (up to several orders of magnitude) in the amount of background noise radiation detected by the receiver.

It is also worthwhile to mention some practical considerations pertaining to the use of operation of optical filters. Although the inherent laser linewi.'th is very narrow – less than 10^{-3}Å (Refs. 20, 22) – such narrowband optical filters cannot be implemented yet. As of today, the best demonstrated filters have bandwidths of the order of 10^{-1} Å (Ref. 23). They can also be electronically tuned, which is necessary for compensating wavelength drift due to doppler shifts and temperature variations at the transmitter. (AlGaAs sem'conductor injection lasers have wavelength temperature variations of the order of 0.5 to 4Å/K.)

C. Advantages of Incoherent Power Combining

Incoherent power combining is much easier to implement than coherent power combining, and that is its basic advantage.

No effort has to be made in order to synchronize the lasers, no external optical components (e.g., phase-shifters, isolators) are needed for the array implementation, the thermal performance is potentially better, and two-dimensional configurations are easier to construct. The design of an incoherent array is free from the many constraints imposed by the requirement of phase-locking. However, although incoherent power combining is easier to implement, the significant advantages offered by coherent power combining (namely, improved power directivity and narrower spect:al extent) seem to justify the additional efforts needed to realize devices based on these methods.

IV. Conclusions

Methods of coherent and incoherent power combining of semiconductor lasers have been described. It was found that although incoherent power combining is easier to implement, the significant advantages offered by coherent power combining seem to justify the additional efforts needed to realize devices based on this method. This conclusion is true, particularly in systems which require very high beam directivity and narrow spectral range of the transmitted radiation.

References

1. Pierce, J. R., "Optical Channels: Practical Limits with Photon Counting," *IEEE Trans. Comm.*, COM-26, pp. 1819-1821 (1978).

2. Lesh, J. R., Katz, J., Tan, H. H., and Zwillinger, D., "2.5 Bit/Detected Photon Demonstration Program: Description, Analysis and Phase I Results," *TDA Progress Report 42-66*, pp. 115-132, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1981.

3. Gagliardi, R. M., Vilnrotter, V. A., and Dolinar, S. J., Jr., "Optical Deep Space Communication via Relay Satellite," Publication 81-40, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1981.

4. Katz, J., "High Power Semiconductor Lasers for Deep Space Communications," *TDA Progress Report 42-63*, pp. 40-50, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1981 (and references therein).

5. Katz, J., "Phase-Locking of Semiconductor Injection Lasers," *TDA Progress Report 42-66*, pp. 101-114, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1981.

6. Katz, J., "Phase Cont.ol and Beam Steering of Semiconductor Laser Arrays," *TDA Progress Report 42-68*, pp. 42-49, Jet Propulsion Laboratory, Pasadena, Calif., April 15, 1982 (and references therein).

7. Adler, R., "A Study of Locking Phenomena in Oscillators," *Proc. IRE*, 34, pp. 351-357, 1946.

8. Tung, C. L., and Statz, H., "Phase-Locking of Laser Oscillators by Injected Signal," *J. Appl. Phys.*, 38, pp. 323-324 (1967).

9. Stover, H. L., and Steier, W. H., "Locking of Laser Oscillators by Light Injection," *Appl. Phys. Lett.*, 8, pp. 91-93, 1966.

10. Stern, F., "Saturation in Semiconductor Absorbers and Amplifiers of Light," in *Proc. Physics of Quantum Electronics Conf.*, McGraw-Hill, N.Y., 1966, pp. 442-449.

11 Vuillenmier, R., Collins, N. C., Smith, J. M., Kim, J. C. S., and Raillard, H., "Coherent Amplification Characteristics of a GaAs Phased Array," *Proc. IEEE*, 55, pp. 1420-1425, 1967.

12. Chang, M. B., and Garmire, E., "Amplification in Cleaved-Substrate Lasers," *IEEE J. Quant. Electron.* QE-16, pp. 997-1001, 1980.

13. Katz, J., "Power Efficiency of Semiconductor Injection Lasers," *TDA Progress Report 42-66*, pp. 94-100, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1981.

14. Casey, H. C., Jr., and Panish, M. B., *Heterostructure Lasers*, Academic Press, New York, 1978, Chapter 2.

15. Parrent, C. B., Jr., and Thompson, B. J., "Physical Optics Notebook," SPIE, Redondo Beach, California, 1971, Chapter 3.

16. Lo, Y. T., "A Mathematical Theory of Antenna Arrays with Randomly Spaced Elements," *IEEE Trans. Ant. Prop.*, AP-12, pp. 257-268, 1964.

17. Lo, Y. T., and Simcoe, R. J., "An Experiment on Antenna Arrays with Randomly Spaced Elements," *IEEE Trans. Ant. Prop.*, AP-15, pp. 231-235, 1967.

18. Corcoran, V. J., and Crabbe, I. A., "Electronically Scanned Waveguide Laser Arrays," *Applied Optics*, 13, pp. 1755-1757, 1974.

19. Yariv, A., *Quantum Electronics*, 2nd Edition, Wiley, New York, 1975, Chapter 10.7.

20. Nakamura, M., "Single Mode Operation of Semiconductor Injection Lasers," *IEEE Trans. Circuits and Systems*, CAS-26, pp. 1055-1062, 1979 (and references therein).

21. For example, Rediker, R. H., Scloss, R. P., Mooradian, A., and Welford, D., "External Cavity Controlled Operation of a Semiconductor Diode Gain Element in Series with an Optical Fiber," Topical Meeting on Integrated and Guided Wave Optics, Jan. 6-8, 1982, Pacific Grove, Calif.

22. Takakura, T., Iga, K., and Tako, T., "Linewidth Measurement of a Single Longitudinal Mode AlGaAs Laser with a Fabri-Perot Interferometer," *Japan. J. Appl. Phys.*, 19, pp. L725-L727, 1980.

23. Title, A. M., and Rosenberg, W. J., "Tunable Birefringent Filters." *Optical Engineering*, 20, pp. 815-823, 1981.
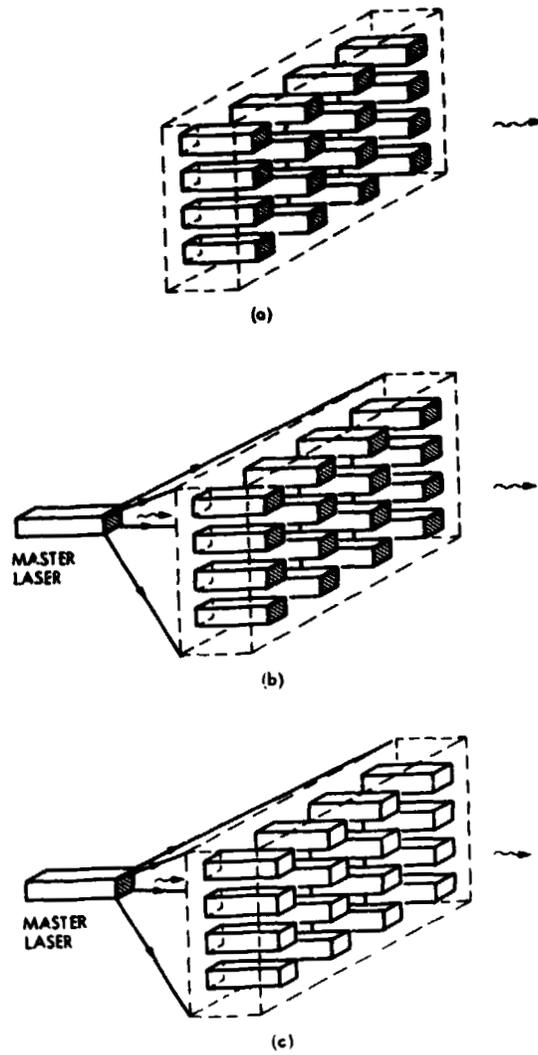
Fig. 1. Schematic configuration of coherent power combining methods: (a) mutual coupling, (b) injection locking (the array elements are lasers), (c) coherent amplification (the array elements are amplifiers).
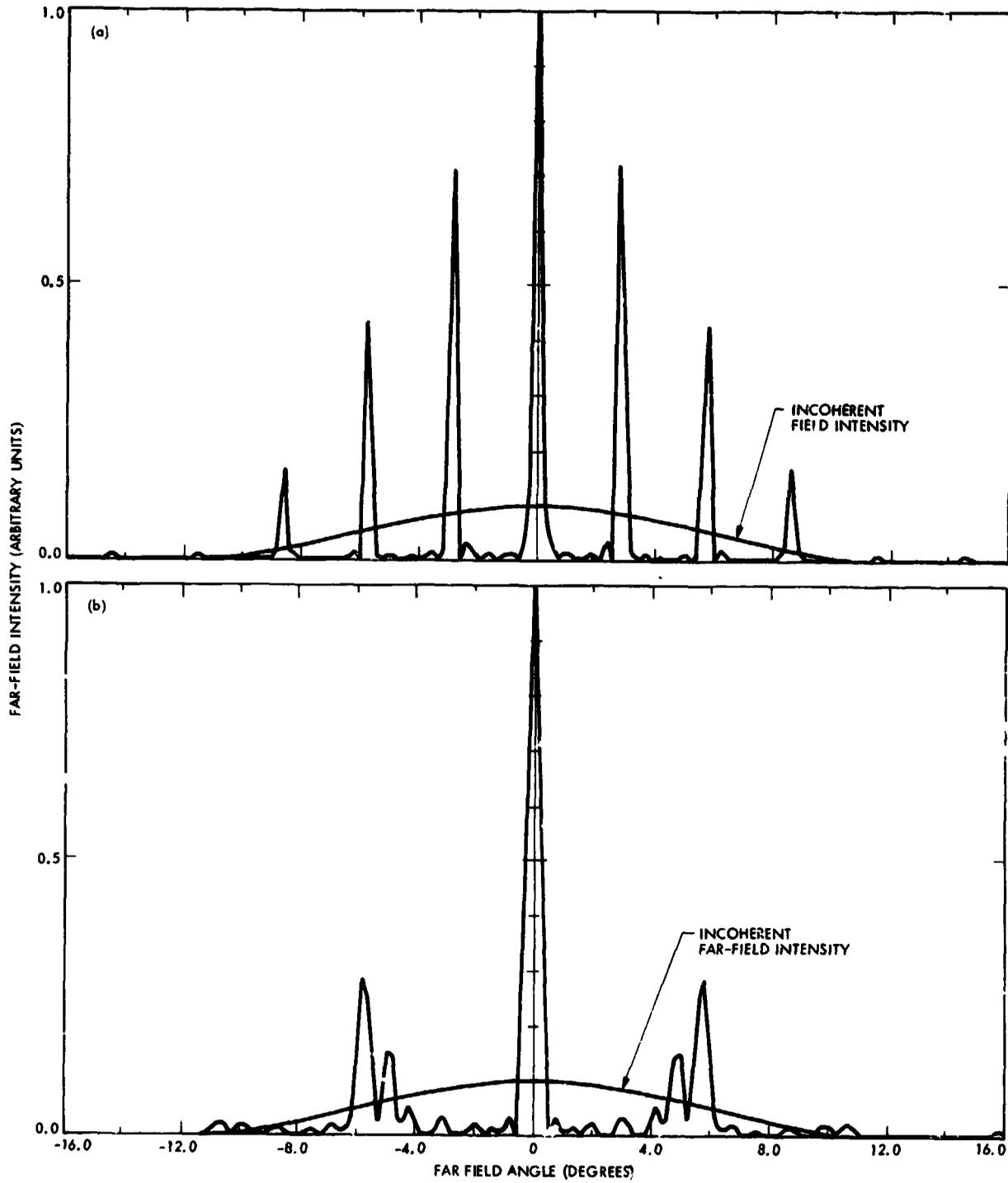
Fig. 2. Approximate far field pattern of a 10-element diode laser coherent array (laser aperture: 4 μm; wavelength: 0.9 μm): (a) array elements are regularly spaced 9 μm apart, (b) array elements are randomly distributed within ± 2 μm of their deterministic location in (a) (also shown is the far-field pattern in the case of an incoherent array)

2.5 Bit/Detected Photon Demonstration Program: Phase II and III Experimental Results

J. Katz

Communications Systems Research Section

This report describes recent progress in the experimental program for demonstrating, in the lab, an energy-efficient optical communication channel operating at a rate of 2.5 bits/detected photon. Results of the uncoded PPM channel performance are presented. These results indicate that the above throughput efficiency can be achieved not only with a Reed-Solomon (255, 191) code as originally predicted, but with less complex (255, 223) code as well.

I. Introduction

The purpose of this report is to describe the progress in the experimental program by demonstrating in the lab an energy-efficient optical communication channel operating at a rate of 2.5 bits/detected photon. The overall scope and analysis of the program, including phase-I experimental results, were described in an earlier report (Ref. 1), and the more general aspects and advantages of free space optical communications can be found in Refs. 2 and 3.

For the sake of completeness, the block diagram of the demonstration system (Ref. 1) is shown in Fig. 1. The optical portion of the system consists mainly of a gallium arsenide semiconductor injection laser and a direct detection photomultiplier tube. Surrounding the optoelectronic components are the modulation and coding hardware, namely, a 256 slot/word PPM modulation/demodulation system and an 8-bit Reed-Solomon encoding/decoding system, respectively. The demonstration program is divided into four phases, as indicated in Fig. 1. Phase I (Ref. 1) involved only the PMT and its associated preamplifier and was concerned with characterizing the

dark current noise distribution of the detection system. This report describes phase II (measuring optical pulse erasure and error statistics) and phase III (measuring PPM word error and word erasure probabilities). The final phase (not described here) will encompass the coding hardware and will demonstrate the 2.5 bits/detected photon goal.

The outline of this report is as follows. In Section II the calibration of the optical link is described. Correct calibration is essential for a relevant comparison between theory and experiment and a meaningful evaluation of the channel performance. Section III is concerned with the pulse detection statistics of the photomultiplier tube (PMT). Finally, Section IV presents the results of the uncoded PPM link. The modulation scheme used is 256 slots/word PPM, transmitted at a rate of 39,062 words/sec, which corresponds to an uncoded data rate of 312 kbits/sec. This rate is more than twice the benchmark rate of Voyager at Jupiter. Projecting from the experimental results of the uncoded PPM into the coded performance, we can predict with greater confidence that the energy efficiency goal of 2.5 bits/detected photon can be achieved.

II. Calibration of Optical Link

The calibration of the optical link is a crucial step in the experiment, since in order to determine the absolute performance of the system in terms of bits/detected photon, the number of detected photons must be determined as accurately as possible.

The optical setup inside the darkroom enclosure is shown in Fig. 2. The light emitter is a GaAs injection laser diode (Mitsubishi, TJS type; model ML-3001). It emits light in a single spatial and longitudinal mode. The lasing wavelength of the laser diode used in this experiment was around 0.81 μm. The current flowing through the laser (i_L) is monitored with a current probe (American Laser Systems, model 711), and the power emitted out of the laser (P_L) is monitored by a photodiode which is included in the laser package (not shown in Fig. 2). These two parameters (i_L and P_L) are not important for this particular experiment, but they have to be monitored so as to not exceed the absolute maximum range ratings of the device. The light emitted by the laser is collimated by a lens and passed through an iris diaphragm which limits the spatial extent of the beam to dimensions smaller than those of the photomultiplier tube (PMT) photocathode. A cube beam-splitter diverts part of the laser radiation into a calibrated photodiode (UDT, model PIN-10), which monitors the actual amount of light entering the photomultiplier tube. The other portion of the light is attenuated by neutral density filters (and, to some extent, also by other glass surfaces which are present in front of the photocathode, e.g., PMT faceplate, PMT housing window). The signal of the calibrated photodiode is amplified and displayed on an oscilloscope. The output signal of the PMT is amplified and fed into a counter (HP5370A). By a straightforward calculation one can find that the number of photoelectrons counted per second N_{pe} is related to the photodiode voltage signal displayed on the oscilloscope V_{pd} by the following formula

$$N_{pe} = \frac{\eta\lambda\left(\dfrac{T}{R}\right)}{ZA\sigma hcL} (DF) P_d V_{pd} \qquad (1)$$

where η is the quantum efficiency of the PMT's photocathode, λ is the radiation wavelength, (T/R) is the ratio between the intensity of the wave transmitted by the beamsplitter to the intensity of the wave reflected by it, Z is the impedance seen by the photodiode, A is the amount of amplification of the photodiode signal, σ is the responsivity of the photodiode, h is Planck's constant, c is the vacuum light velocity, L is the total amount of attenuation (i.e., the ratio between the power transmitted by the beamsplitter and the power incident on the PMT photocathode). DF is the duty factor of the light signal, and P_d is the probability of counting a photoelectron once it

is released by the photocathode. In our case $\lambda = 0.81$ μm, $(T/R) = 32/68$, $Z = 50\Omega$, $A = 100$, $\sigma = 0.3$ A/W, and $\eta = 0.16$. Using these parameters, Eq. (1) reduces to

$$N_{pe} = 2.05 \cdot 10^{14} \frac{(DF)P_d}{L} V_{pd} \qquad (2)$$

The attenuations of the individual filters were measured and calibrated separately. The overall attenuation – which was typically of the order of 50 to 70 dB – was obtained by using a stack of filters. Several sets of filters of different makes and types (both absorption and reflection) were used in order to assure that the results do not depend on a particular set where interference-type interactions between the elements might change the overall attenuation. All the calibrators of the beam-splitter and filters were done at the actual laser wavelength. Also, in order to prevent errors due to undetected problems in certain devices, the measurements were repeated for two photomultiplier tubes, two laser diodes, and the calibrated photodiode response was compared to that of another calibrated photodiode. Finally, in order to virtually eliminate the effect of P_d, the system operated in the region where $P_d \sim 1$. This corresponds to the experimental condition of setting the gain of the PMT as high as possible ($\leq 10^7$), while, at the same time, reducing the counter threshold as much as possible but still without having significant contribution of thermal-Gaussian noise.

Before concluding this section we want to comment on the strength of the optical signal used. In order to make intrinsic noise contributions insignificant and to increase the quality of the average estimates, the number of signal photons was made much larger than the number of dark counts. The upper limit on the signal strength was set by PMT reliability considerations (the absolute maximum rating is about $6 \cdot 10^5$ photoelectrons/second) and by the need to minimize the probability that two detection events partly or totally overlap so they are counted as one event. Typical values were around $10^4 - 5 \cdot 10^4$ photoelectrons/second.

The experimental calibration measurements were in accordance with the calculated results. The experimental error is about 10 - 15%, and is due mainly to inaccuracies of the measurements of the optical attenuation and the estimate on the quantum efficiency, especially due to its dependence on temperature.

III. Single Pulse Detection Statistics

This section describes the experiment of measuring the probability P_{ds} of correctly detecting the presence of incident laser light during a time slot. The basic experimental setup is

shown in Fig. 3. The laser driver is a HP8003A pulse generator. The optical system is the same as that shown in Fig. 2. The photomultiplier tube is an RCA model C31034A, which has a GaAs photocathode and is the best commercially available PMT for the 0.85-μm region of the spectrum. The preamplifier is an ORTEC model 9301 (gain ≈ 10, beandwidth ≈ 150 MHz), and the final amplifier is one specifically designed and built around a μA733 video amplifier. It can operate in different gains and bandwidths, but the nominal values used in the experiment were a gain of 80 and a bandwidth of 80 MHz. Its RMS input noise over a 100-MHz bandwidth is 30 μV, and although it has a smaller bandwidth than the HP461 amplifier used in phase 1 of the experiment (Ref. 1), its overall performance is better because of its higher dynamic range and better saturation characteristics. In some experiments we also used a Comlinear model CLC-102 video amplifier, which has a 250-MHz bandwidth, and the results obtained were similar. The delay unit used was a HP8013A pulse generator. It is needed to synchronize the time-slot clock with the received light pulse. The computing controller used for averaging the counts over long periods was a HP9845C. The detector unit used was one specifically designed and built for this experiment. In Ref. 1 the performance of a detector which is based on an integrate-and-dump procedure was suggested and analyzed. Because of two reasons we did not employ this detector structure in our experiment. First, it is difficult to realize integrate-and-dump circuitry at the needed speeds. Secondly, and more important, the integrate-and-dump scheme is not the optimum detection method, since typically the signal is present only over a small fraction of the time slot, while the noise is integrated over the entire slot. The actual detector circuit used in our experiment employed hard-decision in each time-slot, and it produced reasonable results. The subject of the optimum detection scheme for this type of received signal is still an open issue and is under investigation.

In the experiment the laser diode was pulsed in a duty cycle of 1/256, with pulses of 100-nS duration (i.e., 39,062 pulses per second). Figure 4 shows several examples of the amplified PMT output under illumination intensity level corresponding to approximately one photoelectron per 100-nS time slot. The different signals occurring in each case are due to the fact that they are sample functions of the generating Poisson process. The experimental value of P_{ds} was determined by dividing the average number of time slots per second where a signal was detected by 39,062.

In Fig. 5, the probability P_{ds} of detecting the incident light is shown as a function of the average intensity of light measured in detected photons (i.e., photoelectrons) per slot. The parameter on the curves in this figure is the PMT gain. The threshold value of the detector was set at 80 mV, which is just above the value where thermal noise becomes significant. We see that for

N_s = 3.2 detected photons/slot, which corresponds to 2.5 bits/detected photon when using 8-bit PPM, we can obtain detection probabilities P_{ds} exceeding 90%. The experimental results are upper bounded by

$$P_{ds} \approx 1 - e^{-N_s} \qquad (3)$$

which is the result for the ideal counter.

Figure 5 should be compared with the receiver operating curves of Ref. 1. Since an erroneous noise variance was used in Ref. 1, we are using for comparison the corrected results of Ref. 4. In particular, Fig. 7 of Ref. 4, combined with our experimental results from Fig. 5, is shown in Fig. 6. It is interesting to note that the experimental results – using the hard decision detector – are very similar to the theoretical results – using the integrate-and-dump scheme. The subject of the theoretical analysis of the hard decision detector is under current investigation.

IV. Uncoded PPM Performance

This part of the experiment constitutes the third phase of the experimental program. The experimental setup is shown in Fig. 7. The synthesizer used as the master clock is a SYNTEST model SI-102. The frequency was 39,062 Hz, which corresponds to 100-ns time slots in a 256-slots/word PPM configuration. The PPM modulator/demodulator is an instrument designed and built specifically for the 2.5-bits/detected photon program, and its functions and performance are the subject of a separate report (Ref. 5). Since it contains almost all the necessary performance diagnostics, no additional equipment (except for the "AND" gate) was needed for the error rate measurements. These diagnostics include indications of PPM word errors as well as indications of the number of slots detected during each word period (i.e., 0 (erasure), 1 (single), 2 (double) or >2 (overflow)). The "AND" gate is needed in order to synchronize these indications with the demodulator "strobe" output. The remainder of the equipment used in this setup was described in the last two sections.

In the experiment the laser diode was pulsed (1/256 duty-factor; 100-ns slot time), and the number of the various events (errors, erasures, etc.) was counted. We found that in our case we are limited by erasures, which outnumber errors by more than one order of magnitude. The validity of this condition needs to be checked in any new situation (e.g., space-based receivers), since the performance of the Reed-Solomon decoder degrades when the ratio of erasures to errors decreases. It should be noted that only the information was transmitted optically, the synchronization signals were hard-wired between the modulator and the demodulator.

The results of the measurements are shown in Fig. 8, which depicts the bit error probability P_b as a function of the average number of detected photons per slot N_s, with the PMT gain as a parameter. Also shown on the graph is the lower bound

$$P_b = \frac{256}{510} e^{-N_s} \sim \frac{1}{2} e^{-N_s} \qquad (4)$$

which is the expression of the ideal photon counter. We see that the experiment results are not too far from this bound.

Figure 9 compares the three results, i.e., the ideal photon counter, experimental results, and the integrate-and-dump detector theoretical results from Ref. 1 (but with the correct noise variance), under two PMT gains: 10^6 and $3 \cdot 10^6$. For $G = 10^6$ (Fig. 9a), the experimental hard-decision results are somewhat better than the theoretical integrate-and-dump result. As we increase the gain ($G = 3 \cdot 10^6$, Fig. 9b), both results move closer to the ideal counter, with the (theoretical) integrate-and-dump results somewhat better than the (experimental) hard-decision results.

In order to predict the coded channel performance on the basis of our measurements, we need the relation between the coded and uncoded performance. This relation is shown in Fig. 10. The various curves associated with a given code correspond to different combinations of errors and erasures.

Comparing the results of Figs. 8 and 10, we see that the needed energy efficiency of 2.5 bits/detected photon can be easily achieved with the rate 3/4 code at the desired bit error probability of $5 \cdot 10^{-3}$. From these graphs it also seems that if the ratio of erasures to errors is not too low, operation at 2.5 bits/detected photon can be achieved even with the rate 7/8 code which has less complexity.

V. Conclusions

The uncoded performance of a laboratory optical channel has been demonstrated and evaluated. The results agree with the previously generated theoretical analysis, thus substantially increasing the confidence that the energy-efficient operation of 2.5 bits/detected photon will be achieved using a rate 3/4 Reed-Solomon code, as well as possibly with a less complex rate 7/8 code.

Acknowledgment

References

1. Lesh, J. R., Katz, J., Tan, H. H., and Zwillinger, D., "2.5 Bits/Detected Photon Demonstration Program Description, Analysis, and Phase I Results," *TDA Progress Report 42-66*, pp. 115-132, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1981

2. Pierce, J. R., "Optical Channels: Practical Limits with Photon Counting," *IEEE Trans. Commun.*, COM-26, No. 12, pp. 1819-1821, Dec. 1978.

3. Katz, J., "High Power Semiconductor Lasers for Deep Space Communications," *TDA Progress Report 42-63*, pp. 40-50, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1981

4. Tan, H. H., "A Statistical Model of the Photomultiplier Gain Process with Applications to Optical Pulse Detection," *TDA Progress Report 42-68*, pp. 55-67, Jet Propulsion Laboratory, Pasadena, Calif., April 15, 1982

5. Marshall, W. K., "A PPM Modulator and Demodulator for the 2.5 Bit/Detected Photon Demonstration," *TDA Progress Report 42-68*, pp. 50-54, Jet Propulsion Laboratory, Pasadena, Calif., April 15, 1982
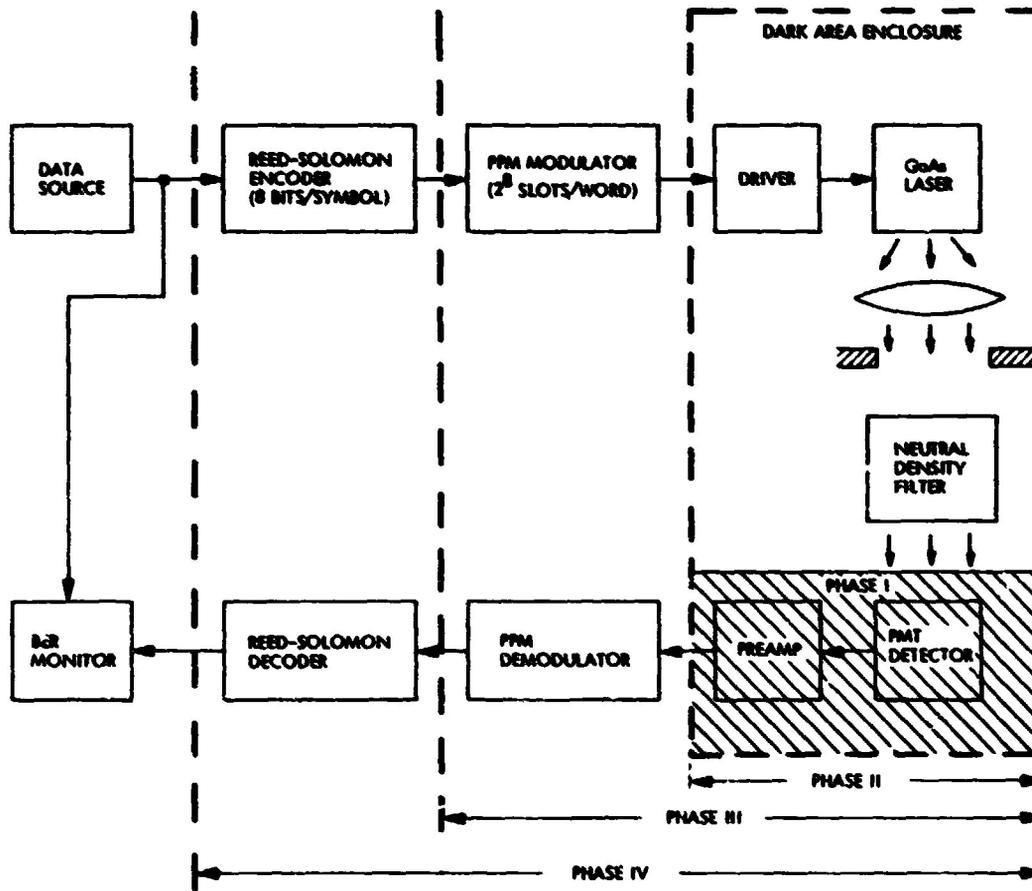
Fig. 1. Block diagram of 2.5 bit/detected photon demonstration system with the various experimental phases
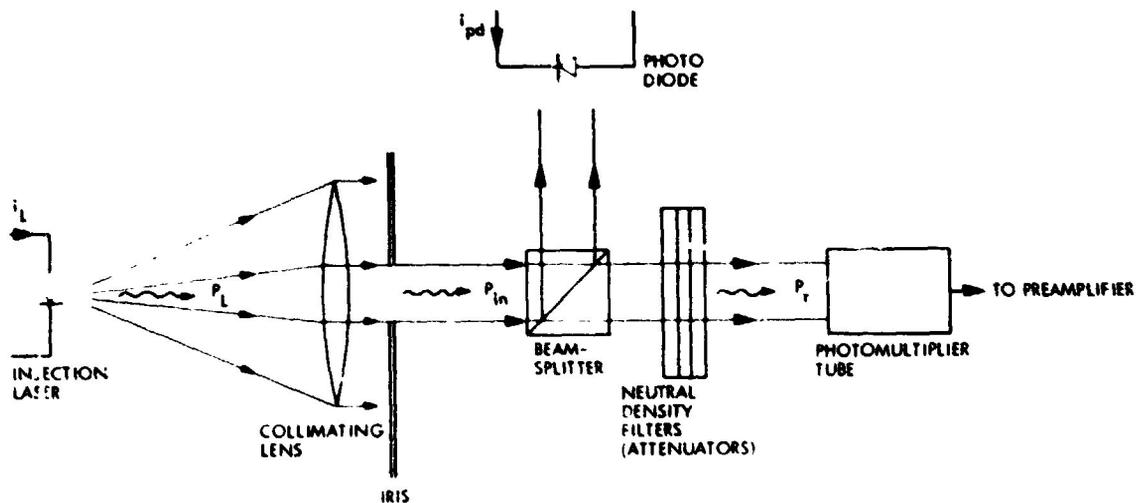
Fig. 2. Schematic view of the electrooptical components in the experiment
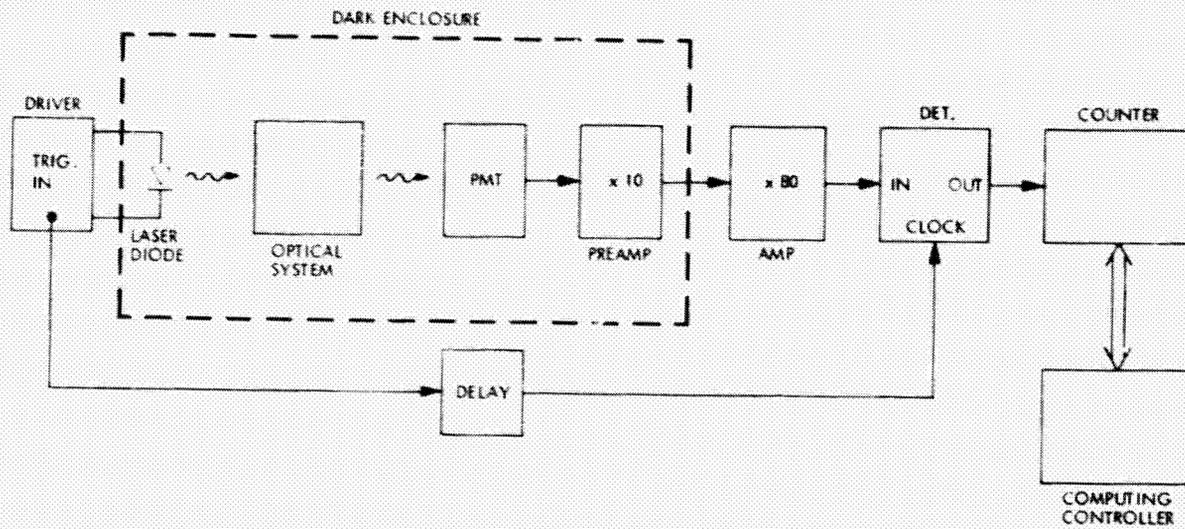
Fig. 3. Block diagram of experimental setup for measurement of optical pulse detection probability P_{ds}
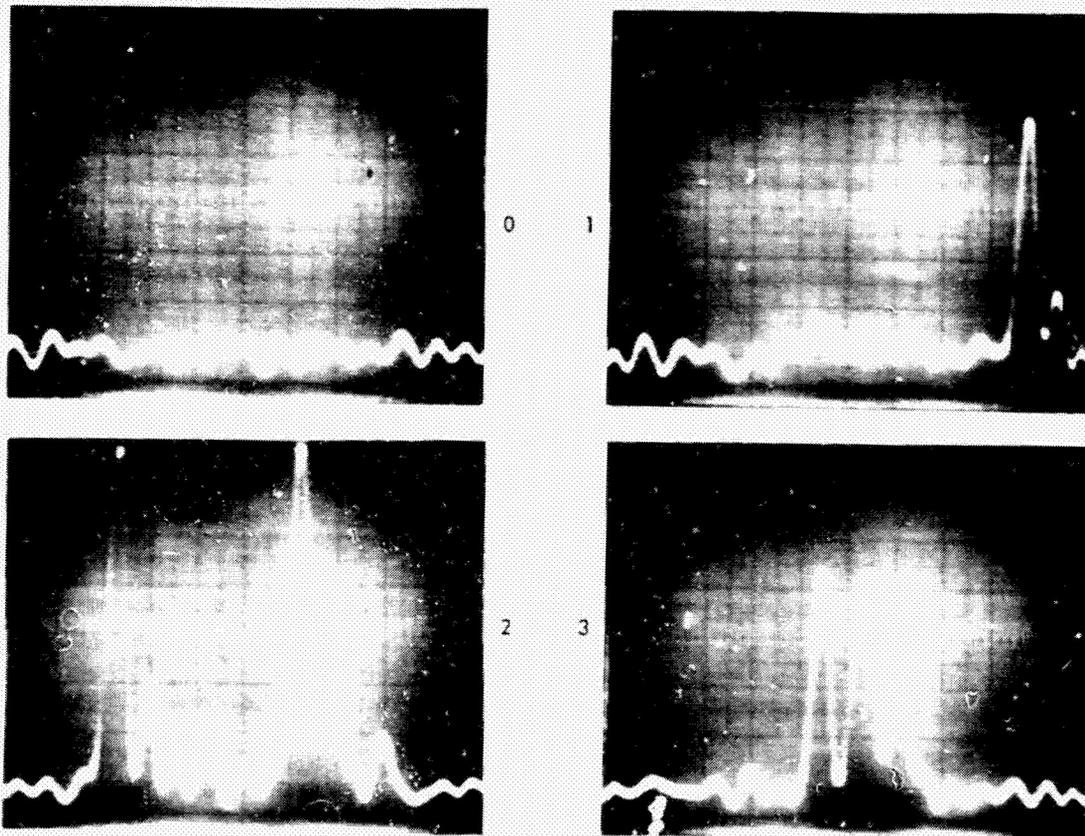
Fig. 4 Examples of various amplified PMT output oscilloscope traces during incidence of a light pulse with $\bar{N}_s = 1$ detected photon/slot (time slot is 100 ns). Shown are traces with 0, 1, 2 and 3 photoelectron events.
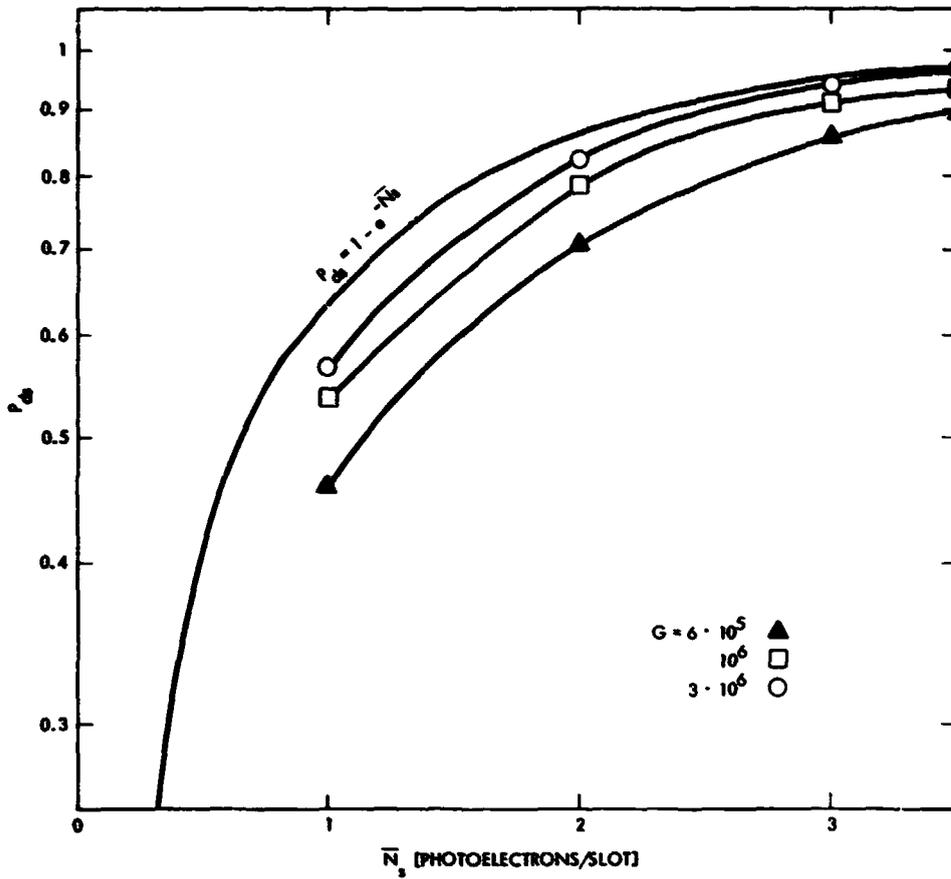
Fig. 5. Measured optical pulse detection probability P_{ds} as a function of $\bar{N}_s$ (average number of photoelectrons per pulse) for several values of PMT gain. Also shown is the ideal upper bound $P_{ds} = 1 - e^{-\bar{N}_s}$
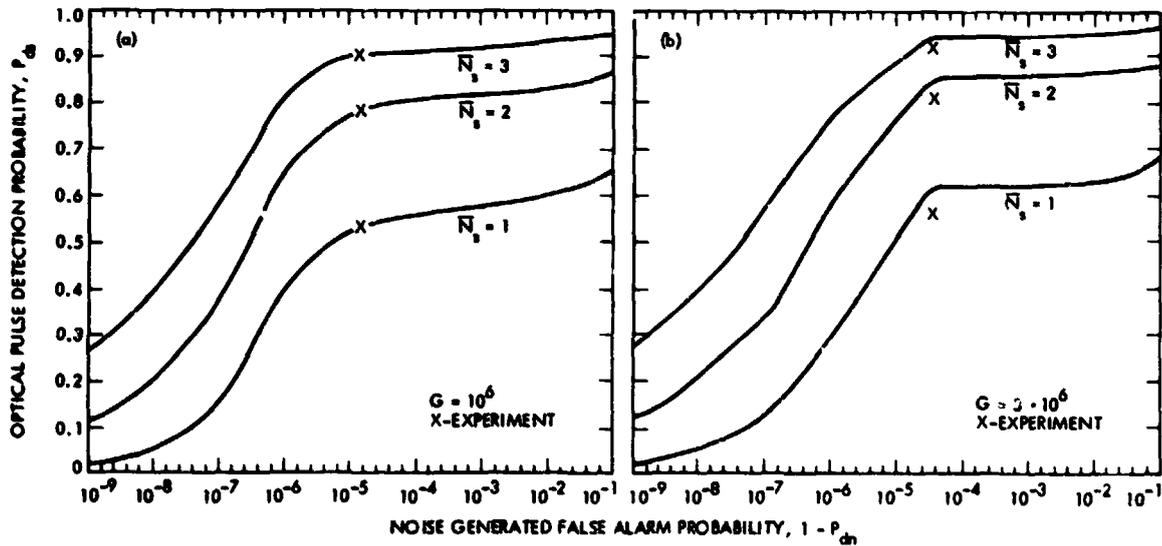
Fig. 6. Binary communication receiver operating curve as a function of $\bar{N}_s$: comparison between theory and experiment
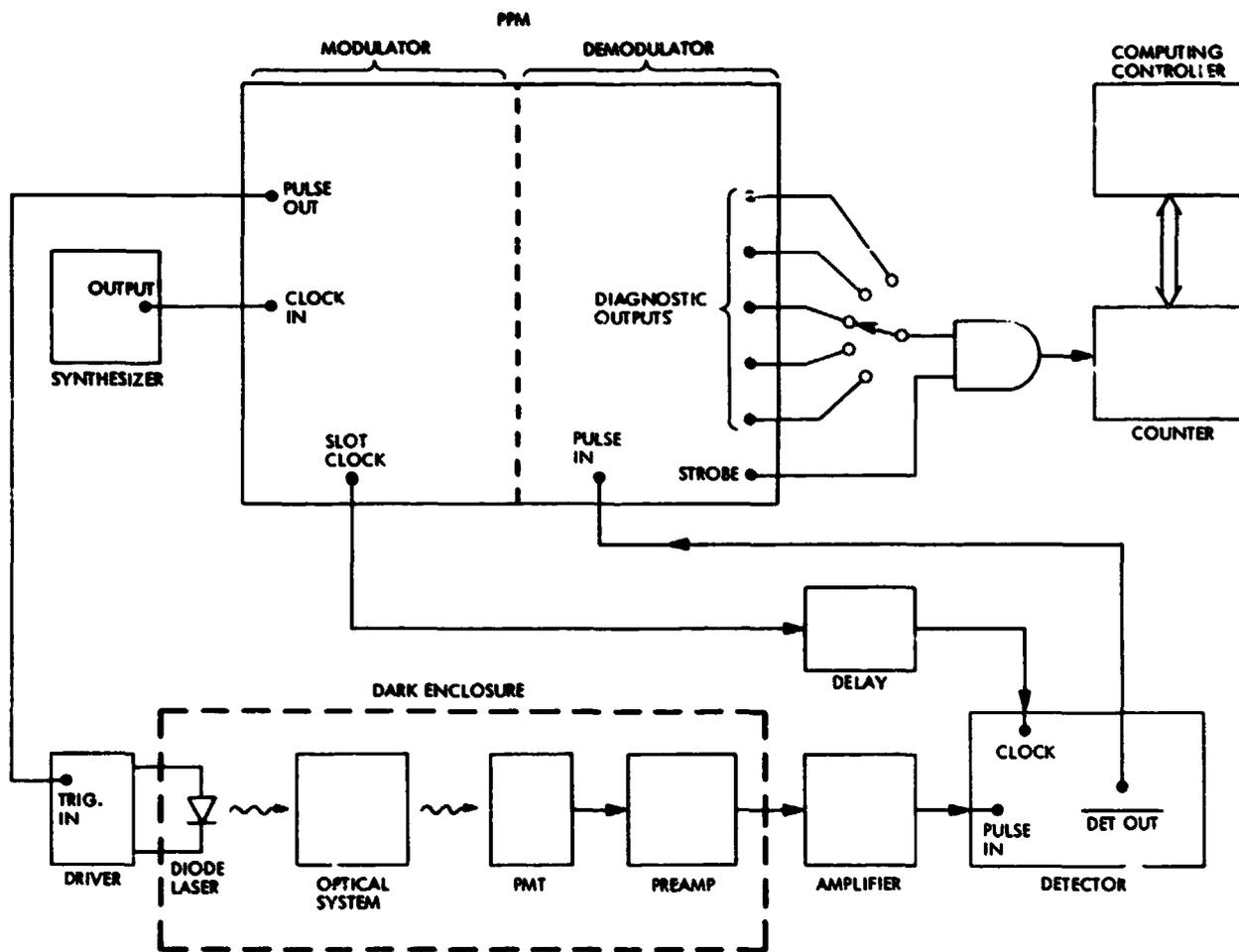
Fig. 7. Block diagram of experimental setup for measuring the uncoded, PPM-modulated error performance of the optical channel
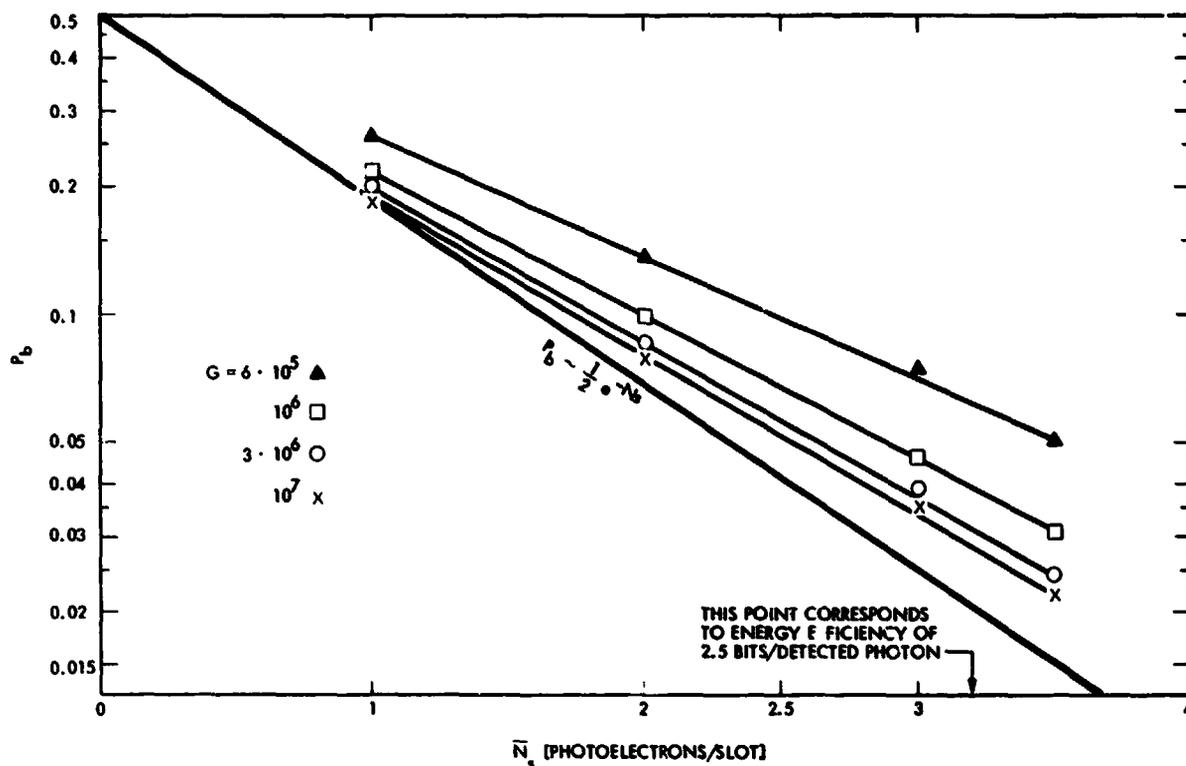
Fig. 8. Measured bit error performance for uncoded 256-ary PPM as a function of $\bar{N}_s$, the average number of detected photons/slot, for several values of PMT gain. $\bar{N}_s = 3.2$ corresponds to operation with energy efficiency of 2.5 bits/detected photon
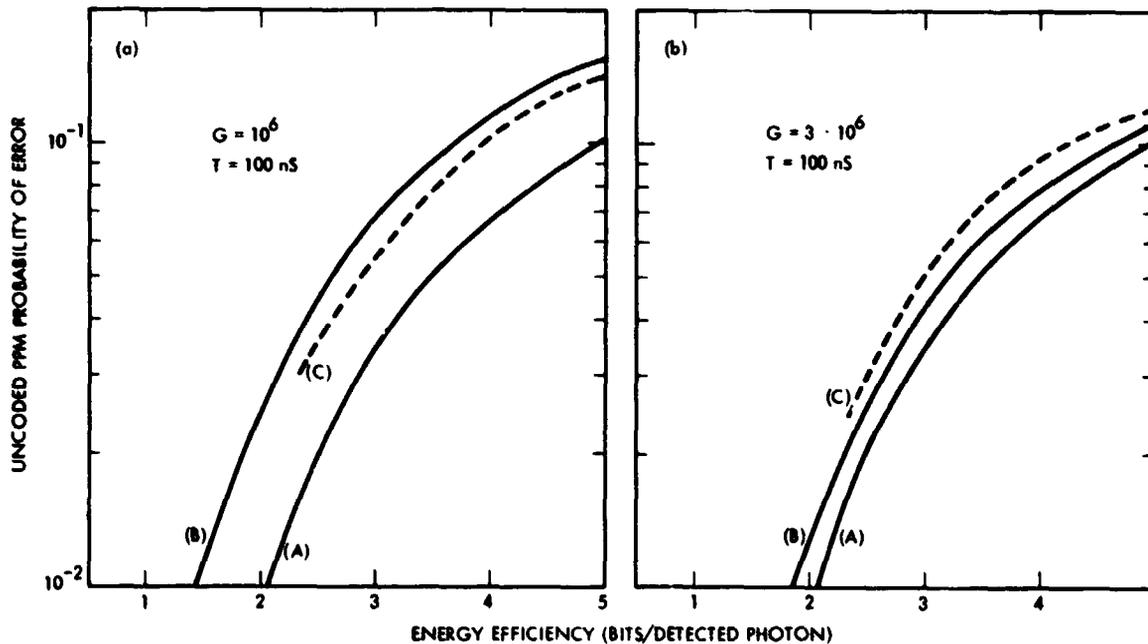
Fig. 9. Uncoded PPM bit error probability vs energy throughput efficiency (in bits/detected photon) for (a) PMT gain = 10^6 and (b) PMT gain = $3.3 \cdot 10^6$. In every graph, (A) indicates the ideal counter, (B) the integrate-and-dump theoretical performance, and (C) the experimental results using hard decision, as taken from Fig. 8
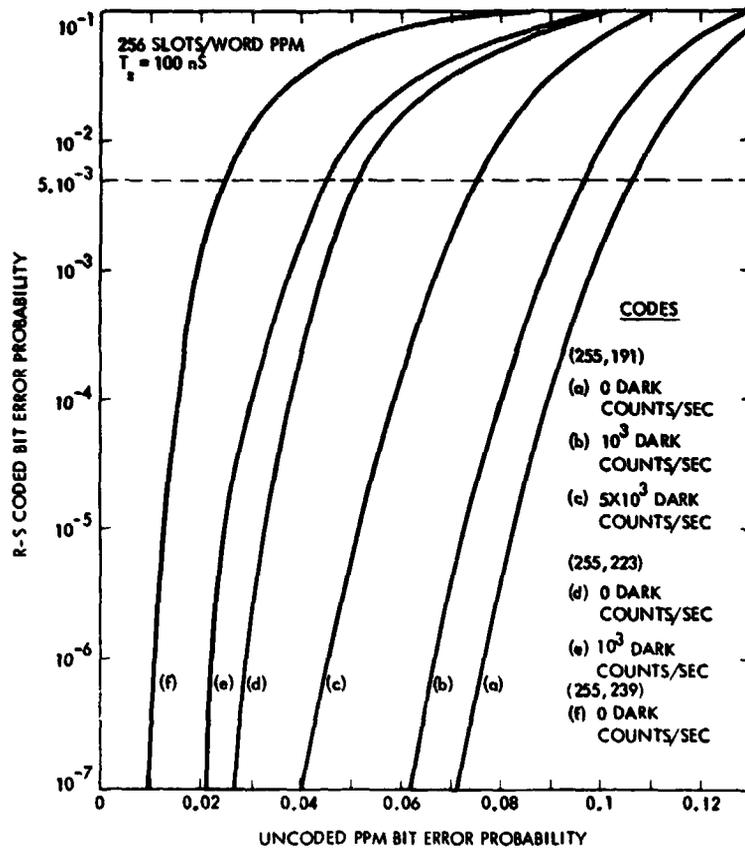
Fig. 10. RS coded bit error rate as a function of the uncoded PPM bit error rate for (255, 223) and (255, 191) codes and for the extreme combinations of errors and erasures

A Note on Deep Space Optical Communication Link Parameters

S. J. Dolinar and J. H. Yuen
Communications Systems Research Section

Communication link analysis at the optical frequencies differs significantly from that at microwave frequencies such as the traditional S- and X-ba·· ₃ used in deep space applications, due to the drastically different technology of transmitter, antenna, modulators, receivers, etc. In addition, the important role that quantum noise plays in limiting system performance is quite different than that of thermal noise. In this paper optical communication is discussed in the context of a deep space communication link. The optical link design is put in a design control table format similar to a microwave telecom link design. Key considerations unique to the optical link are briefly discussed.

I. Introduction

We have made a preliminary attempt to present the parameters influencing the performance of a deep space optical communication system in the same "design control table" format as that used in the design of microwave systems. This form of presentation facilitates comparison between the two types of systems. However, the optical and microwave systems are not completely analogous, and thus the presentation must be issued with several caveats to prevent misunderstanding.

The free space optical link differs from familiar microwave links in that its performance is limited by intrinsic quantum mechanical measurement uncertainty (loosely termed "quantum noise") and, occasionally, by background light levels, rather than by receiver thermal noise. The quantum noise contributes generally non-Gaussian statistics, and consequently analyses of the optical and microwave links are quite different. Performance in the case of a Gaussian noise-limited microwave link is completely summarized by a signal-to-noise ratio E_s/N_0, where E_s is the received signal

energy per bit and N_0 is the (single-sided) noise spectral density level. Unfortunately, for a quantum noise-limited optical link, there is no comparably handy ratio that fully characterizes performance.

It is nonetheless convenient to go ahead and normalize the received optical energy relative to a reasonable measure of the quantum noise level. In the optical communications literature, it is standard to normalize optical signal energies to units of photons. This can be loosely interpreted as a signal-to-noise ratio, to the extent that the "amount" of quantum noise is roughly indicated by the energy $h\nu$ of a single photon (h = Planck's constant, ν = optical frequency). It must be remembered, however, that this ratio is not sufficient by itself to specify performance, even when only quantum noise is present. Performance of the optical system depends in a complicated way on the number of detected signal photons and background photons, and also on the kinds of signal modulation, receiver structure, and information coding that are used.

II. Sample Optical Design Control Table

A nominal design control table (DCT) for a sample deep space optical link is given in Table 1. The sample link consists of a free space downlink from the vicinity of Jupiter at 5 AU to an Earth-orbiting relay station. Parameter values appearing in the table are largely drawn from Refs. 1 and 2, which analyzed the optical deep space link in some detail.

A brief overview of the sample optical link DCT is helpful. The first seven entries calculate the detected power at the receiver due to the transmitted signal. The assumed values for transmitter power, antenna gains, and receiver losses correspond to similar assumptions in Refs. 1 and 2, and they represent current or foreseeable technological capabilities. Entries 8-10 in the table estimate the net detected power at the receiver due to typical sources of background light. In this example, the receiver's field of view is assumed to take in light from either a typical point source (weak star, magnitude +6) or a typical distributed source (Jupiter at opposition). Background sources as strong as these may or may not be present in an actual application; stronger sources (e.g., bright stars, sun, skylight for ground-based receivers) might also cause problems in certain configurations. Entries 11-13 normalize the signal and background power relative to the bit rate. Entries 14-16 further normalize these bit energies to units of photons. The last two entries calculate the link performance and margin for the assumed modulation, coding, and detection schemes.

A brief annotation of each of the individual entries in the sample DCT follows:

(1) The assumed transmitter power value of 1 watt refers to the total power broadcast from the transmitting antenna; i.e., it includes internal transmitter inefficiencies as well as losses in coupling the transmitter to the antenna.

(2) The transmitting antenna gain is computed as $4\pi A_t/\lambda^2$, where the transmitter wavelength λ is taken as 1 micron and the effective transmitting area A_t is taken as $1/4\,\text{m}^2$. This value of A_t requires 56-cm-diameter optics if diffraction limited.

(3) The 2-dB pointing loss was computed for $1/2$-μrad rms error from curves in Section 2.7 of Ref. 2. The $1/2$-μrad rms error level corresponds to approximately $1/4$ beam width. This level was chosen as a threshold beyond which performance degrades very rapidly, and as such it represents a stringent requirement on pointing accuracy.

(4) Space loss is determined from the formula $(4\pi R/\lambda)^2$, where the assumed range is $R = 5$ AU.

(5) The receiving antenna gain is computed as $4\pi A_r/\lambda^2$, where the receiving area A_r is taken to be $10\,\text{m}^2$. This corresponds to 3.6-m-diameter receiving optics, not necessarily diffraction-limited.

(6) Total losses at the receiving end are listed as 8 dB. Three contributions to the figure are itemized separately. The atmospheric loss entry of 0 dB is included just to illustrate one of the advantages of a deep space relay link as compared to familiar direct links to Earth. The -1 dB receiver transmission loss and -7 dB detector quantum efficiency correspond to factors $\zeta_r = 0.8$ and $\eta_r = 0.2$ used in Refs. 1 and 2. The factor ζ_r accounts for receiving optical losses, and the factor η_r refers to the probability of detecting individual photons at the receiver.

(7) The net detected signal entry is simply the sum (in dB) of entries 1 through 6.

(8) The background intensity of -97 dBm is taken from Fig. 1-4 and Eq. (1-4) of Ref. 2, assuming wavelength $\lambda = 1$ μm, optical predetection bandwidth $\Delta\lambda = 10$Å, and receiving area $A_r = 10\,\text{m}^2$, for either of two cases:

(a) weak star, magnitude +6, or

(b) Jupiter at opposition, as seen with receiver field of view $\theta_r = 2$ μrad.

The assumed field of view (for the distributed source case) is taken to be the same as the transmitted beamwidth; it does not require diffraction limited receiving optics.

(9) The same losses at the receiving end apply to both signal and background power, and therefore entry 6 is repeated here.

(10) The net detected background power entry is the sum (in dB) of entries 8 and 9.

(11) The assumed bit rate of 1 Mbps is approximately 9 times the capability of the Voyager system from Jupiter.

(12) Detected signal energy per bit E_s is obtained by dividing detected signal power by bit rate.

(13) Detected background energy per bit E_b is obtained by dividing detected background power by bit rate.

(14) 'Quantum noise' energy is measured by $h\nu$, as discussed above.

(15) The "signal-to-quantum noise ratio" $E_s/h\nu$ is obtained from entries 12 and 14. In the optical literature it is conventional to use the photon information rate $\rho = (E_s/h\nu)^{-1}$ rather than $E_s/h\nu$.

(16) The "background-to-quantum noise ratio" $E_b/h\nu$ is obtained from entries 13 and 14.

(17) Required $E_s/h\nu$ represents the net effect of many different system paramete .. The calculation here assumes uncoded 64-ary PPM modulation and a direct detection receiver. A value of required $E_s/h\nu = 1$ (0 dB) to achieve a bit error rate of 5×10^{-3} is listec in the table. Additional performance results are discussed in the next section.

(18) The nominal link margin of 3 dB is obtained from entries 15 and 17.

III. Key Uncertainty Areas and Tradeoff Considerations

Table 1 demo: "rates the potential feasibility of communicating over a 5-AU free space optical link at a rate of 1 Mbps, assuming the parameter values listed. We have attempted to choose values which ar: not overly optimistic or conservative for near-future optical systems. However, because of the relative immaturity of optical technology, these numbers are stated with much less certainty than the corresponding parameters in a microwave system.

There are several key areas of uncertainty concerning parameters which directly affect the amount of signal power obtained at the receiver:

(1) The assumed transmitted power of 1 watt is beyond current technological capabilities, and further development of efficient, high-power, narrow-beam optical sources is needed. Advances in optimizing the power efficiency of semiconductor injection lasers (Ref. 3) and in phase locking laser arra' 'o produce a strong coherent source (Ref. 4) are currently underway.

(2) The assumed optical antenna dimensions are modest compared to those of corresponding microwave antennas or of Earth-based telescopes, but the technology of low weight, spaceborne optical antennas is still in its infancy. Improvements are expected, with the experience gained from such projects as the Infrared Astronomy Satellite (IRAS) (Ref. 5).

(3) Very precise pointing and tracking systems need to be developed. To keep pointing loss reasonably low, pointing errors must be limited to submi, radian levels. The nominal 2-dB loss assumed in the table could be increased radically if this level of accuracy is not obtainable.

(4) Required $E_s/h\nu$ depends on many different system parameters, including the desired bit error rate, the amount of background noise, and the kinds of signal modulation, receiver structure, and information coding that are used. The 0-dB value assumed in the table corresponds to a photon information rate of 1 bit/photon. This v Jue may be raised or lowered significantly if changes are made in the system parameters. For example, eliminating the assumed background noise entirely would ... reo 'ired $E_s/h\nu$ to -1 dB, whereas higher bac!·2·· d ..cls might raise t' ; required $E_s/h\nu$ intolerably. A tighter error tolerance would require higher $E_s/h\nu$, for instance, $E_s/h\nu = 4$ dB for a bit error rate (BER) of 10^{-6} in the absence of background. The requirement at this BER could be drastically reduced via coding (e.g., to a required $E_s/h\nu = -3$ dB with a (63, 32) Reed-Solomon code) or by using a larger number of PPM slots (e.g., required $E_s/h\nu = 1$ dB for $M = 4096$). Ultimate capacity of the quantum limited PPM/direct detection channel is unbounded, and thus in principle the required $E_s/h\nu$ may 'e made arbitrarily small at any BER but practical limits on coding complexity and on laser peak power levels[1] generally restrict these gains to a few dB relative to Table 1. Presently, a laboratory effort (Ref. 6) is in progress to demonstrate the feasibility of communicating at 2.5 bits/photon with currently available devices. Heterodyne and homodyne receiver structures applied to the quantum limited channel have finite capacities of 1 nat/photon and 2 nats/photon, respectively (corresponding to finite lower bounds on $E_s/h\nu$ of -1.6 dB and -4.6 dB), but these structures may be preferable to direct detection in certain applications.

The following table illustrates some of the tradeoff issues involved in the determination of the required $E_s/h\nu$. For the purpose of this illustration, a direct detection receiver i used and background noise is assumed to be negligible. Required $E_s/h\nu$ is given as a function of the number of PPM slots (M) and the required BER for the two cases of uncoded transmission and rate $1/2$ $(M - 1, M/2)$ Reed-Solomon coding. By way of comparison, the 2.5 bits/photon $(E_s/h\nu = -4$ dB) laboratory demonstration (Ref. 6) uses $M = 256$ and a rate 3/4 (255, 191) Reed-Solomon code.

---

[1]At a fixed average power level (e.g., 1 watt in Table 1), the peak power required of the transmitting laser increases directly with M, the number of PPM slots.

References

1. Vilnrotter, V. A. and Gagliardi, R. M., "Optical-Communication Systems for Deep Space Applications," Publication 80-7, Jet Propulsion Laboratory, Pasadena, Calif., Mar. 15, 1980.

2. Gagliardi, R. M., Vilnrotter, V. A. and Dolinar, S. J., "Optical Deep Space Communication via Relay Satellite," Publication 81-40, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1981.

3. Katz, J., "Power Efficiency of Semiconductor Injection Lasers," *TDA Progress Report 42-66*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 94-100, Dec. 15, 1981.

4. Katz, J., "Phase-locking of Semiconductor Injection Lasers," *TDA Progress Report 42-66*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 101-114, Dec. 15, 1981.

5. "Infrared Astraonomicai Satellite Mission (IRAS) Joint Project Requirements," National Aeronautics and Space Administration, Publication No. 623-6 Rev. A, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, Calif., Mar. 1981.

6. Lesh, J. R. et al., "2.5-Bit/Detected Photon Demonstration Program: Description, Analysis, and Phase I Results," *TDA Progress Report 42 66*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 115-132, Dec. 15, 1981.

Table 1. Jupiter to earth-orbiting relay optical link

1.	Transmitted power (1 watt)		30 dBm
2.	Transmitting antenna gain (A_t = 1/4 m², λ = 1 μm)		125 dB
3.	Pointing loss (1/2 μrad rms error)		-2 dB
4	Space loss (R = 5 AU, λ = 1 μm)		-380 dB
5	Receiving antenna gain (A_r - 10 m², λ = 1 μm)		141 dB
6.	Losses at receiving end		-8 dB
	Atmospheric loss	-0 dB	
	Receiver transmission loss	-1 dB	
	Detector efficiency	-7 dB	
7	Net detected signal power		-94 dBm
8.	Background intensity (Jupiter at opposition or weak star, 10A bandwidth, A_r = 10 m², $θ_r$ = 2 μrad)		-97 dB
9	Losses at receiving end		-8 dB
10	Net detected background power		-105 dBm
11	Bit rate (1 Mbps)		60 dB Hz
12	Detected signal energy/bit (E_s)		-154 dB mJ
13	Detected background energy/bit (E_b)		-165 dB mJ
14	"Quantum noise" (hν)		-157 dB mJ
15	$E_s/hν(=1/\rho)$		3 dB
16	$E_b/hν$		-8 dB
17	Required $E_s/hν\ (=1/\rho)$		0 dB
18	Margin		3 dB

Table 2. Required $E_s/hν$ for quantum limited direct detection of PPM signals

Required $E_s/hν$ (in dB), No Coding					Required $E_s/hν$ (in dB), (M - 1, M/2) R-S coding				
M \ Bt R	16	32	64	256	M \ Bt R	16	32	64	256
10^{-7}	5.9	4.9	4.1	2.8	10^{-7}	1.3	-1.1	-3.0	-5.8
10^{-6}	5.2	4.2	3.4	2.1	10^{-6}	0.9	-1.4	-3.3	-6.0
10^{-5}	4.3	3.4	2.6	1.3	10^{-5}	0.3	-1.9	-3.6	-6.1
10^{-4}	3.3	2.3	1.5	0.3	10^{-4}	0.4	-2.4	-4.0	-6.4
10^{-3}	1.9	0.9	0.2	-1.1	10^{-3}	-1.2	3.0	-4.5	-6.6
10^{-2}	0.1	1.1	1.9	3.1	10^{-2}	-2.3	3.9	-5.1	-7.0

N82 32541

Thermal Analysis of Antenna Backup Structure
Part I — Methodology Development

D. Schonfeld and F. L. Lansing

DSN Engineering Section

An analytic method is devised for predicting the temperature distribution in typical antenna structural back-up members. The results are in good agreement with those obtained by a numerical shooting method. The analytic method developed in this work has shown a good potential in greatly simplifying the thermal analysis process for complex back-up antenna structures.

I. Introduction

Future requirements for dee, space navigation and telemetry link communications point to the need for higher frequency bands. At these higher frequencies, data rates are better but antenna surface accuracy requirements are more severe than those of the presently used S- or X-bands. Guided by the accuracy needs, studies are now in progress on the feasibility of constructing large Ka-band antennas which will use frequencies at 32 GHz or higher (Ref. 1).

The construction requirements for these Ka-band antennas are much more demanding than those for presently used devices. For example, the reflective surfaces must be positioned and aligned much more accurately and must be maintained during the life of the antenna against varying gravity, wind, and thermal loadings. Environmental changes such as temperature variations can have deleterious effects on the antenna performance. Field thermal measurements (Ref. 2) and analytic investigations are being conducted to study the environmental effects on the structural members of the large 64-m antenna. In the analytical investigation, which is the subject of this article, we are interested in simulating the temperature distribution throughout the complex back-up

structural members given their physical properties, geometric arrangement, and environmental conditions, such as air temperature, wind velocity, and solar irradiation (insolation).

The simulation of the temperature pattern of a structural member can be done by "conventional" numerical finite difference methods in heat transfer which divide the member under consideration into nodes and then apply heat balance equations to each node. Rather than solving for an excessively large network of nodes for a complex antenna back-up structure, a new method is developed in this paper to save effort and time. The method is analytical in nature and relies on deriving a universal relationship for the temperature variations and heat fluxes within each member. The methodology is described only in this article and will be followed by additional applications in subsequent TDA reports.

II. Methodology Development

Consider a simple bar, of length L, as sketched in Fig. 1 subjected to solar radiation, conduction, convection, and radiation heat exchange with the ambient air. The cross section of the bar, A, and its material properties are assumed

constant along the bar's length. For an element of incremental length dx along the bar axis, whose temperature is $T(X)$, the convection losses are given by:

$$dq_c = h P dx (T - T_a)$$ (1)

the radiative losses by:

$$dq_r = \epsilon \sigma \mathscr{F} P dx (T^4 - T_a^4)$$ (2)

the absorbed portion of the solar irradiation by:

$$dq_i = \alpha I P dx$$ (3)

and the change in conductive heat transfer stored in the element by:

$$dq_k = \frac{d}{dx} \left(kA \frac{dT}{dx} \right) dx$$ (4)

The complete heat balance equation at steady-state conditions can be written as:

$$\frac{d^2 T}{dx^2} = \frac{hP}{kA} (T - T_a) + \frac{\epsilon \sigma P \mathscr{F}}{kA} (T^4 - T_a^4) - \frac{\alpha I P}{kA}$$ (5)

Equation (5), together with the boundary conditions

$$\text{at} \quad x = 0 \quad , \quad T = T(0)$$ (6)

$$\text{at} \quad x = L \quad , \quad T = T(L)$$ (7)

forms a nonlinear boundary value problem for $T(X)$. In general, such problems cannot be solved analytically and numerical methods must be employed (Refs. 3 and 4). Several of these methods are presented in the paragraphs which follow.

A. The "Shooting" Method

An outline of the "shooting" method and its use is given in Ref. 4. Using the subroutine described in that reference, Eq. (5) can be written as a system of equations in the form:

$$Y_1' = Y_2$$ (8)

$$Y_2' = C_1 Y_1^4 + C_2 Y_1 + C_3$$ (9)

where Y_1 stands for the temperature, T, while Y_2 and Y_2' for dT/dx and $d^2 T/dx^2$ respectively, and the superscript (') for d/dx. The constants C_1, C_2 and C_3 are defined as:

$$C_1 \equiv \frac{\epsilon \sigma P \mathscr{F}}{kA}$$ (10)

$$C_2 \equiv \frac{hP}{kA}$$ (11)

$$C_3 \equiv -\frac{hP}{kA} T_a - \frac{\epsilon \sigma P \mathscr{F}}{kA} T_a^4 - \frac{\alpha I P}{kA}$$ (12)

This numerical method will serve as a useful check on the analytical results to be developed later.

B. Perturbation Method

This analytic method of solving Eq. (5) using perturbation requires that the equation must become linear in the limit of some selected small parameter. The parameter C_1 in Eq. (10) is small because of the Stefan-Boltzman constant σ of the order 10^{-10}, while C_2 and C_3 from Eqs. (11) and (12) are of the order 1 and greater. The perturbation method assumes that the solution can be written in the form:

$$T(x) = T_0(x) + C_1 T_1(x) + C_1^2 T_2(x) + \cdots$$ (13)

Then, by inserting (13) into (5), one obtains a set of linear equations such as:

$$\frac{d^2 T_0}{dx^2} - C_2 T_0 - C_3 = 0$$ (14)

$$\frac{d^2 T_1}{dx^2} - C_2 T_1 - C_3 = T_0^4$$ (15)

$$\frac{d^2 T_2}{dx^2} - C_2 T_2 - C_3 = T_1^4$$ (16)

$$\vdots$$

which must be solved sequentially.

The disadvantage of applying this method for our case is that once an expression for T_0 is found by solving Eq. (14), subsequent equations will become very complicated due to the presence of terms such as T_0^4, T_1^4, etc. For this reason, the perturbation method was not examined further.

111

C. Linearization Method

Equation (5) will be reduced to a closed form solution for $T(X)$ by linearizing the radiation term in Eq. (2) by writing it as:

$$dq_r = h_r P \, dx \, (T - T_a) \tag{17}$$

By comparing Eq. (17) with Eq. (2) one obtains:

$$h_r = \epsilon \sigma \, \mathscr{F}(T^2 + T_a^2)(T + T_a) \tag{18}$$

The initial value of h_r can be obtained by making the approximation $T \cong T_a$. This assumption is valid in cases where one would not expect too much difference between the temperature of the structure and that of the ambient air. Note, however, that such an assumption cannot be used for a structure in space where the temperature of the structure is much different from the surrounding space temperature, ($T_a = 0$ K). As the temperature pattern is known, a modified value of h_r can be obtained at the average link temperature.

By using Eq. (17) in Eq. (5), one obtains the linearized differential equation:

$$\frac{d^2 T}{dx^2} = \left[\frac{(h_r + h_c)P}{kA} \right] T - \left[\frac{(h_r + h_c)P}{kA} T_a + \frac{\alpha IP}{kA} \right] \tag{19}$$

and by making use of the abbreviations

$$\psi \equiv \frac{(h_r + h_c)P}{kA} \tag{20}$$

$$\xi \equiv \frac{(h_r + h_c)P}{kA} T_a + \frac{\alpha IP}{kA} \tag{21}$$

one obtains the solution of Eq. (19) as:

$$T(x) = \frac{\left[T(L) - \frac{\xi}{\psi} \right] - \left[T(0) - \frac{\xi}{\psi} \right] \cosh \sqrt{\psi} L}{\sinh \sqrt{\psi} L} \sinh \sqrt{\psi} x$$

$$+ \left[T(0) - \frac{\xi}{\psi} \right] \cosh \sqrt{\psi} x + \frac{\xi}{\psi} \tag{22}$$

The ratio ξ/ψ from Eqs. (20) and (21) is:

$$\frac{\xi}{\psi} = T_a + \frac{\alpha I}{h_r + h_c} \tag{23}$$

Hence, the physical meaning of ξ/ψ represents a balance between the incoming insolation and the rate of radiative and convective losses, all with respect to the ambient temperature, T. Therefore, ξ/ψ is an "equilibrium" temperature which will be denoted by T_e. In terms of T_e, Eq. (22) becomes

$$T(x) = \frac{[T(L) - T_e] - [T(0) - T_e] \cosh \sqrt{\psi} L}{\sinh \sqrt{\psi} L} \sinh \sqrt{\psi} x$$

$$+ [T(0) - T_e] \cosh \sqrt{\psi} x + T_e \tag{24}$$

III. Numerical Example

To check the validity of Eq. (24) versus the more accurate shooting method, let us consider a numerical example. In this example we assume a steel bar having the following geometry:

Perimeter, $P = 0.305$ m (1 ft)

Length, $L = 0.91$ m (3 ft)

Cross sectional area, $A = 5.81 \times 10^{-3}$ m^2 (0.0625 ft^2)

The two end temperatures are kept at:

$$T(0) = 311 \text{ K} (560°\text{R})$$

$$T(L) = 322 \text{ K} (580°\text{R})$$

while the ambient temperature is:

$$T_a = 294 \text{ K} (530°\text{R})$$

For these temperatures, and assuming typical Goldstone values for wind speed [16.1 m/hr (10 mph)] and insolation (800 W/m^2), we will use the following heat transfer coefficients:

conductivity, $k = 45.0$ W/mK (26.0 BTU/hr ft°R)

radiation heat transfer coefficient $h_r = 3.1$ W/m^2 K (0.55 BTU/hr ft^{2}R)

convective heat transfer coefficient, $h = 22.7$ W/m^2 K (4.0 BTU/hr ft^2°R)

emissivity, $\epsilon = 0.5$

view factor, $\mathcal{F} = 1.0$

absorptivity, $\alpha = 0.4$

insolation, $I = 800$ W/m^2 (254 BTU/hr ft^2)

For these values, $\psi = 30.1$ m^{-2} (2.8 ft.$^{-2}$) and $T = 308$ K (552°R). The results of the shooting method and the linearization method in Eq. (24) are shown in Tables 1 and 2. A plot of the results of Table 2 is made in Fig. 2. The excellent agreement between the analytical and numerical results validates the use of Eq. (24). Note that the temperature distribution within the bar need not be monotonic; in this particular case, the temperature within the bar reaches a point where it is lower than either of the end temperatures.

IV. The Multi-link Problem

Consider now a system of 3 bars linked together as illustrated in Fig. 3. The physical and geometric properties of each bar can be different and so can their exposure to solar radiation. The temperatures at the junctions are assumed to be unknown. The problem is to find a simple way of determining these temperatures and then, by using Eq. (24), to obtain the temperature distribution throughout the links.

The solution relies on the fact that, at each junction of two or more members, the heat flux balance at steady state is written as:

$$\Sigma q_{i,j,k, \ldots} = 0$$

Since the only heat transfer at the junction points is through conduction, for the case we are considering, the following equations are valid:

$$q_{1,2} = \left[-k_1 A_1 \left(\frac{dT}{dx} \right)_0 \right]_1 - \left[k_2 A_2 \left(\frac{dT}{dx} \right)_L \right]_2 = 0 \quad (25)$$

$$q_{2,3} = \left[-k_2 A_2 \left(\frac{dT}{dx} \right)_0 \right]_2 - \left[k_3 A_3 \left(\frac{dT}{dx} \right)_L \right]_3 = 0 \quad (26)$$

$$q_{3,1} = \left[-k_3 A_3 \left(\frac{dT}{dx} \right)_0 \right]_3 - \left[k_1 A_1 \left(\frac{dT}{dx} \right)_L \right]_1 = 0 \quad (27)$$

In Eqs. (25) through (27), the square brackets are subscripted by the link number, while the derivitive dT/dx is to be taken at the "beginning" (0) or the "end" (L) of the respective link. The assignment of (0) and (L) to each link is completely arbitrary. Note that, based on Eq. (24):

$$\left(\frac{dT}{dx} \right)_L = \frac{-\sqrt{\psi}}{\sinh \sqrt{\psi} L} [(T(0) - T_e) - (T(L) - T_e) \cosh \sqrt{\psi} L] \quad (28)$$

$$\left(\frac{dT}{dx} \right)_0 = \frac{\sqrt{\psi}}{\sinh \sqrt{\psi} L} [(T(L) - T_e) - (T(0) - T_e) \cosh \sqrt{\psi} L] \quad (29)$$

By substituting Eq. (28) and (29) into Eqs. (25) through (27), one gets three equations and six unknowns, the unknowns being the temperatures $T(L)_i$ and $T(0)_i$ ($i = 1, 2, 3$). However, three of these temperatures are redundant because of the junction conditions:

$$T(0)_1 = T(L)_2 \quad (30)$$

$$T(0)_3 = T(L)_1 \quad (31)$$

$$T(L)_3 = T(0)_2 \quad (32)$$

We choose to solve for $T(0)_i$ ($i = 1, 2, 3$) and to use the following abbreviations:

$$\Theta_i \equiv k_i A_i \frac{\sqrt{\psi_i}}{\sin h \sqrt{\psi_i} L_i} \quad (33)$$

$$\beta_i \equiv \Theta_i T_{e_i} (\cosh \sqrt{\psi_i} L_i - 1) \quad (34)$$

$$\delta_i \equiv \Theta_i \cosh \sqrt{\psi_i} L_i \quad (35)$$

Then, the system Eqs. (25) through (27), by using Eqs. (30) through (32) and (24), can be written in matrix form as:

$$\begin{pmatrix} \delta_1 + \delta_2 & \theta_2 & -\theta_1 \\ -\theta_2 & \delta_2 + \delta_3 & -\theta_3 \\ -\theta_1 & -\theta_3 & \delta_1 + \delta_3 \end{pmatrix} \begin{pmatrix} T(0)_1 \\ T(0)_2 \\ T(0)_3 \end{pmatrix}$$

$$= \begin{pmatrix} \beta_1 + \beta_2 \\ \beta_2 + \beta_3 \\ \beta_1 + \beta_3 \end{pmatrix} \tag{36}$$

The solutions of Eq. (36) are given by:

$$T(0)_1 = \frac{\begin{bmatrix} (\beta_1 + \beta_2) & \theta_2 & -\theta_1 \\ (\beta_2 + \beta_3) & \delta_2 + \delta_3 & -\theta_3 \\ (\beta_1 + \beta_3) & -\theta_3 & \delta_1 + \delta_3 \end{bmatrix}}{\mathcal{D}} \tag{37}$$

$$T(0)_2 = \frac{\begin{bmatrix} \delta_1 + \delta_2 & (\beta_1 + \beta_2) & -\theta_1 \\ -\theta_2 & (\beta_2 + \beta_3) & -\theta_3 \\ -\theta_1 & (\beta_1 + \beta_3) & \delta_1 + \delta_3 \end{bmatrix}}{\mathcal{D}} \tag{38}$$

$$T(0)_3 = \frac{\begin{bmatrix} \delta_1 + \delta_2 & \theta_2 & (\beta_1 + \beta_2) \\ -\theta_2 & \delta_2 + \delta_3 & (\beta_2 + \beta_3) \\ \theta_1 & -\theta_3 & (\beta_1 + \beta_3) \end{bmatrix}}{\mathcal{D}} \tag{39}$$

where:

$$\mathcal{D} = \begin{bmatrix} \delta_1 + \delta_2 & \theta_2 & -\theta_1 \\ -\theta_2 & \delta_2 + \delta_3 & -\theta_3 \\ -\theta_1 & -\theta_3 & \delta_1 + \delta_3 \end{bmatrix} \tag{40}$$

V. Conclusions

In this first report we have outlined a simple, analytic method for obtaining the temperature distribution in a typical arrangement of antenna structural back-up member. The results were verified by the numerical shooting method. Good agreement between the two methods was obtained. Further work will deal with setting up the matrices for more complex structures, with radiation exchange, and with comparison between simulations and actual field measurements.

References

1. Baars, J. W., "Technology of Large Radio Telescopes for Millimeter and Submillimeter Wavelengths", *Reviews in Infrared and Millimeter Waves*, Vol. 2, K. J. Button ed., Plenum Publishing Corp., N.Y.

2. Glazer, S., and Gayle, G., "Thermal Measurement Technique of Rib Elements on DSN Antenna Structure", *The Telecommunications and Data Acquisition Progress Report 42-66*, Jet Propulsion Laboratory, Pasadena, Calif., December 15, 1981.

3. Smith, J. P., *SINDA User's Manual*, RO-71-810, TRW Systems Group, April 1971.

4. Schonfeld, D., "Analysis of Capacitive Heat Exchangers, Part II", *The Telecommunications and Data Acquisition Progress Report 42-69*, Jet Propulsion Laboratory Pasadena, Calif., June 15, 1982.

Table 1. Example of temperature distribution obtained by shooting method (B)	
x	T
0.00	0.560+03
0 10	0 559+03
0.20	0.558+03
0.30	0.557+03
0.40	0.557+03
0.50	0.556+03
0.60	0.556+03
0.70	0.555+03
0.80	0.555+03
0.90	0 555+03
i 00	0.555+03
1.10	0.555+03
1.20	0.555+03
1.30	0 555+03
1 40	0.555+03
1.50	0 555+03
1.60	0.555+03
1 70	0.556+03
1 80	0.556+03
1 90	0.557+03
2.00	0.558+03
2 10	0.559+03
2 20	0.560+03
2.30	0.561+03
2.40	0.562+03
2 50	0.564+03
2.60	0.566+03
2 70	0 569+03
2 80	0.572+03
2 90	0.576+03
3.00	0.580+03

Table 2. Example of temperature distribution using the linearization method (C)	
x	T
0.00	0.560+03
0.10	0.559+03
0.20	0 558+03
0.30	0.557+03
0.40	0.556+03
0 50	0.556+03
0.60	0.555+03
0.70	0.555+03
0.80	0 555+03
0.90	0.555+03
1.00	0.554+03
1 10	0.554+03
1 20	0.554+03
1.30	0 555+03
1 40	0.555+03
1 50	0.555+03
1 60	0 555+03
1.70	0.556+03
1 80	0.556+03
1 90	0.557+03
2.00	0.558+03
2 10	0 558+03
2.20	0.560+03
2 30	0.561+03
2 40	0.562+03
2.50	0.564+03
2 60	0 566+03
2 70	0.569+03
2.80	0.572+03
2 90	0.576+03
3.00	0.580+03

Fig. 2. Temperature distribution in a simple bar using
linearization method

Fig. 1. Physical system for a simple bar
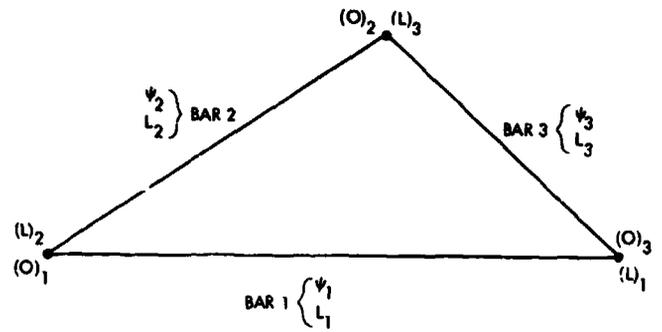
Fig. 3. Three-bar linkage

Performance of an Optical Relay Satellite Using Reed-Solomon Coding Over a Cascaded Optical PPM and BPSK Channel

D. Divsalar

Communications Systems Research Section

F. Naderi

Telecommunications Systems Section

The concept of using a relay satellite which receives information from deep space vehicles over an optical channel and relays this information to Earth over a microwave channel has been considered in the past. An important consideration in such a system is the nature of the optical/microwave interface aboard the relay satellite. In order to allow for the maximum system flexibility, without overburdening either the optical or RF channel, this paper considers the option of demodulating the optical channel on board the relay satellite but leaving the optical channel decoding to be performed at the ground station.

This not only removes some degree of complexity from the relay satellite but more importantly it circumvents restricting all deep space vehicles to a specific channel coding for which the decoder is provided on board the relay. For this scheme to be viable, the occurrence of erasures in the optical channel must be properly treated. A hard decision on the erasure (i.e., the relay selecting a symbol at random in case of erasure occurrence) seriously degrades the performance of the overall system. In this paper, we suggest coding the erasure occurrences at the relay and transmitting this information via an extra bit to the ground station where it can be used by the decoder. Many examples with varying bit/photon energy efficiency and for the noisy and noiseless optical channel have been considered. It is shown that coding the erasure occurrences dramatically improves the performance of the cascaded channel relative to the case of hard decision on the erasure by the relay.

I. Introduction

Optical communication systems have been studied in the past as a means of improving and/or expanding the capabilities of NASA's current Deep Space Network (Ref. 1). The principal advantage in communicating with optical frequencies is the potential increase in the information that can be transmitted to Earth from a deep space vehicle (DSV) having limited power and a structurally small antenna. However, Earth's atmosphere and adverse weather effects introduce attenuation and possible outages which limit the reliability of an Earth-based optical system.

In order to eliminate the atmospheric effects detrimental to optical links, an orbiting relay satellite might be employed. Such a satellite outside the Earth's atmosphere at the geosynchronous orbit would receive signals from the deep space vehicle over an optical link and subsequently relay the signals to Earth via a conventional microwave link (Ref. 2). Under this concept one would exploit advantages afforded by an optical link while eliminating some of its negative attributes. Figure 1 shows the block diagram for the optical/microwave communication system.

In the above, an important consideration is the nature of the optical/microwave link interface. Two alternatives were discussed in an earlier report (Ref. 2) and are briefly discussed here. Figure 2 shows the interface for option 1. Here the output of the photodetector is demodulated and decoded to recover the baseband data, which then modulates an RF carrier prior to transmission to Earth. The overall bit error probability at the ground station is given by the expression

$$PE = PE_{RF} + PE_o - 2PE_{RF} PE_o \approx PE_{RF} + PE_o \quad (1)$$

where PE_{RF} and PE_o are the bit error probabilities in the microwave and optical links respectively. The principal disadvantage of this scheme, which we shall refer to as a "demodulate and decode," is that by placing the demodulator and the decoder on the relay, the system is dictating the modulation and channel coding schemes to be used by all deep space vehicles which want to communicate through the relay satellite.

To mitigate this problem, the satellite can be operated in the "bent pipe" mode. Figure 3 shows the optical/microwave interface for this option where the analog output of the photodetector, $x(t)$, directly modulates (e.g., FM) an RF carrier. At the ground station, the RF carrier is demodulated to get an estimate of $x(t)$, which is then routed to an optical demodulator and a channel decoder to recover the baseband data. While this scheme offers the most flexibility, it also

imposes a wide bandwidth requirement on the microwave link. The performance of this system is being studied.

As a compromise between the above two alternatives, a third option is discussed and evaluated in this paper. In order to eliminate the wide bandwidth requirement of the bent pipe option and yet retain some degree of flexibility in the system, consider the case where the output of the photodetector on the relay satellite is demodulated but not decoded. The demodulated bit stream modulates an RF carrier for transmission to the ground station where the carrier is demodulated and decoded to recover the baseband data.

As an example, the optical link may use Reed-Solomon (RS) coding followed by a pulse position modulation (PPM) to transmit data from deep space to the relay satellite where the received signal is demodulated and the ensuing bit stream modulates an RF carrier using binary phase shift keying (BPSK). In the next section we analyze the performance of this system.

II. Performance of Reed-Solomon Coding Over Cascaded Optical PPM and BPSK Channel

In this section we analyze the performance of an (N, K) RS code over cascaded M-ary optical PPM and binary PSK channels. The mathematical model for the system under consideration is shown in Fig. 4. Allowing for erasure, the PPM channel can be modeled as an M-ary input $(M + 1)$-ary output discrete memoryless channel (DMC), characterized by symbol error probability P_s, erasure probability P_e and correct symbol probability P_c, shown in Fig. 5.

Two cases are considered. In the first case the erasure is eliminated on the relay satellite by hard decision (i.e., a symbol is randomly selected in case of an erasure occurrence). In the second case, the occurrence of an erasure is properly coded and transmitted to the ground station where this information is used by the decoder.

Case A — Hard Decision on Erasures at the Relay

In the case of hard decision, the symbol error probability P_{sh} and correct symbol probability P_{ch} are given by

$$P_{sh} = P_s + \frac{M-1}{M} P_e \quad (2)$$

and

$$P_{ch} = P_c + \frac{1}{M} P_e \quad (3)$$

After PPM demodulation (and the hard decision) RS symbols must be sent over a binary symmetric channel (BSC) characterized by bit error probability P_b given as

$$P_b = 1/2 \text{ erfc} \left(\sqrt{\frac{E_b}{N_0}} \right) \qquad (4)$$

where E_b is energy per bit and N_0 is the one-sided noise spectral density at the ground station. Since each RS symbol can be represented by $n = \log_2 M$ bits, transmission of an RS symbol can be modeled by a $(BSC)^n$ channel as depicted in Fig. 6.

The equivalent cascaded M-ary PPM channel, with hard-decided erasure symbols, and $(BSC)^n$ channel is shown in Fig. 7. From this figure, the symbol error probability of the equivalent channel is given by

$$\pi_{sh} = 1 - \left[\left(1 - \frac{MP_{sh}}{M-1} \right) \; (1 - P_b)^n + \frac{r_{sh}}{M-1} \right] \qquad (5)$$

If an (N, K) RS code is used in conjunction with the cascaded channel, a word error occurs when there are more than $(N-K)/2$ symbol errors in the N symbol received code word; then the word error probability of the RS code is given by

$$P_{wh}(RS) = \sum_{k=\frac{N-K}{2}+1}^{N} \binom{N}{k} \pi_{sh}^k \; (1 - \pi_{sh})^{N-k} \qquad (6)$$

where $N = M - 1$. Note that for large alphabet size RS codes in which we are interested, since the mass of spheres (with radius of one-half of the code minimum distance) around the code words is much smaller than the mass of total signal space, for a practical range of bit error rates, the probability of an incorrect decoding event can be ignored. Thus when the RS code fails to decode, we may have a bit in error. Under these conditions a bit error occurs when a bit in a received symbol is in error and there are $(N-K)/2$ or more symbol errors in the remaining $N-1$ symbols in a received code word. Then the bit error rate of the RS code is given by

$$P_{bh}(RS) = \left[\frac{M\pi_{sh}}{2(M-1)} \right] \sum_{k=\frac{N-K}{2}}^{N-1} \binom{N-1}{k} \pi_{sh}^k \; (1 - \pi_{sh})^{N-1-k} \qquad (7)$$

In (7), the expression in brackets represents the bit error probability before RS decoding and the summation in (7)

represents the probability of making $(N - K)/2$ or more symbol errors in $N - 1$ symbols of received code word.

Case B — No Hard Decision on Erasures at the Relay

We now consider the case where erasure information is encoded and relayed to the ground station and compare the performance with that of hard-decided erasure on the relay.

In order to transmit the erasure symbols over the downlink BSC, an extra bit is appended to the RS symbols. (We shall see shortly that for the range of E_b/N_0 of interest, one extra bit is sufficient to achieve acceptable performance). This means that symbols sent over the BSC channel are of lengths $n + 1$ bits. At the ground station the decoder examines the $(n + 1)th$ bit; if it is zero, the decoder accepts the first n bits as the RS symbol. However, if the appended bit is one, the receiver declares an erasure symbol and disregards the first n bits. For this case for each transmission of RS or erasure symbols the channel can be modelled as $(BSC)^{n+1}$. The equivalent cascaded PPM and $(BSC)^{n+1}$ channel is shown in Fig. 8. From this figure, the symbol erasure probability of the equivalent channel is:

$$\pi_e = P_c P_b + P_s P_b + P_e (1 - P_b) \qquad (8)$$

and the symbol error probability is

$$\pi_s = P_c \left[1 - (1 - P_b)^n \right] (1 - P_b) + \frac{M-1}{M} \; P_e P_b$$

$$+ \frac{1}{M-1} r_s [M - 2 + (1 - P_b)^n] (1 - P_b) \qquad (9)$$

While the correct symbol probability is

$$\pi_c = P_c (1 - P_b)^{n+1} + \frac{1}{M} P_e P_b + \frac{1}{M-1} P_s$$

$$\cdot [1 - (1 - P_b)^n] (1 - P_b) \qquad (10)$$

From the above three equations, the word error and the bit error probability of the RS code over the equivalent channel of Fig. 8 are given by (Ref. 3).

$$P_{ws}(RS) = \sum_{i=0}^{N} \sum_{j=\Delta}^{N-i} \binom{N}{i} \binom{N-i}{j} \pi_s^i \; \pi_e^j \; \pi_c^{N-i-j} \qquad (11)$$

and

$$P_{bs}(RS) = \frac{M}{2(M-1)} \sum_{i=0}^{N} \sum_{j=\Delta}^{N-i} \binom{N}{i} \binom{N-i}{j} \frac{i+j}{N}$$

$$\cdot \pi_s^i \pi_e^j \tau_c^{N-i-j} \qquad (12)$$

where

$$\Delta = MAX(N-K+1-2i, 0) \qquad (13)$$

III. Numerical Results

In order to compare the performance of the two cases discussed in the previous section, we now consider specific examples. Consider using M-ary PPM with $M = 256$ and three RS codes, namely (255, 223), (255, 191) and (255, 127).

Both noisy and noiseless optical channels are considered. Assuming no background noise, the PPM channel can be viewed as a purely erasure channel with

$$P_e = e^{-K_s} \qquad (14)$$

and

$$P_c = 1 - e^{-K_s} \qquad (15)$$

where K_s is the average number of photon counts per PPM frame of the optical channel. For the case of hard-decided erasures, using Eqs. (14) and (15) above in Eqs. (2) and (3) with P_s set to zero, we have

$$P_{sh} = \frac{M-1}{M} e^{-K_s} \qquad (16)$$

and

$$P_{ch} = 1 - \frac{M-1}{M} e^{-K_s} \qquad (17)$$

Using these results, we have plotted the bit error rate of the end-to-end system (i.e., $P_{bh}(RS)$) for the hard-decided erasure from Eq. (7) and $P_{bs}(RS)$ for the coded erasure from Eq. (12) as a function of E_b/N_0 of the microwave channel for various energy efficiency of ρ bits/photon where ρ is given by

$$\rho = \frac{K \log_2 M}{N K_s} \qquad (18)$$

The results are given in Figs. 9-11.

Next the procedure is repeated for the case of a noisy optical channel. In the presence of background noise, a PPM threshold detector (Ref. 3) is used where for each slot in the PPM frame, the number of received photons is compared with a threshold γ to decide the presence or the absence of the signal.

Assume Poisson distribution in the number of received photons in each slot, with a mean K_b in the absence of the signal and mean $K_b + K_s$ in the presence of a signal in th slot. Then the probability of correctly detecting the presence of the signal is given by

$$P_{ds} = \sum_{k=\gamma+1}^{\infty} \frac{(K_s + K_b)^k}{k!} e^{-(K_s+K_b)} \qquad (19)$$

and the probability of correctly detecting the absence of the signal (or presence of the noise) is given by

$$P_{dn} = \sum_{k=0}^{\gamma} \frac{(K_b)}{k!} e^{-K_b} \qquad (20)$$

Using Eqs. (19) and (20) we have

$$P_s = (M-1)(1-P_{ds})(1-P_{dn})P_{dn}^{M-2}$$

$$P_c = P_{ds} P_{dn}^{M-1} \qquad (21)$$

$$P_e = 1 - P_c - P_s$$

These results are next used in Eqs. (2) and (3) to calculate P_{sh} and P_{ch} and in Eqs. (8) through (10) to calculate π_e, π_s and π_c, which are subsequently used to calculate $P_{bh}(RS)$ and $P_{bs}(RS)$. The threshold γ has been optimized to give minimum bit error rates. $P_{bh}(RS)$ and $P_{bs}(RS)$ are shown in Figs. 12-20.

IV. Conclusion

This paper has considered an optical relay satellite system which has been modeled as a cascaded optical PPM and microwave BPSK channel. In order to maintain maximum system flexibility, the relay satellite demodulates but does not decode the optical channel; the decoding is performed by the ground station. The occurrence of optical erasures is properly coded by the relay and transmitted to the ground decoder using an extra bit. It is shown that this improves the overall system performance dramatically relative to the case where the relay makes hard decision on the erasures, thereby screening the ground decoder from the knowledge of such erasure occurrences.

References

1. Vilnrotter, V. A., and Gagliardi, R. M., "Optical-Communication Systems for Deep Space Applications," Publication 80-7, Jet Propulsion Laboratory, Pasadena, California, March 15, 1980.

2. Gagliardi, R. M., Vilnrotter, V. A., and Dolinar, S. J., "Optical Deep Space Communication via Relay Satellite," Publication 81-40, Jet Propulsion Laboratory, Pasadena, California, August 15, 1981.

3. Lesh, J. R., Katz, J., Tan, H. H., and Zwillinger, D., "2.5-Bit/Detected Photon Demonstration Program, Analysis and Phase 1 Results," TDA Progress Report 42-66, Jet Propulsion Laboratory, Pasadena, California, pp. 115-132, December 15, 1981.
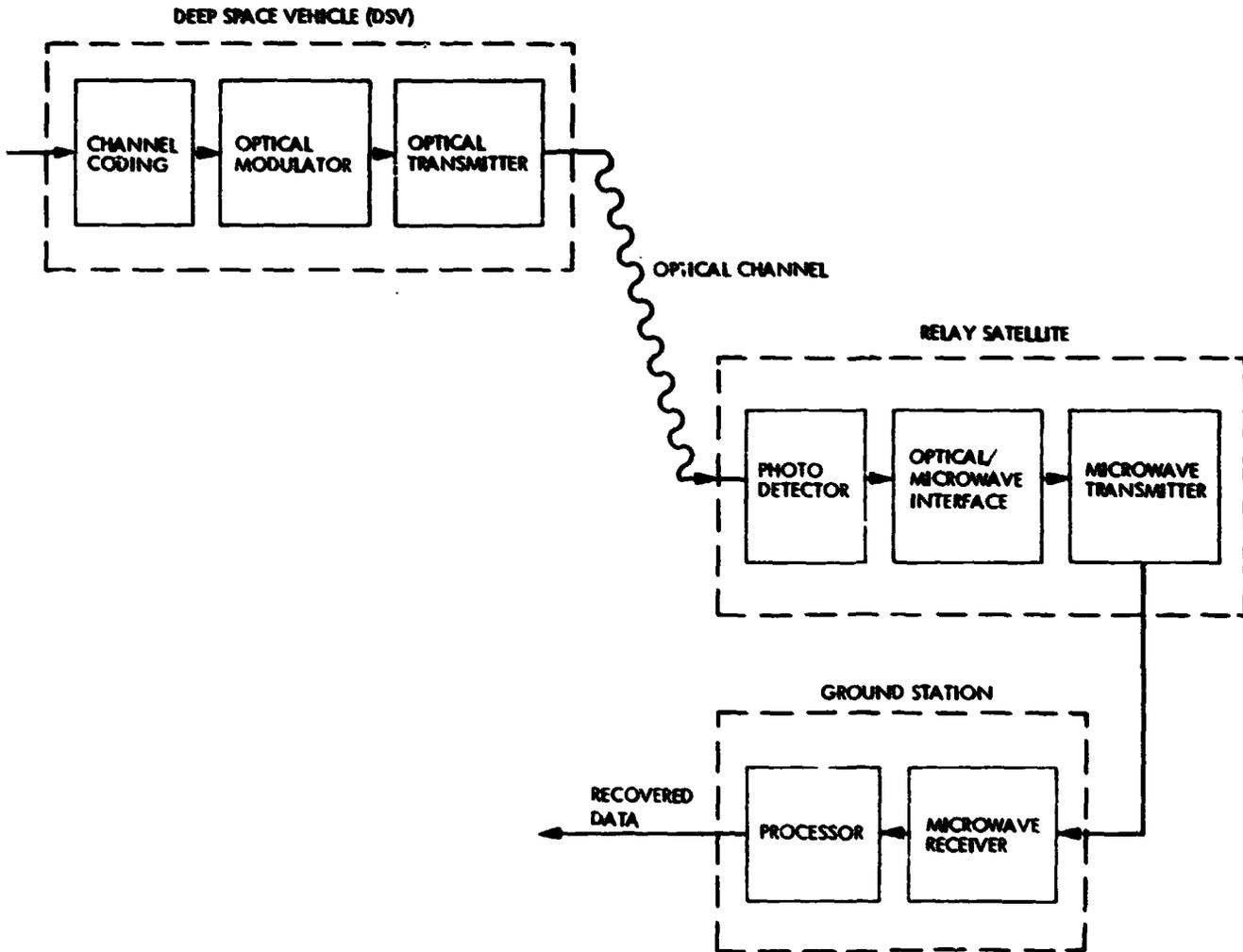
Fig. 1. Block diagram of an optical relay satellite system

OPTICAL/MICROWAVE INTERFACE

PHOTO DETECTOR → OPTICAL CH. DEMODULATOR → OPTICAL CH. DECODER → MICROWAVE CH. MODULATOR → MICROWAVE TRANSMITTER →

(a)
OPTICAL/MICROWAVE INTERFACE ON THE RELAY SATELLITE

← RECOVERED DATA ← MICROWAVE DEMODULATOR ← MICROWAVE RECEIVER ←
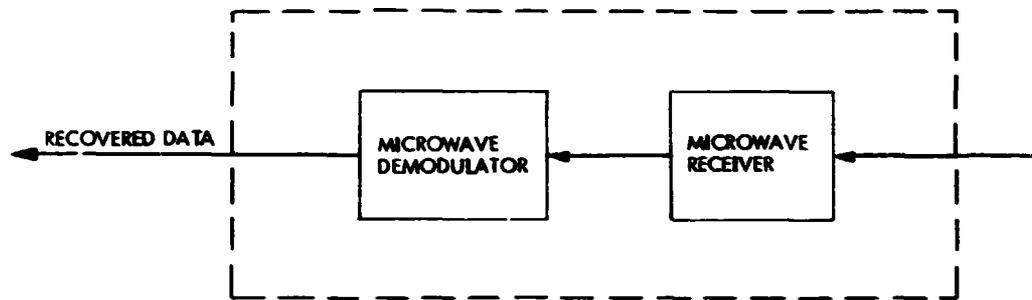
(b)
GROUND STATION PROCESSING

Fig. 2. Block diagram of "demodulate and decode" option
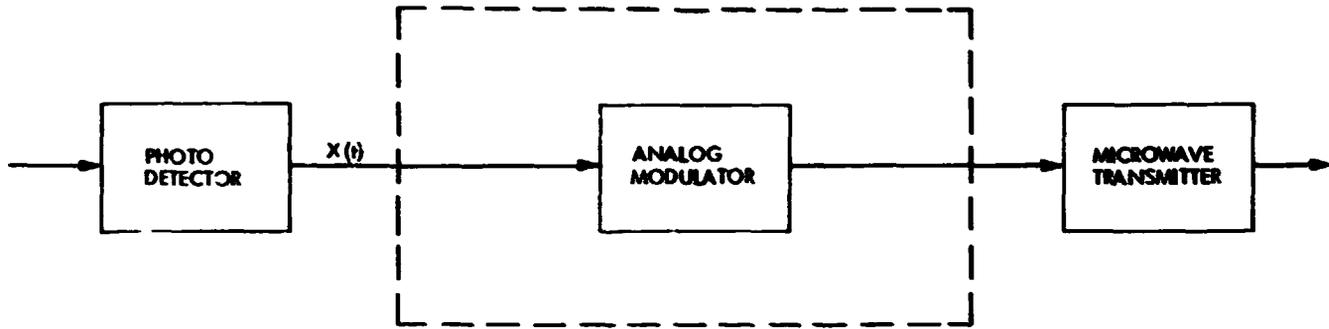
(a)
OPTICAL/MICROWAVE INTERFACE ON THE RELAY SATELLITE
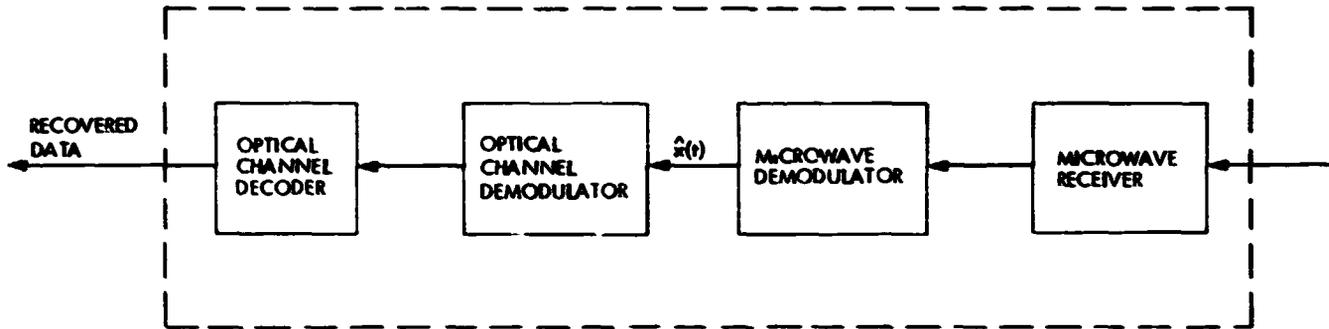
(b)
GROUND STATION PROCESSING

Fig. 3. Block diagram of the "bent pipe" option
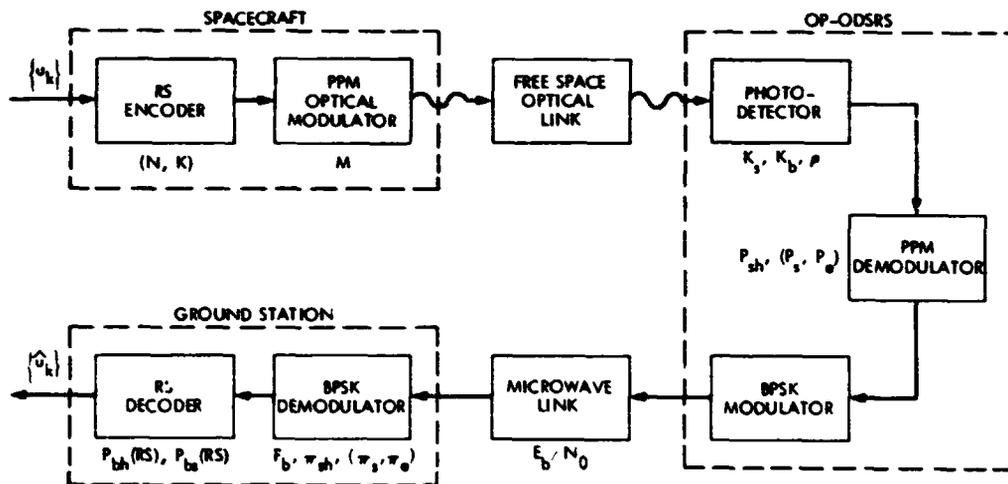
Fig. 4. Block diagram of the end-to-end system
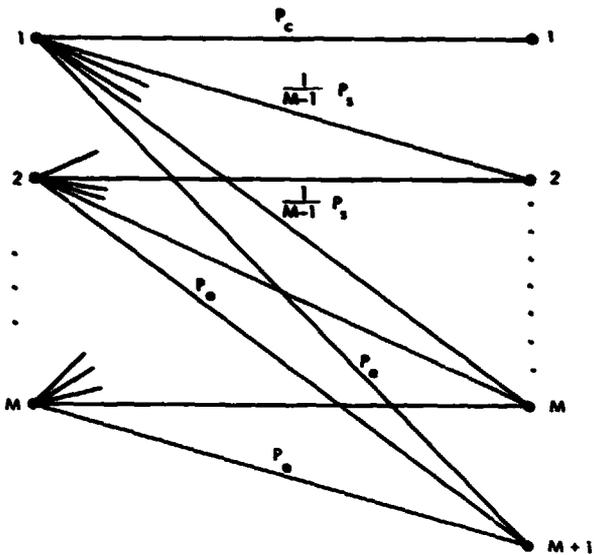
Fig. 5. PPM channel with erasure
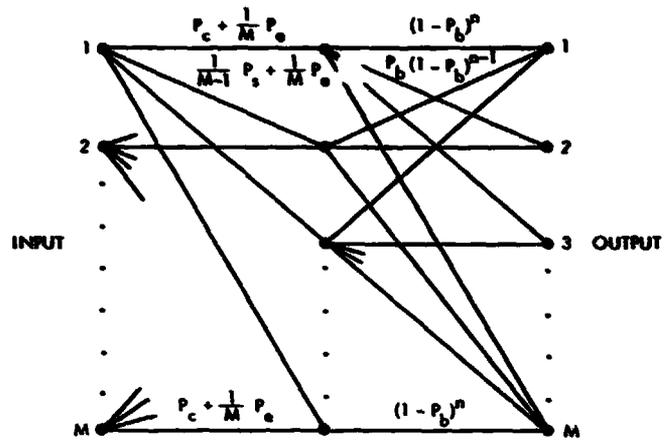
Fig. 7. Equivalent hard-decided cascaded channel
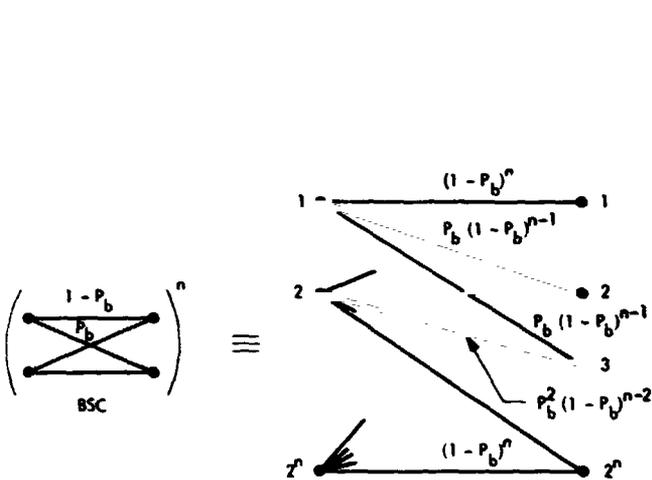
Fig. 6. (BSC)^n channel
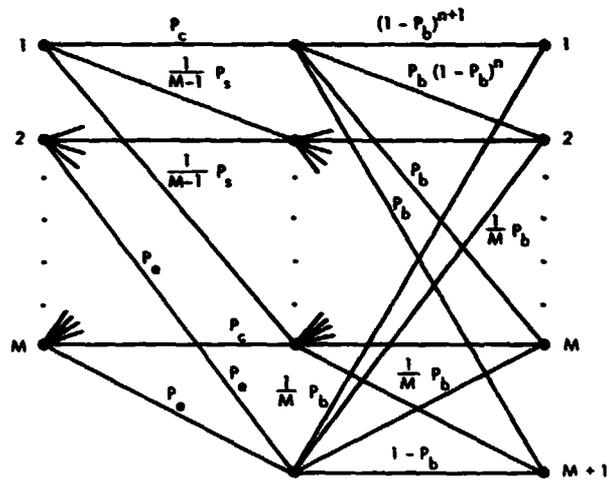
Fig. 8. Equivalent soft-decided cascaded channel
(sending one extra bit for erasure symbol)
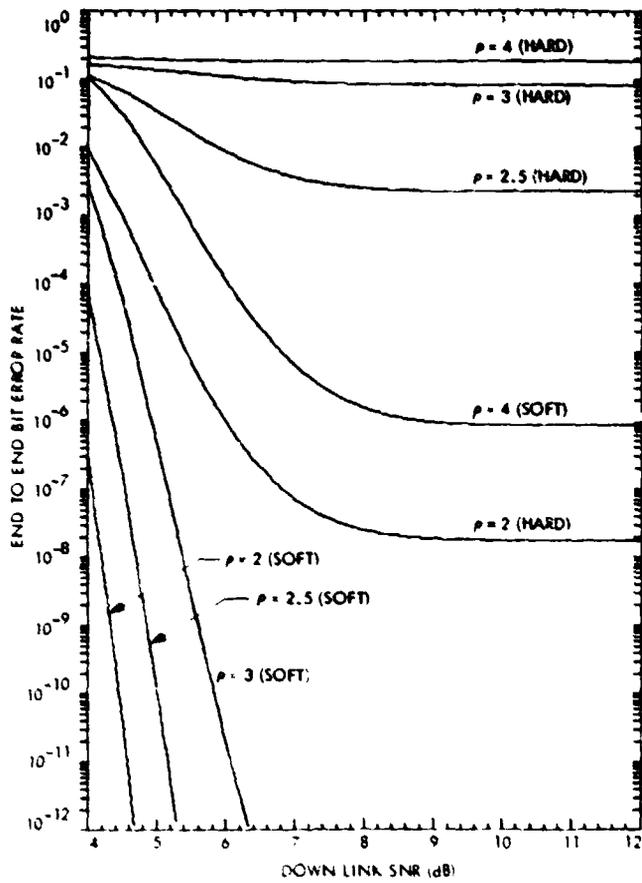
Fig. 9. P_{bh} (RS) and P_{bs} (RS) vs downlink E_s/N_0 for RS(255, 127), code rate 1/2 with no background noise, for various energy efficiency ρ (bits/photon)
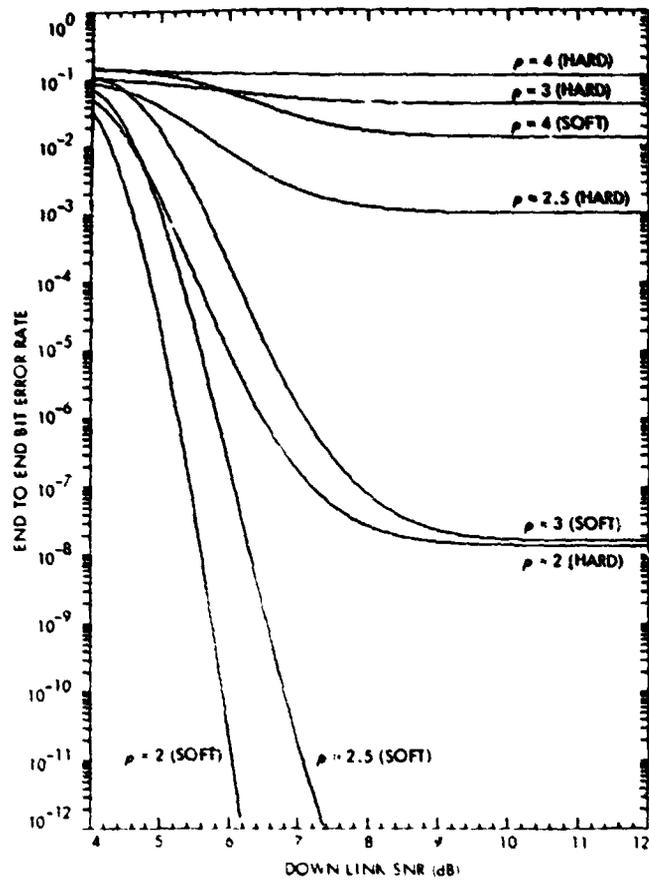
Fig. 10. P_{bh} (RS) and P_{bs} (RS) vs downlink E_s/N_0 for RS(255, 191), code rate 3/4 with no background noise, for various energy efficiency ρ (bits/photon)
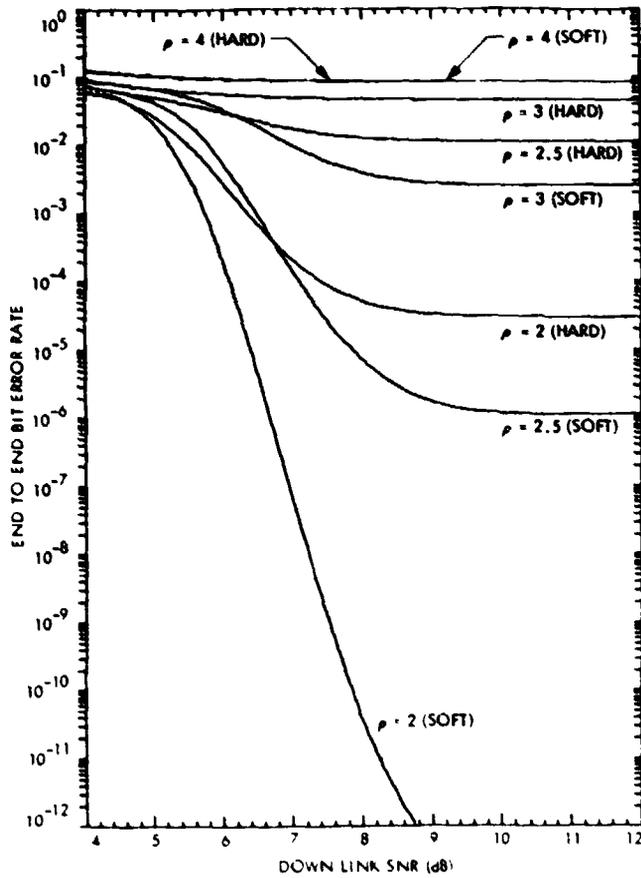
Fig. 11. P_{bh} (RS) and P_{be} (RS) vs downlink E_b/N_0 for RS(255, 223), code rate ~ 7/8 with no background noise, for various energy efficiency ρ (bits/photon)

Fig. 12. P_{bh} (RS) and P_{be} (RS) vs downlink E_b/N_0 for RS(255, 127), code rate ~ 1/2 with background noise $K_b = 10^{-2}$, for various energy efficiency ρ (bits/photon)

127
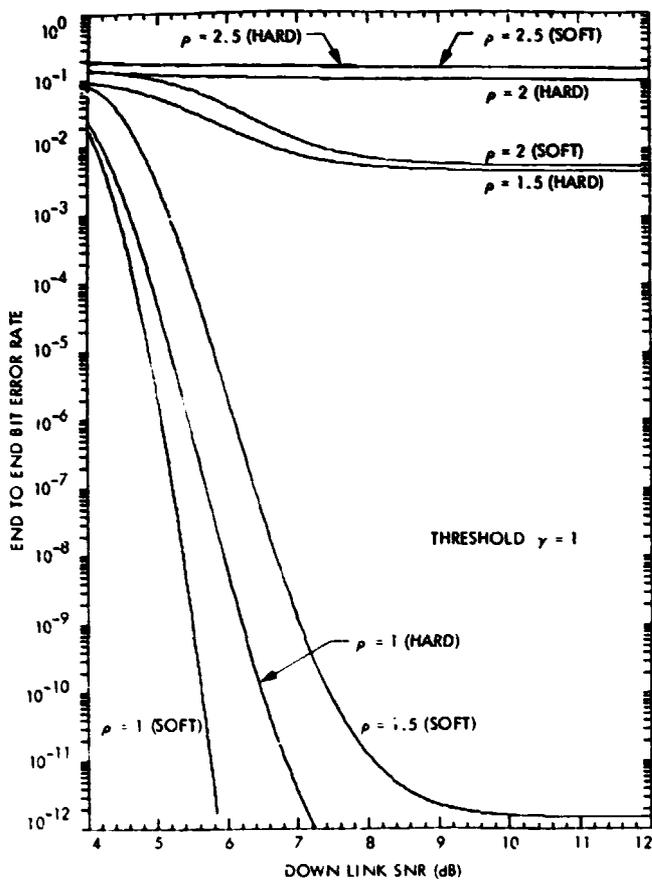
Fig. 13. P_{bh} (RS) and P_{bs} (RS) vs downlink E_b/N_0 for RS(255, 191), code rate ≈ 3/4 with background noise $K_b = 10^{-2}$, for various energy efficiency ρ (bits/photon)
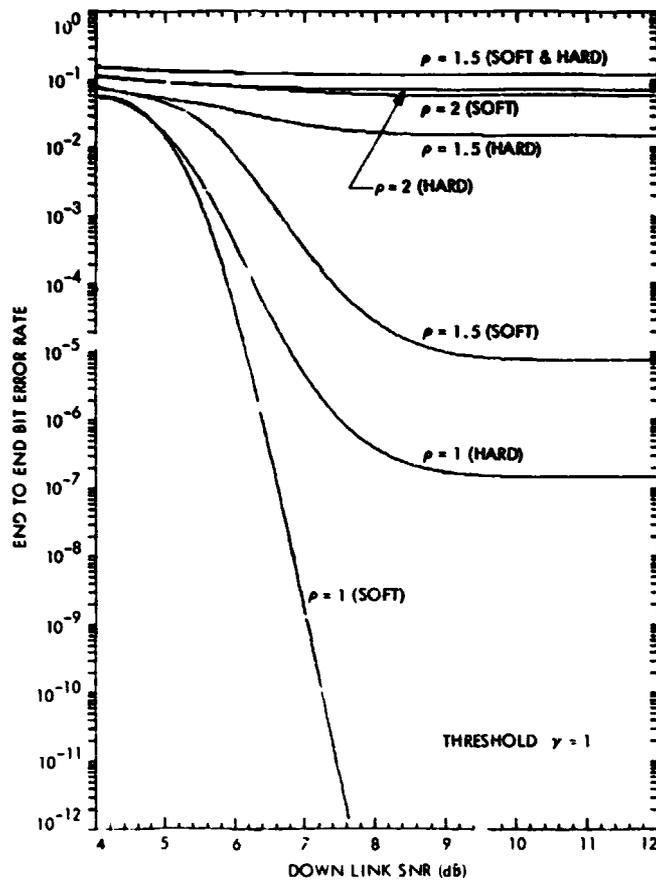
Fig. 14. P_{bh} (RS) and P_{bs} (RS) vs downlink E_b/N_0 for RS(255, 223), code rate ≈ 7/8 with background noise $K_b = 10^{-2}$, for various energy efficiency ρ (bits/photon)

Fig. 15. P_{bh} (RS) and P_{be} (RS) vs downlink E_b/N_0 for RS(255, 127), code rate = 1/2 with background noise $K_b = 10^{-3}$, for various energy efficiency ρ (bits/photon)

Fig. 16. P_{bh} (RS) and P_{be} (RS) vs downlink E_b/N_0 for RS(255, 191), code rate = 3/4 with background noise $K_b = 10^{-3}$, for various energy efficiency ρ (bits/photon)
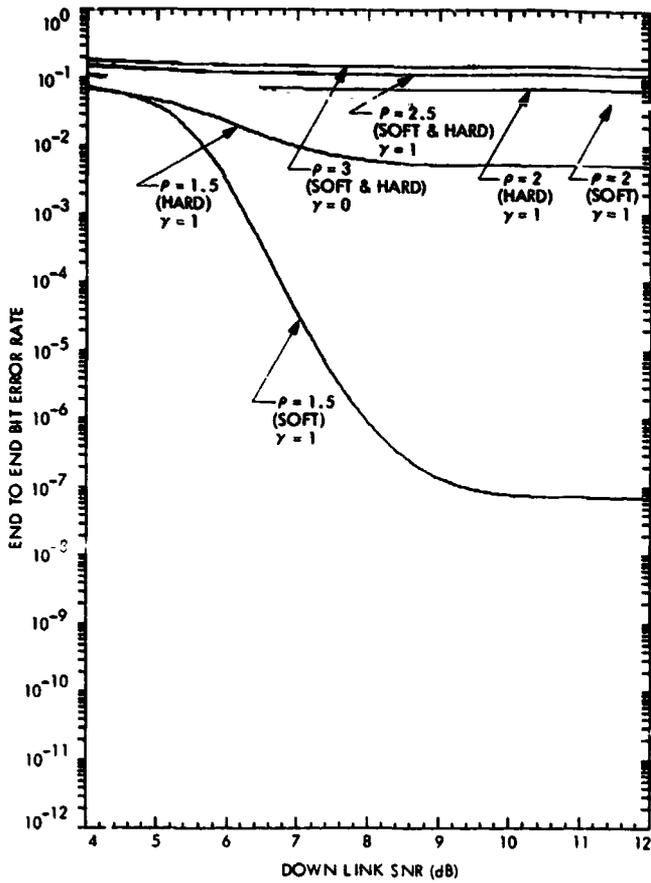
Fig. 17. P_{bh} (RS) and P_{be} (RS) vs downlink E_b/N_0 for RS(255, 223), code rate = 7/8 with background noise $K_b = 10^{-3}$, for various energy efficiency ρ (bits/photon)
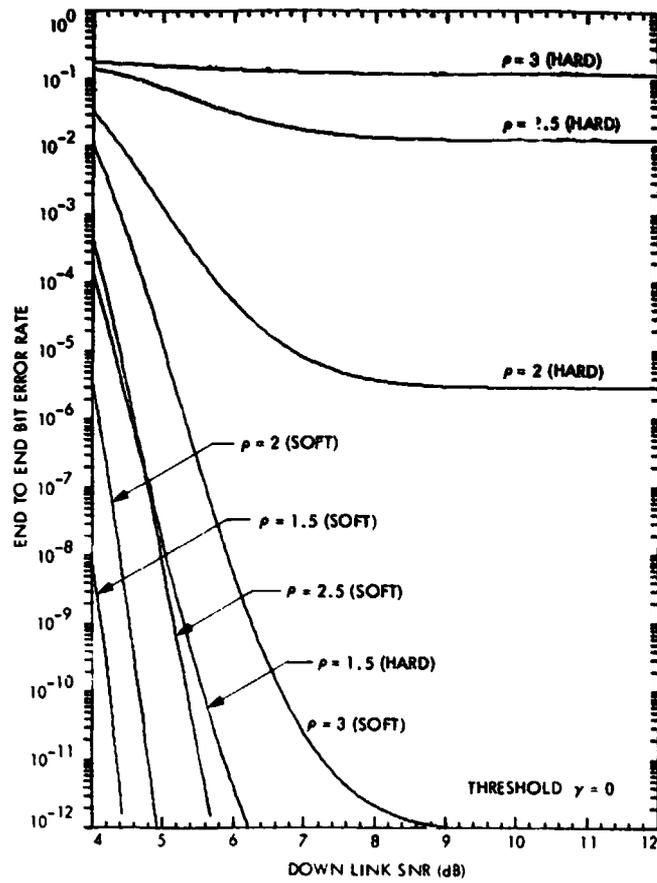
Fig. 18. P_{bh} (RS) and P_{be} (RS) vs downlink E_b/N_0 for RS(255, 127), code rate = 1/2 with background noise $K_b = 10^{-4}$, for various energy efficiency ρ (bits/photon)
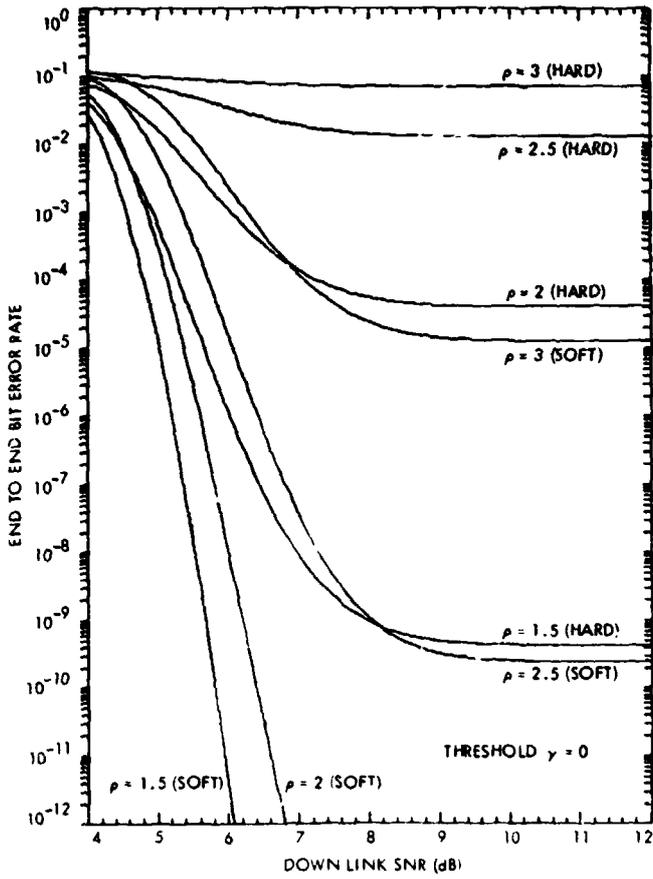
Fig. 19. P_{bh} (RS) and P_{be} (RS) vs downlink E_b/N_0 for RS(255, 191), code rate ~ 3/4 with background noise $K_b = 10^{-4}$, for various energy efficiency ρ (bits/photon)

Fig. 20. P_{bh} (RS) and P_{be} (RS) vs downlink E_b/N_0 for RS(255, 223), code rate ~ 7/8 with background noise $K_b = 10^{-4}$, for various energy efficiency ρ (bits/photon)

131

DSN Data Systems Software Methodology

C. K. Hung

Data Systems Section

This report presents a software methodology for JPL Deep Space Network (DSN) data systems software implementations through transfer and delivery. The DSN Data Systems Software Methodology is compatible with and depends on DSN software methodology and also incorporates the characteristics of real-time program development in a DSN environment. The DSN Data Systems software implementation consists of a series of six distinct phases: (a) planning and requirements, (b) design definition, (c) design and production, (d) combined subsystem testing, (e) acceptance test and transfer, (f) operation and maintenance. The unit (function) demonstration tests du. .g the design and production phase will be planned early in the design definition phase. Each demonstration will serve as a significant milestone and will show an increased program capability. An Independent Group (IG) is responsible for verification and validation of the DSN Data Systems software during development phases.

The DSN Data Systems Software Methodology is being applied to all development software provided for or by the DSN Data Systems Section in Mark IV where there is a desire for reliability, maintainability, and usability within budget and schedule constraints.

I. Introduction

The Deep Space Network (DSN) Data Systems Software Methodology is to provide standard methods and policies for the orderly implementation and management of the DSN Data Systems software through transfer and delivery. The DSN Data Systems Software Methodology is compatible with and depends on DSN software methodology standards (Ref. 1), and also incorporates the characteristics of real-time program development in a DSN environment. The overview of the DSN Data Systems software development techniques are as follows:

(1) The DSN Data Systems software implementation plan is a series of six distinct phases:

    (a) P' ..ning and requirements

    (b) Design definition

    (c) Design and production

    (Combined subsystem testing

        Acceptar   ' and 'ransfer

    (f) Operations and maintena ice.

Before the acceptance testing phase, combined subsystem testing will be performed. The combined subsystem test phase will demonstrate that:

(a) That the program meets functional, performance, and interface requirements in its real-time environment and

(b) Proper concurrent operation of combined subsystems.

(2) The unit (function) demonstration tests during the design and production phase will be planned early in the design definition phase and will appear on the detailed Work Breakdown Structure (WBS) Each demonstration thus:

(a) Shows an increased program capability

(b) Proves module compatibility

(c) Serves as a significant milestone (quality and quantity)

(d) Produces test plans and procedures which often become a part of the Software Test and Transfer Document (STT).

(3) An Independent Group (IG) is responsible for verification and validation of the software during development phases.

II. Software Classification

The extent to which DSN Data Systems software methods and management policies are applied to the implementation of an individual software program is dependent upon the classification of that software into one of three software classes.

The individual software classes are defined as follows:

(1) Class 1—New operation software

(2) Class 2—Modified operation software (new functions added or major revisions) where implementation cost exceeds one-half man year or $25,000

(3) Class 3—Modified operation software (corrections or minor revisions) where implementation cost is less than one-half man year or $25,000.

Table 1 lists the implementation documentation requirements that are applicable to the three software classes.

III. Software Design and Development Process

A. Implementation Phases

This paragraph presents each phase of the software design and development process. It identifies all of the responsible people, their function and it relates to each phase of the process, the products that will be generated, and the monitoring, reviews, and control involved. Figure 1 summarizes the DSN Data System Software Management and Implementation Plan.

The phases of DSN Data Systems software implementation involve:

(1) Planning and requirements

(2) Design definition

(3) Design and production

(4) DSN Data Systems combined subsystem testing

(5) Acceptance test and transfer

(6) Operations and maintenance.

B. Program Language

Coding of the software design is performed using a standard high-level language approved for real-time and nonreal-time programs. The HAL/S language is a currently approved language used for all new minicomputer subsystem software development. The PL/M language is a currently approved language used for all Control and Computation Module (CCM) subsystem software/firmware.

The use of inherited code is not subject to the constraints and does not require a waiver to use an existing language.

C. Milestones

Based on the activities (major documents, design and test reviews) discussed in Paragraph III.A., eight major milestones are established which allow the overall software implementation project to be planned and its in-progress development to be monitored. The milestones, in order of occurrence, are as follows:

(1) Software Requirements Document (SRD) approved and DSN Level D review completed

(2) Project Data Flow and Interface Design Document (LFD) approved

(3) Software Definition Document (SDD), Preliminary Software Operator's Manual (PSOM), and Development Test Plan (DTP) approved and architectural design review (Level E) completed

(4) Final demonstration test review completed

(5) Combined Subsystem Test (CST) review and pre-acceptance test review completed. Combined Subsystem Test Plan and Test Procedure (CTT) approved

(6) Software Operator's Manual (SOM) approved

(7) Software Specification Document (SSD) approved

(8) Software Test and Transfer (STT) Document approved. Transfer Agreement signed, deliverable transferred to Operations.

D. Reviews

Technical reviews are required to assess the technical quality and progress of the development during DSN Data Systems Section software implementation process. Reviews are held at the end of each implementation phase. These reviews are of two types: DSN Data Systems Section internal reviews and DSN formal review. The DSN formal reviews are described in Ref. 1.

The reviews, in order of occurrence, are as described in the paragraphs which follow.

1. DSN subsystem function design review (Level D). The DSN formal Level D review for each subsystem's software and the common software shall be held at the end of software planning and requirements phase.

2. Peer design review. The peer design review shall be held after the DSN formal Level D review, at the end of design definition phase, to allow for corrective action and before the DSN formal subsystem detail design (Level E) review. The DSN Data Systems peer design review procedure is shown in Figure 2. The peer design review criteria shall include:

(1) Requirement traceability

(2) Architectural design

(3) Man/machine interface

(4) Testing factors

(5) Management information

(6) Hardware and software development systems

The detailed review form is shown in Ref. 1. Members of the peer design teams shall consist of the following:

(1) Software Manager (Chairman)

(2) Independent Group representative

(3) Data flow and interface engineer

(4) Combined subsystem testing engineer

(5) Senio software engineer

Also, the DSN subsystem engineer and operational representative participate in the peer review.

3. DSN subsystem detail design review (Level E). The DSN formal Level E review for each subsystem's software and the common software shall be held after the subsystem peer design review and before substantially entering the design and production phase.

4. Unit (function) demonstration test review. The unit (function) demonstration test review for each subsystem and the common software is held at the end of each demonstration test during the design and production phase.

The technical group supervisor shall be responsible for calling the meeting and for assuring resolution of all open items. Members of review teams shall include the Software Manager and the cognizant development engineer and his/her staff. The Independent Group common software engineer may also participate, if required. The review focuses primarily on:

(1) Performance against design. The review

   (a) Systematically lists all program parameters such as data queue, Input/Output (I/O) queues, buffer pool (released buffer), etc.

   (b) Creates boundary conditions for these parameters and systematically checks that all cases are performed properly against design.

(2) Functional performance against functional requirement. The review

   (a) Systematically lists all program parameters and system parameters such as 1200-bit-per-second line, error polynomial, etc.

   (b) Creates boundary conditions for these parameters with test driver/other programs and checks systematically that all cases are performed properly.

(3) Tests (creates) all possible hardware failures and their error recoveries. Also tests all peak load where all interrupt signals occur simultaneously.

5. **Final demonstration test review.** The final demonstration test review (section internal) shall be held at the end of the design and production phase.

6. **Combined subsystem test review.** The combined subsystem test review (section internal) shall be held at the end of the DSN Data Systems combined subsystem testing phase.

7. **DSN subsystem transfer to operations review (Level F).** This formal DSN review is optional and is held normally only for subsystem transfers.

E. Documentation

Documentation is a baselined product of each of the software implementation phases, and is produced concurrently with the design and other implementation activities conducted in each phase. Typical outlines for documentation are shown in Refs. 1, 2, and 3. Table 2 summarizes the documents and their responsibility.

F. Quality Control

The Independent Group shall assure the traceability of software requirements through architectural design, design and production, integration combined subsystem test, and transfer. The Independent Group conducts and observes the pre-acceptance test and combined subsystem tests, and is a signatory to test reports.

Before completion of the transfer of the software from implementation to operations, DSN Quality Assurance (QA) certifies the status of the SSD (including the code). Detailed DSN QA requirements and procedures are contained in Ref. 1.

G. Configuration Management

All documentation and software products are maintained under configuration control by Software Production Management and Control (SPMC). SPMC provides security, integrity, - ' controlled access to material within its custody and enforces configuration management practices.

The SRD, SDD, SSD, SOM, and STT are subject (ultimately) to DSN configuration management. The TRA, DTP, PSOM, DFD, CTT, and SMP are DSN Data Systems Section project documents, and are subject to DSN Data Systems Section change control only.

1. **Section (internal) change control.** The TRA, DTP, PSOM, DFD, CTT, and SMP are maintained under DSN Data

Systems Section configuration control in SPMC files, and are updated whenever a change request is approved. All change requests need to be justified and need to be accompanied by the modified change pages reflecting the change at the same or lower level or detail as was included in the original approved document.

2. **DSN change control.** The SRD and SDD are not maintained beyond program transfer to operations. The SSD, SOM, and STT are separate as-built documents delivered along with the completed program and are maintained throughout the operational life of the program. Therefore, the SSD, SOM, and STT are subject to DSN change control procedures.

Detailed DSN Configuration Management requirements and procedures are contained in Ref. 1.

3. **Waiver.** Deviations from the Software Management plan may appear in the SRD or through the waiver procedure. All approved waivers shall be formally documented by the SPMC.

H. Methodology and Tool Utilization

There are two main categories of computers used operationally and for software development support:

(1) minicomputers

(2) CCM-based microprocessors. Minicomputer applications are intended to be developed on Intel MDS systems and/or a CCM-based Development Work Station (DWS). The development tool summary is shown in Table 3.

IV. DSN Data Systems Mark IV Software Implementation

Over the last several years (1978-1980), the DSN Data Systems Ground Communication Group has employed software development techniques such as unit (function) demo tests, combined subsystem testing, etc., for developing reliable real-time software for the Ground Communication Facility (GCF)-Network Operation Control Center (NOCC) Reconfiguration Project (Ref. 4 and 5).

The DSN Data Systems Section is playing a major role in the implementation of the Network Consolidation Project (NCP) (Ref. 6 and 7). (Referred to in this paper as the Mark IVA Implementation Project.) This software methodology applies to all development software in Mark IVA provided for, or by, the DSN Data Systems Section in the Mark IVA

era where there is a desire for reliability, maintainability, and usability within budget and schedule constraints.

As of mid-July 1982, all DSN Data Systems computer programs are in the architecture design phase or the design and production phase. Table 4 contains the software classifica-

tions for Mark IVA Interim software. Table 5 contains the software classification for the Mark IVA final software.

The status of the DSN Data System software implementation for the NCP will be described in a subsequent issue of the TDA Progress Report.

References

1. "Standard Practices for the Implementation of Computer Software," JPL 78-53, Jet Propulsion Laboratory, Pasadena, California, September 1, 1978.

2. "Preparation Guide for Class B Software Specification Documents," JPL 79-56, Jet Propulsion Laboratory, Pasadena, California, October 1, 1979.

3. Robert C. Tausworthe, Standardized Development of Computer Software, Part 1 (1976), Part 2 (1978), JPL SP-43-29, Jet Propulsion Laboratory, Pasadena, California.

4. Bremner, D. S., and Hung, C. K., "Ground Communication Facility and Network Operations Control Center Reconfiguration," The Telecommunications and Data Acquisition Progress Report 42-58, pp. 108-109, Jet Propulsion Laboratory, Pasadena, California, August 15, 1980.

5. McClure, J. P., "GCF-NOCC Reconfiguration," The Deep Space Network Progress Report 42-55, pp. 86-89, Jet Propulsion Laboratory, Pasadena, California, February 1980.

6. Gatz, E. C., "Network Consolidation Program System Design," The Telecommunications and Data Acquisition Progress Report 42-63, pp. 150-153, Jet Propulsion Laboratory, Pasadena, California, June 15, 1981.

7. Yeater, M. L., and Herrman, D. T., "Network Consolidation Program," The Telecommunications and Data Acquisition Progress Report 42-65, pp. 19-24, Jet Propulsion Laboratory, Pasadena, California, October 1981.

8. Robert C. Tausworthe, "Deep Space Network Software Cost Estimation Model," JPL 81-7, Jet Propulsion Laboratory, Pasadena, California, April 15, 1981.

Table 1. Software classifications

Document	Product for each program	Software classifications		
		Class 1	Class 2*	Class 3*
SRD	Software Requirements Document	Yes	Yes**	Yes**
SDD	Software Definition Document	Yes	Yes	Yes***
PSOM	Preliminary Software Operator's Manual	Yes	Yes (Redline SOM)	Yes (Redline SOM)
DTP	Development Test Plan	Yes	Yes	Yes
SOM	Software Operator's Manual	Yes	Yes	Yes
SSD	Software Specification Document	Yes	Yes	Yes
STT	Software Test and Transfer Document	Yes	Yes	Yes
TRA	Test Requirements Analysis Report	Yes	If critical	No

*Existing Documents updated for Changes
**ECR + Softcost (Ref. 8) = SRD on Changes
***ECO + WBS = SDD on Changes

Table 2. Documentation

Phase	Document	Responsibility	Approved By	Review By
1	SRD	CDE	DSN Line Mgrs (810-13) S/W Mgr (Concur)	IG
2	SDD	CDE	DSN Line Mgrs (810-13) S/W Mgr (Concur)	IG
2	Prelim SOM	CDE	Supervisor, S/W Mgr	IG
2	DTP	CDE	Supervisor, S/W Mgr	IG
3, 4	SOM	CDE	DSN Line Mgrs (810-13)	S/W Mgr IG
3, 4	SSD	CDE	DSN Line Mgrs (810-13)	
3, 4, 5	STT	CDE	DSN Line Mgrs (810-13)	IG
1, 2	DFD	Data Flow and Interfaces Eng	Subsystem Eng Mgr Section Mgr	S/W Mgr, CDFs IG/CST Eng
2, 3, 4	CTT	CST Eng	Integration Mgr Section Mgr	S/W Mgr IG, TGSs, CDEs
1, 2, 3	TRA	IG	CST Eng, Integration Mgr Subsystem Eng Mgr	S/W Mgr CDEs
1	SMP	S/W Mgr	Section Mgr	TGSs, CDEs, Cost Eng Integration Mgr

Table 3. Development Tool Summary

Tool	Phases	Utilization	
Cost model	1	CDE/Supervisor	; Scheduling
Word processor	1, 2, 3, 4, 5	CDE, SPMC	; Documentation
Work Breakdown Structure (WBS)	2, 3, 4, 5	CDE/Supervisor	; Scheduling and milestones
Program Design Language (PDL)	2*, 3, 4, 5	CDE, SPMC, QA	; Documentation
HAL/S	3, 4, 5	CDE	; Modcomp high-order language
PL/M	3, 4, 5	CDE	; CCM S/W/firmware language
Developmen Version Control System (DVCS)	3, 4, 5	CDE, SPMC, QA	; Version control
Multiple terminal for software code	3, 4, 5	CDE/Implementer	; Code development
Anomaly Report System (ARS)	3, 4, 5	CDE, COE, IG	; Anomaly Reporting System
Source Editor (SED)	3, 4, 5	CDE/Implementer	; Modcomp Source Editor
Assembler*	3, 4, 5	CDE/Implementer	; Modcomp language
Test generator*	3, 4, 5	CDE/Implementer	; Debugger

*Option

Table 4. Software classification for Mark IVA interim software

Subsystem	Program	S/W classification	Code estimates (X 1000 lines)
Command	Deep Space Station (DSS) Command Modulator Assembly (CMA) firmware	1	1.6
	DSS Command Processor Assemoly (CPA)	2	4.4
	Network Operation Control Center (NOCC) Command Real-Time Monitor (NCD)	2	0.2
Telemetry	DSS Telemetry Processor Assembly (DTM)	2	3.2
	NOCC Telemetry Real-Time Monitor (NTM)	2	0.12
NOCC Display	NOCC Display Subsystem (NDS)	2	2.4
	Video Assembly Processor (VAP)	2	0.3
Common Software	NOCC Common Software	2	5.6

Table 5. Software classification for Mark IVA final software

Subsystem	Program	S/W classification	Code estimates (X 1000 lines)
Command	Deep Space Communications Complex (DSCC) Command Processor Assembly (CPA)	2	0.8
	Network Operations Control Center (NOCC) Command Real-Time Monitor (NCD)	2	2.0
Test Support	DSCC Test Support Assembly (TSA)	1	31.7
	NOCC Test and Support Assembly (NTS)	2	12.0
Antenna Mechanical	DSCC Antenna Pointing Assembly (APA)	1	25.5
	DSCC Antenna Control Subassembly (ACS)	1	12.0
Tracking	DSCC Metric Data Assembly (MDA)	2	21.0
	NOCC Network Tracking Assembly (NTK)	2	25.0
Radio Science	DSCC Occultation Data Assembly (ODA)	2	16.1
Monitor and Control	DSCC Complex Monitor and Control (CMC)	1	40.0
	DSCC Link Monitor and Control (LMC)	1	35.0
	NOCC Monitor and Control (NMC)	1	20.0
Telemetry	DSCC Telemetry Processor Assembly (DTM)	2	20.0
	NOCC Telemetry Real-Time Monitor (NTM)	2	25.1
Ground Communication Facility	DSCC Area Routing Assembly (ARA)	1	22.0
	Central Communication Terminal (CCT) Error Correction and Switching Assembly (ECS)	2	12.0
	CCT Data Record Generator Assembly (DRG)	2	8.0
	CCT Central Communication Monitor (CCM)	2	7.2
	NOCC Network Communication Equipment (NCE)	3	8.0
Display	NOCC Display Subsystem (NDS)	2	0.3
	NOCC Video Assembly Processor (VAP)	2	0.3
Network Support	Star and VAX Interface Adapter (SVIA)	1	0.5
	Network Support Subsystem (NSS) Control Input/Output	1	10.0
	NSS Sequence of Events (SOE)	1	14.0
	NSS Standard and Limit (S&L)	1	16.0
	NSS System Performance Record (SPR)	1	16.0
Common Software	Functional Independent Data Module (FIDM), IEEE 488, STAR and FTS handler	1	4.0
	Local Area Network (LAN) RS-232 device handler for microprocess	1	0.8

ITEM / PHASES	SOFTWARE PLANNING AND REQUIREMENTS	SOFTWARE DESIGN DEFINITION (ARCHITECTURAL DESIGN PHASE)	SOFTWARE DESIGN AND PRODUCTION

MAJOR ACTIVITY BY COGNIZANT DEVELOPMENT ENGINEER (CDE)/IMPLEMENTORS	• GENERATE SRD
	  • REQ. FOR S/W
	  • INHERITED PGM CONSIDERATIONS
	• COST MODEL - CLASS C WBS IN MADNET DATA BASE
	• REVIEW
		• GENERATE SDD, PRELIM SOM AND DTP
		  • FUNCTIONS OF EACH PGM AND DATA STRUCTURE
		  • ARCHITECTURAL DESIGN AND ORDER OF DEV.
		  • MAN/MACHINE INTERFACE
		  • DATA FLOW AND INTERFACE (INTERNAL AND EXTERNAL) DESIGN
		  • UNIT (FUNCTION) DEMO TEST PLAN AND H/W EQUIPMENTS NEED DATES
		• CLASS B WBS
		• REVIEW
			• DETAIL DESIGN, CODING AND TESTIN
			  • UNIT (FUNCTION) DEMO TESTS –
			  OF INCREASING PGM FUNCTION
			  • FINAL DEMO TEST AND REVIEW
			  • ANOMALIES – W B S
			• GENERATING FINAL PRELIM. SOM.
			• REVIEW AND DISTRIBUTE SOM, SSD, STT

MAJOR ACTIVITY INDEPENDENT GROUP (IG)	• VERIFY AND VALIDATE S/W REQUIREMENTS AGAINST FRD/FDD • INITIATE A TEST REQ. ANALYSIS REPORT (TRA) • PARTICIPATE IN REVIEW FUNCTIONS AND INTERFACE WITH CDEs IN FORMAL SCHEDULED MEETINGS. THE TGSs SHOULD SCHEDULE THESE MEETINGS	• EVALUATE ADEQUACY OF S/W ARCHITECTURE DESIGN TO MEET ALLOCATED REQUIREMENTS • PREPARE AND ISSUE FIRST DRAFT OF TRA REPORT • PARTICIPATE IN REVIEW FUNCTIONS	• COMPLETE THE TRA AND SUPPORT CST IN PREPARING CTT • REVIEW PRELIM. CTT AND STT TO ENSU MEET FUNCTIONAL REQUIREMENTS • OBSERVE DEMO TESTS AND REVIEW TES • PARTICIPATE IN REVIEW FUNCTIONS
MAJOR ACTIVITY BY COMBINED SUBSYSTEM TEST ENGINEER	• ESTABLISH THE TEST PHILOSOPHY FOR H. W AND S/W	• INITIATE COMBINED SUBSYSTEM TEST PLAN AND TEST PROCEDURE (CTT)	• GENERATING OF CTT
MAJOR ACTIVITY BY DATA FLOW AND INTERFACES ENGINEER	• PREPARE AND ISSUE PRELIM. DFD • COORDINATE INTERFACES	• GENERATE DFD • ENSURE CONSISTENCY/INTEGRITY OF END-TO-END DATA FLOW FOR ALL DSN SYSTEMS	• ENSURE CONSISTENCY/INTEGRITY END-TO-END DATA FLOW FOR ALL DSN SYSTEMS
MAJOR ACTIVITY BY COMMON SOFTWARE ENGINEER	• IDENTIFY COMMON S/W ELEMENTS • ESTABLISH AN IMPLEMENTATION PLAN FOR COMMON S/W AND ASSOCIATED WBS's ESTABLISH PRODUCTION PERSONNEL THROUGH TGSs	• COORDINATE PREPARATION OF FUNCTIONAL SPEC. OF NEW COMMON S/W ELEMENT • PROVIDE "USER" DOCUMENTATION AND TRAINING AND SUPPORT SERVICE IN SPMC	• COORDINATE PRODUCTION OF CON AND IS RESPONSIBLE FOR FUNCTION TESTING
MAJOR ACTIVITY BY SOFTWARE MANAGER	• GENERATE SOFTWARE MANAGEMENT PLAN (SMP) • MONITOR AND MANAGE	• ASSURE PROPER INTER SUBSYSTEM COORDINATION IN SUPPORT OF SECTION INTERNAL/EXTERNAL S/W I/F TESTS, . . . ETC, • MONITOR AND MANAGE	• MONITOR AND MANAGE (WBS, DEM • COORDINATE H/W REQ. WITH H/W N • RESOLVE TECHNICAL CONFLICTS
MAJOR DOCUMENTS FOR EACH PROGRAM	SOFTWARE REQUIREMENTS DOCUMENT — SRD	SOFTWARE DEFINITION DOCUMENT — SDD PRELIM SOFTWARE OPERATOR MANUAL — PRELIM SOM DEVELOP TEST PLAN — DTP	SOM PRELIM. SOFTWARE SPECIFICATI DOCUMENT
MAJOR PROJECT DOCUMENTS FOR SECTION (338) SUBSYSTEMS	SOFTWARE MANAGEMENT PLAN — SMP PROJECT DATA FLOW AND INTERFACES DESIGN DOCUMENT — DFD PRELIM.	TEST REQUIREMENTS ANALYSIS — TRA PRELIM. DFD	COMBINED SUBSYS TEST PLAN AND TEST PROCEDURE
DESIGN REVIEWS ▲ DENOTES FORMAL △ DENOTES SECTION INFORMAL		DSN LEVEL D ▲ 810-10	LEVEL E PEER ▲ DSN ▲ TGSs	DEMO TEST REVIEW △ - - - △ GROUP INTERNAL
ANOMALY REPORT	N/A	N/A	APPLICABLE (338 WBS
MILESTONES	1. SRD APPROVED AND DSN LEVEL D REVIEW COMPLETED (CDE) ▽1	2. DFD APPROVED (DFE) 3. SDD, PRELIM. SOM, DTP APPROVED AND ARCHITECTURAL REVIEW COMPL "ED (CDE, TGS) ▽2 ▽3	4. FINAL DEM TESTS REVIE COMPLETED
DEVELOPMENT TOOLS USED	• SPMC JACQUARD WORD PROCESSOR • COST MODEL - CLASS C WBS IN MADNET DATA BASE	• SPMC JACQUARD WORD PROCESSOR • CLASS B WBS IN MADNET DATA BASE • PDL FOR TOP-LEVEL DESIGN (OPTIONAL)	• PDL • HAL/S • PL/M • DVCS • 7870 • WBS • ARS • SED • JACQUARD WORD PROCESSOR • ASSEMBLER (IF REQUIRED) • STAR
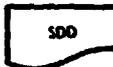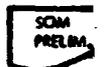
SOFTWARE DESIGN AND PRODUCTION	SECTION 338 COMBINED SUBSYSTEM TESTING	ACCEPTANCE TEST AND TRANSFER TO OPERATION	OPERATION AND MAINTENANCE	
DETAIL DESIGN, CODING AND TESTING • UNIT (FUNCTION) DEMO TESTS - PROVE DEMO OF INCREASING PGM FUNCTION CAPABILITY • FINAL DEMO TEST AND REVIEW • ANOMALIES — W B S GENERATING FINAL PRELIM. SOM, SSD, STT REVIEW AND DISTRIBUTE SOM, SSD, STT	• SUPPORT CST EGR AND IG FOR TESTING • DEMONSTRATE PROGRAM MEETS SRD AND DFD IN REAL TIME ENVIRONMENT • DEMONSTRATE PROPER CONCURRENT OPERATION OF COMBINED SUBSYSTEMS • FIX ANOMALIES - STT, CTT • REVISE STT, SOM, SSD • SUBMIT PRELIM, SSD & PGM TO QA • ANOMALIES —WBS, ARS	• RESPONSIBLE FOR THE OVERALL ACCEPTANCE TEST PROCESS AT DSS 11 • SUPPORT COE FOR ACCEPTANCE TESTING, SPT AND SOAK TESTS • FIX ANOMALIES AND REVISED STT • TRANSFER • SUBMIT REVISED SSD & PGM TO QA • ANOMALIES - ARS , W B S	• REMOVE LIENS • ANOMALIES - ARS	1.
COMPLETE THE TRA AND SUPPORT CST EGR N PREPARING CTT VIEW PRELIM. CTT AND STT TO ENSURE THEY MEET FUNCTIONAL REQUIREMENTS OBSERVE DEMO TESTS AND REVIEW TEST RESULTS PARTICIPATE IN REVIEW FUNCTIONS	• CONDUCT THE PRE-ACCEPTANCE TEST IN ACCORDANCE WITH STT AT 338 LAB, SIF/CTA 21 OR DSS 11 • SUPPORT CST EGR FOR CST • ANOMALIES — A R S	• MONITOR AND SUPPORT S/W ACCEPTANCE TEST FOR CONFORMANCE WITH STT • EVALUATE CONFORMITY OF TEST RESULTS WITH TRA		2.
GENERATING OF CTT	• CONDUCTING THE CST IN ACCORDANCE WITH THE CTT AT SECTION LAB, SIF/CTA21 OR DSS 11 • REVISED CTT • ANOMALIES	• ENSURE CONSISTENCY/INTEGRITY OF COMBINED SUBSYSTEMS		3.
ENSURE CONSISTENCY/INTEGRITY OF END-TO-END DATA FLOW FOR ALL SYSTEMS	• ENSURE CONSISTENCY/INTEGRITY OF END-TO-END DATA FLOW FOR ALL DSN SYSTEMS	• ENSURE CONSISTENCY/INTEGRITY OF END-TO-END DATA FLOW FOR ALL DSN SYSTEMS		
COORDINATE PRODUCTION OF COMMON S/W AND IS RESPONSIBLE FOR FUNCTIONAL TESTING	• COORDINATE PRODUCTION OF COMMON S/W AND IS RESPONSIBLE FOR FUNCTIONAL TESTING	• RESPONSIBLE FOR FUNCTIONAL TESTING		
MONITOR AND MANAGE (WBS, DEMO TESTS) COORDINATE H/W REQ. WITH H/W MGR RESOLVE TECHNICAL CONFLICTS	• REVIEW CTT • MONITOR AND MANAGE	MONITOR AND MANAGE		

Fig. 1. DSN data systems softwar

FOLDOUT FRAME 2

|	ACCEPTANCE TEST AND TRANSFER TO OPERATION	OPERATION AND MAINTENANCE	REMARKS
⁝ ⁝ ᴛION ⁚ ⁚	• RESPONSIBLE FOR THE OVERALL ACCEPTANCE TEST PROCESS AT DSS 11 • SUPPORT COE FOR ACCEPTANCE TESTING, SPT AND SOAK TESTS • FIX ANOMALIES AND REVISED STT • TRANSFER • SUBMIT REVISED SSD & PGM TO QA • ANOMALIES - ARS, WBS	• REMOVE LIENS • ANOMALIES – ARS	1. SENIOR S/W EGR FROM EACH TECHNICAL GROUP WILL ADVISE ON V&V DURING PLANNING AND REQUIREMENTS PHASE AND DESIGN DEFINITION PHASE
ANCE ⁚	• MONITOR AND SUPPORT S/W ACCEPTANCE TEST FOR CONFORMANCE WITH STT • EVALUATE CONFORMITY OF TEST RESULTS WITH TRA		2. SENIOR S/W ENGINEERS WILL SERVE AS TECHNICAL ADVISOR FOR SECTION S/W DEVELOPMENT DURING IMPLEMENTATION PHASES
⁚ 11	• ENSURE CONSISTENCY/INTEGRITY OF COMBINED SUBSYSTEMS		3. COE CONDUCTS ACCEPTANCE TEST
⁚	• ENSURE CONSISTENCY/INTEGRITY OF END-TO-END DATA FLOW FOR ALL DSN SYSTEMS		
	• RESPONSIBLE FOR FUNCTIONAL TESTING		
	MONITOR AND MANAGE		
	[SOM] [SSD] [STT]		4. SSD, SOM AND STT ARE UNDER DSN CHANGE CONTROL
			5. TRA UNDER. SMP INTERNAL DFD CHANGE CTT CONTROL
	LEVEL F		6. HARDWARE AVAILABILITY
ANCE ⚠	TRANSFER ⚠		
D/I.G.	COE		
		• ANOMALIES REPORTED TO COE	3
	APPLICABLE (DSN)	APPLICABLE (DSN)	FOLDOUT FRAME
		• CSE ANALYSIS & CATEGORIZATION	
▽5	6. SOM APPROVED (CDE) 7. SSD APPROVED (CDE) 8. STT APPROVED; DELIVERABLES TO OPERATIONS (CDE) ▽6 ▽7 ▽8		ORIGINAL PAGE IS OF POOR QUALITY
	• PDL • HAL/S • PL/M • DVCS • 7870 • WBS • ARS • SED • JACQUARD WORD PROCESSOR • ASSEMBLER (IF REQ) • STAR TEST GENERATOR	• DSN ARS	

Fig. 1. DSN data systems software management and implementation

Fig. 2. DSN data systems peer design review procedure

A Federated Information Management System for the Deep Space Network

E. Dobinson

DSN Data Systems Section

This paper examines some general requirements for an information management system for the Deep Space Network (DSN). It also presents a concise review of available database management system technology. It then recommends that a federation of logically decentralized databases be implemented for the Network Information Management System of the DSN. Overall characteristics of the federation are specified, as well as reasons for adopting this approach.

I. Introduction

There have been many advances in database management systems over the last decade. Faced with the task of modelling a particular application environment, an organization today must make important choices. Off-the-shelf products are commercially available for a wide selection of computers. Nonprocedural query languages, report-writers, forms-based interfaces, programming languages, and graphics are but some of the tools offered for today's applications. However, before selecting a particular product, certain fundamental issues of database organization must be considered. The functional requirements of the application environment must be analyzed and then carefully matched to an information system architecture best suited to meet those requirements.

The purpose of this paper is to study the information management needs of the Deep Space Network of NASA, and to recommend a database management system architecture which will meet those needs most effectively. We begin with an overview of the Deep Space Network, describing the way in which the organization is administered, and the ways in which various elements of this administration interact with each other and with the rest of the Jet Propulsion Laboratory. From this discussion we formulate some general requirements for an information management system [Section II]. We then turn to an examination of information management systems available today [Section III]. By matching the characteristics of these systems with our requirements, we recommend an approach for the DSN, give a top-level design of the system using a small, but representative, subset of data, and indicate how this system can be expanded to serve adequately the whole of the DSN and/or JPL [Section IV]. We then analyze the benefits and costs of our choice in comparison with the benefits and costs of an alternative proposal [Section V]. Lastly, we conclude with a brief summary of the paper [Section VI].

This paper is of particular importance because of its timely coincidence with plans for the Network Information Management System for the DSN already underway. As no major software decisions for this system have as yet been made, the recommendations contained in this paper may be considered.

II. The Deep Space Network

We begin this section with an overview of DSN operations. We then focus in some detail on several key administr :ive activities. This leads us to formulate general requirements for a DSN infurmation management system.

A. An Overview

The Deep Space Network (DSN) of the National Aeronautics and Space Administration (NASA) is responsible for the guidance and control of all of NASA's unmanned spacecraft at planetary and interplanetary distances, as well as for the receiving and processing of the vast amounts of information these spacecraft acquire and send back to Earth. This network is made up of tracking stations around the world, a central control organization at the Jet Propulsion Laboratory in Pasadena, California, and the ground communication linking them together. The three station groups, called Deep Space Communications Complexes (DSCC), are located approximately 120 degrees apart in longitude, so that a spacecraft is always in view of at least one antenna as the Earth rotates. These locations are Goldstone, California; Madrid, Spain; and Canberra, Australia. These stations function as autonomous organizations under management at JPL. This management is decentralized at JPL in various locations both on and off the laboratory, including a secondary site at Hill Street in Pasadena.

A variety of administrative activities in the DSN require the management of data. These activities are presently supported by separate application systems, each of which has its own set of data. However, for many of these activities the data overlap. At present, there are few automatic mechanisms for these activities to share data. It is cumbersome as well for an activity to span several systems. In addition, it is difficult for the three Deep Space Stations to cooperate in the performance of these administrative functions or to interchange data amongst themselves.

The need for improving this situation has been recognized by JPL management. To this end, an extensive study has been undertaken, under outside contract, which has resulted in a proposal of a hardware and communications configuration for improved operations. The proposed system is called the Network Information Management (NIM) System. It is for this system that we will address our database design. The NIM assembly is described in detail in Refs. 1 and 2. The proposed worldwide network will initially consist of four nodes, one at JPL and one at each of the three station complexes. In addition, each node is itself an internal local network. Each NIM node will have hardware, software, and communications to provide a distributed computing environment for the DSN.

The NIM study, undertaken by the Aaron-Ross Corporation for JPL, has produced an extensive survey of all of the components of the DSN (Ref. 3). This survey identifies the responsibilities and requirements of various DSN activities in terms of their database needs. We shall avail ourselves of its contents throughout this paper in formulating our own recommendations.

B. Some Important Administrative Activities

In this discussion we will focus on some important administrative activities of the Deep Space Network in order to determine a design for an information management system which will permit these activities to function efficiently, and which will give management the overview and knowledge it needs to do its job well. Because the totality of these activities is much too great for the scope of this report, we will concentrate on several important operations which span the entire organization.

1. **Engineering Change Management (ECM).** Engineering Change Management is a complex, far-reaching DSN activity. It is coordinated at present by a group in section 377 at JPL, directed by a Change Control Board, and involves a large number of personnel throughout the JPL/DSN organization. The process of instituting an engineering change involves the initiation of an Engineering Change Request (ECR). This request is carefully assessed by representatives from all other systems that might be affected by the change. The assessments are then brought before the Change Control Board, which passes judgement on the request. The request may either be denied, approved, or sent back for further evaluation. Once a request is approved, one or more Engineering Change Orders (ECOs) are issued to design and implement the change. Each ECO is then planned in detail, with costs and schedules developed for each phase. At this point the evaluation phase is complete and the design and implementation phase begins.

During the design and implementation phase the ECM group functions to collect status information about the actual schedule performance, and to alert management if any rescheduling will be required. A general awareness of the progress of the ECO is needed by everyone involved, including logistics, maintenance, and mission planning personnel who need to know when the installation will actually occur. As the Aaron-Ross survey points out "... an ECR has the potential to affect nearly every aspect of DSN operations and support, ranging from mission performance analysis to spare parts provisioning and from maintenance personnel scheduling to DSN utilization forecasts. As a consequence, there is a large constituency of personnel, with widely varying needs, all of whom absolutely require or can benefit by a conveniently obtained status and schedule forecast for ECOs.

When the design and implementation of the ECO are complete, the installation at the tracking sites begins. The scheduling of the installation of the equipment must be coordinated with the tracking schedules, so as not to interfere with any mission, and yet be there in time for any future missions that require it. There are also, in addition, some temporary ECOs whose removal must be scheduled similarly. When the equipment is finally installed and running, or when the temporary equipment is removed, the ECO is closed out.

The Engineering Change Management is clearly an important activity, having the capability to affect the DSN in many ways. The data representing the initiation and assessment phases are of interest to a variety of people at JPL, while the data for the development, implementation, and installation schedules may be needed by a variety of personnel throughout the entire organization.

2. Equipment and Materials Management. This DSN activity is responsible for the management of JPL property, DSN tracking equipment, repairable spare parts, other maintenance spares, and consumables. Management of equipment and materials involves obtaining them in the first place (provisioning and procurement), keeping track of their location and status (inventory and control), and moving them from place to place (transportation and shipping). These activities are distributed among several organizational elements at JPL and at the Deep Space Stations.

3. Anomaly Reporting Services. The knowledge and control of anomalies occurring from time to time throughout the DSN is an important activity for its well being. To this end the DSN has procedures for the reporting of various anomalies. Two categories of reports which are processed are Discrepancy Reports and Failure Reports. The ultimate goal of these reports is to provide DSN engineering, operations, support, and management the information on which to make changes in equipment, technology, procedures, and policy. There are three basic activities connected with these reports. These are (1) filling them out, (2) validating, investigating, and analyzing them, and (3) summarizing the status of the anomalies to management.

4. Other DSN activities. In addition to the three activi-ies highlighted above there are many more too numerous to list. These include energy management, financial management, personnel management, scheduling, maintenance, production control, as well as activities that pertain to the tracking sites only, such as operations, repairs, maintenance and integration, cabling, etc. For each of these activities the efficient management of data is extremely important.

C. General Requirements for a DSN Information Management System

There are three general requirements for an information management system for the DSN. First of all, the system must permit the necessary interchange of data between the various administrative activities, as well as between the various physical sites of the organization. Secondly, the system must allow each of these activities to develop and function autonomously. And, thirdly, the system must be capable of evolving incrementally over time.

1. The need for sharing. There are some obvious relationships between the various DSN management activities. For example, there is a natural interaction between the ECM system, the equipment management system, and the anomaly reporting system. An ECO almost always will affect the equipment database. Either new equipment will be installed or old equipment will be removed, or both. Anomaly reports can and often do result in the initiation of an ECR. Furthermore, the implementation of an ECO can result in anomalies. And so on. At present none of these interactions can take place automatically. Relationships are maintained manually, if at all.

In addition to the need for having activities share information, JPL management needs to have an overview of DSN operations. This overview requires the integration of data from separate sources within the DSN. JPL might need to know, for example, which stations have completed installation of ECOs for a given ECR, or it might need to compare the cost of the installation from one site to another. The ability to discover unknown relationships is also desirable. There is currently no easy way for management to determine if, for example, a particular piece of equipment causes the same problems at each site where it is used, or for two sites with the same problem to benefit from each other's experience.

To overcome, in part, this problem of management's difficulty in deriving composite information from various sources within the DSN, a pilot system is currently being developed at JPL, which will provide integrated data concerning DSN operations, maintenance, and repairs at Goldstone (Ref. 5). This system, called the Productivity Information Management System (PIMS), will provide its users, both managers and management scientists at JPL, a set of tools for manipulating data in a variety of ways. Management scientists will then have the capability to develop and verify operations research models. The implementation of efficient operational policies can then lead to substantial savings and cost reductions. Because the data that management needs is decentralized, and stored in different forms, using different overall methodologies, a major integration effort such as PIMS is presently the only way to provide the overview so badly needed.

2. The need for autonomy. In addition to the need for sharing information, there is a conflicting need for activities to remain autonomous. The various DSN management activities are separate and distinct applications. They have developed, and continue to develop, independently of each other, and are under autonomous local control. Integrating the data from all of these activities into a single centralized database is restrictive. Local control of the data is an important aspect of the DSN, as is the independence of one activity from another. It is, therefore, neither desirable nor practical to develop a specification of the totality of operational data for the DSN and to design a logically centralized database.

3. The need for evolvability. Coupled with the need for autonomy is the need for evolvability. Administrative activities evolve with the growth of the organization. Some functions are replaced, others are added. The information management of these activities must be capable of evolving also. The database must at all times be an accurate reflection of the organization. It must therefore be dynamic, capable of changing and growing as the DSN changes and grows.

Evolvability of the information management system is important for financial reasons also. Funding comes not all at once, but in small increments over time. The information management system must be capable of incremental development.

III. Database Management Systems

In this section we consider available choices for database management systems in the 1980s. We begin with a brief discussion of data models. We then present a historical development and description of database management system architectures.

A. Data Models

A data model is an abstract representation of the information content of the database. As such, its main function is to insulate the user from the implementation details of the database. Typically, the data in the database is represented using a "conceptual" schema, which is an instance of a given data model. (The relationship of database schema to data model is analogous to that of a program to a programming language.) The data model provides both data structures for representing data and operations for manipulating them. The three best known data models are the hierarchical model, the network model, and the relational model. We now give a brief description of each of these, and site some of the more well-known implementations of each.

1. Hierarchical data model. In the hierarchical data model, the data are represented using trees and links. One designated record type occupies the top node of the tree, while its dependent record types are at nodes on lower levels of the tree. The links connect occurrences of these records. These structures model one-to-many relationships, since every dependent record can have at most one parent record. As an illustration of the use of this model, let us consider the canonical example of suppliers and parts. To represent the relationship of suppliers to parts supplied we would have a forest of trees, with a particular supplier at the top of each tree, and the parts supplied by that supplier at the nodes on the next level.

Some of the longest established database management systems adopt the hierarchical approach to data organization. These include the Information Management System (IMS) of IBM, System 2000 of MRI, and Mark IV of Informatics.

2. Network data model. Many of the relationships inherent in a database are not one-to-many, but many-to-many. To capture these kinds of relationships a more general structure, called a network, was introduced. A network can be viewed as a graph containing nodes and bidirectional links. Although this allows more flexibility than the hierarchical model and is more efficient in some cases, it is considered more complex.

The most important example of network systems is provided by the proposals of the CODASYL Data Base Task Group, DBTG. Two commercial systems based on these proposals are DMS 1100 by Univac, and IDMS by Cullinane. Other network systems include TOTAL by Cincom, an IDS by Honeywell.

3. Relational data model. In the relational model data are organized into tables, called relations, which closely correspond to traditional files. The rows of a table correspond to the records of a file, and the columns correspond to the fields of a record. Associations between the rows are represented solely by data values in columns drawn from a common domain, or pool of values. All of the information in the database, entities as well as relationships, is represented in a single uniform manner, namely in the form of tables. This uniformity of data representation results in a corresponding uniformity and simplicity in data operations.

In contrast to the hierarchical and network models there are no interrecord links in the relational model. This feature gives the relational model an independence from the underlying physical realization of the database. The physical dependence of the hierarchical and network systems stems from the encouraged association between the physical access paths and the logical interrecord links. The absence of such links gives the relational model an added degree of flexibility.

Relational systems are historically the most recent. Some of the better-known relational systems are SQL/DS (System R) and Query-By-Example from IBM, INGRES from Relational Technology, NOMAD from National CSS, ORACLE from Relational Software, and ENCOMPASS from Tandem.

4. Semantic data models. In addition to the three commercially available data models described above, there have recently been developed some higher-level models which allow the meaning, or semantics, of the database to be incorporated more completely into the schema. These models differ from the record-oriented models above by employing constructs that are more user oriented, such as objects, types of objects, and attributes of objects.

There are many semantic models currently in use, but their usage is mainly academic. That is, there are no direct implementations of any of them as products. Some of the more well-known of these models are the Entity-Relationship Model (Chen 1976), the Semantic Database Model SDM (Hammer and McLeod 1978), the Extended Relational Model RM/T (Codd 1979), and the Event Database Model (King and McLeod 1981).

B. DBMS Architectures

It is useful to classify databases according to whether they are logically and physically centralized or decentralized. Using this framework, four classes of data base architectures can be identified. Logically centralized and physically centralized databases include conventional integrated databases. Logically centralized and physically decentralized databases include "distributed databases", as well as a number of recent approaches to composite database support. Logically decentralized and physically centralized or decentralized databases are the domain of federated databases.

1. Logically centralized, physically centralized systems. Nearly all database management systems in use today, including all of those mentioned in Section III-A, manage databases that are both logically and physically centralized. This means that a single conceptual schema, derived from a formal data model such as the ones mentioned above, is used to structure all of the data in the database. It also means that all of the data in the dat____ are stored in one location. There are three main reasons for integrating data thusly, from separate sources and varying applications, into a unified, coherent whole. One reason is that duplication of the data from one source to another is greatly reduced. The second is that the data becomes logically and physically independent of the application programs that use it. This means that the physical details of data storage and access methods are handled by the system. The

entire collection of data in the organization is now an important resource, easy to access and use for a variety of diverse applications. The third reason is that the data are now under control of a centralized authority, who makes decisions for the good of the organization as a whole rather than any one application.

2. Logically centralized, physically decentralized systems. The advantages of integrated databases were widely recognized. However, in some applications, the organization itself is geographically distributed. Having the data stored in one central location means high communications costs and degraded system performance. Therefore, the next step taken in DBMS research was to take the data in a logically centralized database and physically distribute it among the various nodes of a computer network. This physical distribution is totally transparent to the user of the system. That is, to the user of the database it is as if all of the data were in one place. The system's performance is improved because the data is located where it is accessed most frequently. Distribution of the database to optimize parallel processing becomes a key design issue for distributed systems. Another key feature of distributed systems is the possibility of increased reliability. A company can reduce the disaster of a computer failure by duplicating the data at more than one site. These features of distributed systems make them highly desirable for many of today's application environments. Therefore, much research and development on distributed systems is currently taking place. Added complexities involving consistency of redundant data, recovery from a failure at any site, and control of concurrent processing pose some difficult research problems. Prototypes have, however, been built, most notably SDD-1 by Computer Corporation of America. It should not be long before a distributed DBMS will be commercially available.

3. Logically decentralized systems. Both conventional and distributed systems, though they differ in their physical realization, are logically the same. A single conceptual schema defines all of the data in the database, and the control of the database is centralized, even though the data may not be. This can pose, and has posed, some problems. It has been, in some environments, very difficult to integrate data from many applications because the views they have of the data are different. Logical centralization can force the coupling of data where the retention of some individual autonomy is desirable. Each user of a centralized system is forced to surrender the control of the structure of his data to a central authority, who has the task of organizing all of the parts into a coherent whole. This can have drawbacks. Many individuals are very reluctant to relinquish control of their data, so much so that many an attempted database effort has failed for this reason. Even where this is not the case, centralized control often creates a large bottleneck through which all

requests for change must pass. Changes, therefore, occur reluctantly and slowly, resulting in inaccuracies and anachronisms in the database. In addition, the job of the central authority is an enormous one, for this person (or persons) must understand every aspect of the organization thoroughly in order to model the data well, and must also have a thorough knowledge of DBMS software. The database administrator(s) must choose a design for the system which optimizes usage for the whole collection of users, a design which, however, is often much less than optimal for any one user. Thus, the benefits of integration can have a very high cost.

The notion of a federated database architecture was introduced to remedy these problems. A federation is a union of two or more logically decentralized sources of data which may be, but need not be, physically decentralized. The essential difference from the systems above is the logical decentralization. The individual components of a federation remain under autonomous local control with, however, a certain amount of sharing and coordination. One component of the federation is distinguished as the federal controller. It keeps track of the topology of the federation, and aids in the entrance or departure of a component into or from the federation. The components themselves, through the communications facilities provided by the federation, define the system and negotiate their interactions. Each component has its own schema, which states which of its data is private and which is to be shared in the federation. Individual members of the federation may change internally so long as their interface to the federation is maintained. The federated architecture is both dynamic and modular, with components coming and going at any time. It therefore carries with it all of the well-known benefits of dynamic modular systems.

The research on federated systems is relatively new, and to date there is only a small working prototype at the University of Southern California. However, as a compromise between total integration on the one hand and total autonomy on the other, it is highly desirable for many of today's applications. In addition, the trend today away from large mainframe computers toward networks of smaller machines makes the federated approach to database organization particularly appropriate

IV. A Database Management System for the DSN

In this section we bring together the requirements of Section II and the system characteristics of Section III to recommend a system for the DSN. We then describe in some detail the nature of this system.

A. A Federated System for the DSN

In choosing an architecture for the DSN we must satisfy the three previously stated requirements. These are (1) applications must be able to share data and activities; (2) applications must retain individual autonomy and control of their data; and (3) applications must be able to change with time.

Logically centralized systems fail to meet the second requirement. If we were to adopt a centralized database architecture for the DSN, all of the data from all of the applications would have to be under centralized control. As we have seen from the examples in Section II, this is impractical.

Logically centralized systems also do not meet our third requirement very well. Because at any one time the totality of the database must be represented in a single logical schema, changes in the database require a redesign of the schema.

The characteristics of federated databases, on the other hand, seem to be perfectly matched to the needs of our DSN environment. Federations allow for local autonomy, while facilitating the sharing of data and activities. Federations are also capable of evolving over time. Let us take a closer look at what a federated information management system for the DSN would be like.

1. **The topology of the federation.** The components of a federation are the logically autonomous units of an organization that sometimes need to share data or jointly to perform some action. In the case of the DSN, these components are the various administrative entities described earlier, such as Engineering Change Management, Equipment and Materials Management, Anomaly Reporting, Repairs, Cabling, etc. A distinguished component, which can be distributed among the sites or be resident at JPL, is the federal dictionary, whose task it is to record the topology of the federation. The dictionary acts in establishing, maintaining, and terminating the federation, as well as in monitoring structural changes.

Each application that needs autonomy, whether it be ECM or Repairs or anything else, will be a logical component. Some of these components will reside on the same machine, others may reside at other NIM nodes, while still others may be distributed throughout the NIM computer network. The degree and nature of sharing and cooperation among these components will be expressed in the component schemas. The federation provides an integrated set of intercomponent communication facilities. These are data importation for data

sharing, message passing for transaction sharing, and negotiation for cooperative activities.

2. The component schemas. Each component of a federation is a logical entity having its own component schema. This schema describes the information of concern to the component and has three parts: an export schema, which specifies the information it is willing to share with other components; an import schema, which specifies the information in the federation that the component wishes to access; and a private schema, which specifies local information, which the component is unwilling to share at all.

The export schema for the ECM component would likely contain most of its data, since ECM is a network-wide activity. Its import schema would contain the items exported from the equipment database, cabling database, anomalies database, and possibly others. Other components, such as Repairs, would have a larger private schema while exporting relatively less information. The actual content of these schemas will be decided through the negotiation mechanisms of the federation according to the needs of the components.

It is highly possible that one component importing data from another will need to have a different view of the data. The federated model also provides operators for deriving both types and attributes. This means that is is not necessary for components to agree on a common view of the data for sharing to take place.

The federal dictionary component is the repository of information global to all components, which includes information describing the structure of the federation. Its import schema is used to collect this global information, while its export schema is used to share it with the other components. Any component of the federation can find out from the dictionary what components currently constitute the federation, and how it may communicate with them, as well as obtain a summary of the kinds of information available.

3. The data model. The federated architecture requires a common data model to be used throughout the federation, although a component may use any data model of its choice for internal use. Each component uses the federation model to define its export, import, and private schemas. The federal dictionary component uses the model to define the structure of the federation.

While it is poss   e that any data model could be made to work as the federation model, a semantic model, such as the Event Database Model (Ref. 4) is preferable because many kinds of relationships between the data can be represented. In

addition, since the model is not tied to any particular physical representation, the underlying physical implementation of the database can change without _ .ecting its logical expression.

If the logical components of the federation use a different model than the federation model, a translation can be made between the two. This is important if components are to be managed with DBMS software commercially available today. Because of the simplicity and structural independence of the relational model, it is the best commercial choice available today for the components to use.

B. Evolving the Federation

One of the advantages of adopting the federated approach to database organization is that the database can be developed incrementally. Components can come into or depart from the federation at any time. A component can also change internally, so long as it supports its interface to the federation. This evolvability is particularly appropriate for the DSN, as funding is easier to obtain in increments. The federation can grow both within the NIM system and beyond.

1. Within the NIM. The extendibility of the federated architecture means that as the NIM communications and hardware expands, so does the federated information management system. The federation for NIM can evolve from the components themselves. They will each express their own export, import, and private schemas, and will use the negotiation mechanism of the model to achieve a desirable configuration. This configuration need by no means be static. Components can negotiate for their entrance or removal from the federation, as well as restructure themselves internally. This means that as new applications are added to the NIM system they can easily become a part of the federation, and assures that the federation will always be an accurate model of the application environment. The basic lines of autonomy and the patterns of interaction are the governing design principles to be embodied.

2. Beyond the NIM. The federated architecture can also be extended to include components outside the NIM assembly. This is particularly desirable, for the DSN needs to interact with various JPL institutional systems from time to time. These include institutional systems for financial management, personnel management, property management, work scheduling, mission planning, etc. A higher-level federation, one with the NIM federation as one component (the DSN component) along with these other institutional systems, can be envisoned. The principles of design are the same. All that is needed is the necessary hardware and communications to link them together.

V. Comparison of an Alternative Choice

In this final section we consider a proposed alternative to a federated system. In our evaluation, we focus on two basic features. The first of these is the desirability of a general-purpose vs a special-purpose system. The second is the desirability of a dynamic vs a static system. Also, the life-cycle costs of both choices must be considered.

The database system now under consideration for the DSN is a system of separate, independent databases for each application on the NIM. This approach is an electronic counterpart of the situation that exists now, and can be achieved with little research or development effort. As before, each application will own its data. However, because of the communications provided by the NIM network, any application will be able to peruse the data from any other application's database. Nevertheless, one application will not be able to use the data in another database without writing a program to incorporate that data into its own system. The imported data will have to be duplicated, interpreted, and restructured before it can be used. Composite information will still be extremely difficult to obtain. This is because each application will have data in different and incommensurate forms. The task of utilizing these different views of the data is not trivial, and standardizing these views is tantamount to centralization. In addition, keeping redundant copies of data creates a problem of consistency with updates that must be dealt with. The costs of duplicate storage space, and of time to transmit copies back and forth across the network must also be considered.

While it is true that the interactions provided by the federated model can be realized on a case-by-case basis by means of ad hoc application programs, this can become extremely costly in the long run. For, as the number of components increases, the cost of application software to tie them together grows geometrically, whereas the cost of the federated software stays the same. The federation provides a general-purpose system for maintaining autonomy while facilitating sharing.

The distinction between the ad hoc alternative approach and the federated approach is the distinction between generality and flexibility on the one hand, and specificity and rigidity on the other.

Also, if it were possible to state at any one time all the ways in which the data are to be shared amongst the users of the NIM then one could implement the necessary programs to do this. However, obtaining such a specification is unrealistic. Changes are a fact of life, and the ability to respond to changes is a highly desirable feature, saving much cost over the years. Only a federated system offers the ability to change these interapplication, intercomponent relationships dynamically. Therefore, it is the life-cycle cost of each alternative that must be compared. The added time, effort, and dollars necessary to implement the federated information management system, from first principles, is more than likely to pay for itself as time goes by.

VI. Conclusion

In this paper we have examined several key DSN administrative functions. We have seen how these activities need to have a data management system which will allow them to retain their individual autonomy and which will also allow them to share data. We have also seen that these activities need to be able to grow and change independently of each other. They therefore require a data management system that is dynamic.

We feel that the federated approach to database organization is particularly appropriate to this situation. We also feel that the benefits of implementing it far outweigh the costs. The development of a federated information management system is an ambitious undertaking, but one worthy of such an important organization as the DSN. The results are very exciting to contemplate.

Acknowledgments

References

1. "Deep Space Network Subsystem Functional Design: Deep Space Network Information Management (NIM) Assembly," Aaron-Ross Corporation, Glendora, Calif., 835-2, Feb. 1982.

2. "NIM Fourth Progress Review: Level 'D' (Backup Information)," Aaron-Ross Corporation, Glendora, Calif., Feb. 1982.

3. "NIM Program Specification Document," Aaron-Ross Corporation, (endora, Calif., Feb. 1982.

4. King, R., and McLeod, D., "The Event Database Specification Model (A Preliminary Report)." In *Database Modelling*, editors M. L. Brodie. J. Mylopoulos, and J. Schmidt, 1982 (to be published).

5. Hull, R. "An Introduction to the New Productivity Information Management Systems (PIMS)," *TDA Progress Report 42-70*, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1982.

Bibliography

Aaron-Ross Corporation, "Deep Space Network General Requirements and Plans: DSN Network Information Management (NIM) Assembly (1982 through 1989)," 821-6, Revised Dec. 1981, Glendora, Calif.

Date, C. J., *An Introduction to Database Systems*, Addison-Wesley, 1981.

Hammer, M., and McLeod, D., "On the Architecture of Database Management Systems." In *Infotech State-of-the-Art Report on Data Design*, 1980.

Heimbigner, D., and McLeod, D., "Federated Information Bases (A Preliminary Report)." In *Infotech State-of-the-Art Report on Data Design*, 1981.

Heimbigner, D., and McLeod, D., "An Approach to Logical Database Decentralization." Computer Science Department, University of Southern California, Los Angeles, Calif., Dec. 1981.

King, R., and McLeod, D., "Semantic Database Models," *Database Design*, editor S. B. Yao, 1981.

King, R. and McLeod, D., "A Semantics-Based Methodology For Database Design, Use, and Evolution." Computer Science Department, University of Southern California, ' os Angeles, Calif., Dec. 1981.

McLeod, D. and Heimbigner, D., "A Federated Architecture for Database Systems." In *Proceedings of the National Computer Conference*, AFIPS, pp. 283-289, Anaheim, Calif., May 1980.

CONSCAN Implementation for Antenna Control Assembly

J. E. Ohlson

Stanford Telecommunications Inc.

and

K. P. Abichandani

DSN Engineering Section

CONSCAN was previously recommended for implementation in the Antenna Control Assembly. This article presents specifics of this implementation, including calibration, signal cleanup, and system protection. Equations for programming the algorithms are provided.

I. Introduction

Reference 1 discussed the merits of several automatic tracking techniques, and concluded that CONSCAN (conical-scan tracking) was most desirable for DSN antenna application.

Additional effort was necessary to develop the equations and algorithms required for the implementation of CONSCAN. This article provides that detail as the final phase of the study involving automatic tracking techniques for the Antenna Control Assembly (ACA).

This report presents the detailed analysis and instructions for implementing CONSCAN into the ACA.

II. CONSCAN Algorithm

Scan deviation may be defined as deviation of the antenna beam axis from the antenna boresight axis. The deviation is resolved into two orthogonal components called elevation and cross-elevation. Let

$E(k)$ = elevation deviation

$X(k)$ = cross-elevation deviation

$A(k)$ = azimuth deviation (or hour-angle deviation)

E_{ANT} = antenna elevation (or declination)

A_{ANT} = antenna azimuth

To maintain cross-elevation deviation constant with elevation, the azimuth deviation must vary with elevation angle:

$$A(k) = X(k) \sec E_{ANT} \qquad (1)$$

From Fig. 1, the scan equations are

$$E(k) = R \sin \omega k \Delta \qquad (2)$$

$$X(k) = R \cos \omega k \Delta \qquad (3)$$

where

ω = scan rate in rad/s

$\quad = 2\pi/P$

P = scan period, s

R = scan radius, deg (Fig. 1)

k = no. of units of Δ (constant)

Δ = time between updates (for data samples)

Hence, having chosen the time between updates, and indexing time by the number k, the scan deviation may readily be computed.

III. Antenna Pointing Angle Equations

The components of the antenna pointing angles, E_{ANT}, and A_{ANT} are calculated for two cases, CONSCAN *OFF* and *ON*. (Refer to Fig. 2)

(1) With CONSCAN *OFF*,

$$E_{ANT}(k) = E_{PRED}(k) + E_{TAB}(k) + S_E \qquad (4)$$

$$A_{ANT}(k) = A_{PRED}(k) + A_{TAB}(k) + S_A \qquad (5)$$

where the subscripts are

ANT = actual antenna command

$PRED$ = predicted value from ephemerides

$\quad TAB$ = value from systematic error correction table

and where,

S_E = total correction for elevation

S_A = total correction for azimuth

$\quad = S_X \sec E(k)$

S_X = total correction for cross-elevation

(2) With CONSCAN *ON*,

$$E_{ANT}(k) = E_{PRED} + E_{TAB}(k) + S_E + E_C(k) \qquad (6)$$

$$A_{ANT}(k) = A_{PRED} + A_{TAB}(k) + S_A + A_C(k)$$

$$= A_{PRED} + A_{TAB}(k) + S_A$$

$$+ X_C(k) \sec E_{ANT} \qquad (7)$$

where E_c and X_c are the latest (kth) correction in elevation and cross-elevation, respectively. The predicts for azimuth and elevation are calculated from the spacecraft ephemeris or from the known location of a radio star. The Correction Table is a set of stored corrections and provides a first-order correction to previously measured systematic errors.

IV. Coordinate Correction Algorithm

Reference 2 derives the expressions for corrections in each coordinate. These corrections are given in integral form, but are modified to conform to digital processing requirements by expressing them in summation form as follows,

$$E_C = G \sum_{SCAN \, k} V(k) \sin(\omega k \Delta + Z) \qquad (8)$$

$$X_C = G \sum_{SCAN \, k} V(k) \cos(\omega k \Delta + Z) \qquad (9)$$

where $V(k)$ is the kth edited signal sample,

G = gain

Z = phase shift

The phase shift Z is required because of mechanical phase lag of the antenna response and the phase lag of the AGC or radiometer averaging. The gain G and the phase shift Z must be calibrated for each antenna configuration and signal source.

The values of E_C and X_C may be determined by one of the following approaches:

(1) Brute Force correlation.

(2) Fast Fourier transform.

The Brute Force approach is simple and is indicated by the nature of expression. All that is needed is use of a sine and cosine lookup table, multiplication, and accumulation.

The fast Fourier transform (FFT), although it requires more computations, provides more information as a bonus.

Analytically, the resultant correction parameter Y, having two components X_C and E_C, is expressed as

$$Y = X_C + jE_C \qquad (10)$$

where j is the square root of -1. If we use Eqs. (8) and (9) and DeMoivres theorem,

$$Y = G \sum V(k) e^{j(\omega k \Delta + Z)}$$

$$= Ge^{jZ} \sum V(k) e^{j\omega k \Delta} \qquad (11)$$

This is the $Y(1)$ term of the discrete Fourier transform, generally computed by the FFT. The transform is

$$Y(N) = Ge^{jZ} \sum V(k) e^{jN\omega k \Delta} \qquad (12)$$

The term $Y(1)$ gives pointing error. All terms for $N > 1$ are a measure of anomalies. For example, ellipticity in the antenna beam will give a very small but finite contribution to $Y(2)$. Normally, all $Y(N)$ for $N \geqslant 2$ will be very small because $V(k)$ is usually nearly constant. When $V(k)$ has a glitch due to receiver dropout, interference, or a step change of signal such as a spacecraft transmitter mode change, then $Y(N)$ for $N \geqslant 2$ can be much larger than $Y(1)$. Thresholding can be done on a few values of $Y(N)$ to detect anomalous conditions in order to reject potentially absurd calculated "corrections."

The number of signal samples per scan must be chosen to be a power of 2 in order to directly use the FFT. For example, the conveniently sized 512-point FFT provides a scan of 51.2 s when a signal sample rate of 10 Hz is used.

The steps followed in using the FFT are:

(1) Using the $V(k)$ values for a scan, calculate $Y(N)$ by performing the above summation (Eq. 12), for several values of N, say 2, 3, and 4.

(2) Compare each $Y(N)$ thus computed to preselected threshold. If the threshold is exceeded, the correction for this scan is rejected. The threshold that is chosen should be small; it should be on the order of the scan radius. However, it should be large enough to allow the second harmonic generated as a result of elliptic cross section of the beam to be ignored. The rejection rate due to glitches and other spurious content of $Y(N)$ should be less than 5 percent.

V. Tracking Signal Cleanup and Protection During Track

Reference 1 proposed signal cleanup and protection schemes both for radio sources and spacecraft.

The signal inputs are (1) square law detector output for radio sources and (2) AGC (Automatic Gain Control) voltage for spacecraft tracking. The anomalies that are likely to occur that will affect the received signal power may e listed as follows:

(1) Signal dropout due to momentary receiver dropout (out-of-lock), operator error or some unexpected transient (glitch) in the receiving system.

(2) Change in spacecraft transmitter or antenna mode.

(3) Spacecraft antenna pointing direction change with spacecraft limit cycling, or some other change on board the spacecraft causing variation in the downlink signal strength.

(4) In the case of spin-stabilized spacecraft, modulation in the signal level due to spin rate.

Testing is a very essential part of signal processing. Since the data processing requirements are low and the penalty for spacecraft data loss is high, tests should be made as frequently as possible.

When tracking radio sources, a single receiver dropout should cause a scan correction to be skipped. Further, expected signal level should be compared to observed AGC level as another check. AGC can be calibrated at known signal levels, and/or by feeding the expected signal level to ACA. The apparent signal level is continuously compared with the expected (predicted) signal level, and, if deviation is greater than 1 or 2 dB, that scan is rejected.

However, if there is a slow cycling of signal level, a straight line connecting the last and the first point may be subtracted to determine the difference in signal level.

Hence, ignore scan if

$$| V(k) - \text{expected } V(k) | \geqslant \text{threshold} \qquad (13)$$

This is the condition of *out of lock*; if cycling (low) is expected, the straight line test referred to above is analytically expressed as,

$$\left| V(k) - \frac{V(k_o) - V(0)}{(k_o)} k - V(0) \right| \geqslant Th \qquad (14)$$

where

Th = threshold (1 or 2 dB)

k_o = number of points in scan

For testing of radio source tracking anomalies, the system temperature and the expected radio source temperature should be provided so a check can be run. This will entail a precise setup for the radiometer gain to ensure accurate comparisons. A table of temperatures for common radio stars could be easily stored into the system memory.

If T_{op} is the system operating temperature and T_s the radio source (expected) temperature, the scan should be ignored if

$$\text{scaled } [V(k)] < T_{op} \tag{15}$$

or

$$\text{scaled } [V(k)] > T_{op} + T_s + \text{margin} \tag{16}$$

There is the very real likelihood that, even though signal level is continuously and carefully tested, an unusually large correction may sneak through. Thus limit tests become necessary. These may be categorized as first limit and second limit.

(1) First limit: If the desired correction exceeds the first limit, correction only to the extent of the limit value is to be made. Examples of limits that could be set as first limits on elevation and cross-elevation are

(a) For S-band, 0.015 deg

(b) For X-band, 0.005 deg

If the next correction has to be the same limit (because it is equaled or exceeded), no correction should be made; instead, a warning is generated.

(2) Second limit: If the desired correction exceeds the second limit, no correction is made; instead, a warning is generated. Example limits for elevation and cross-elevation that could be set are

(a) For S-band, 0.030 deg

(b) For X-band, 0.010 deg

VI. Antenna and Signal Source Gain and Phase Calibration

Referring to Eqs. (8) and (9), we see the gain parameter is given by G. At this time, an expression for G will be derived to enable calibration.

The closed loop time constant T is given by (Ref. 2)

$$T = -P/\ln r \tag{17}$$

Define

$$r = 1 - hA \tag{18}$$

where h in Ref. 2 has been defined as "selectable" gain and $P = 2\pi/\omega m$.

A is a parameter readily measurable by use of expressions 11 through 21 (Ref. 2). Equation (19) of Ref. 2 expresses it as (for radiometer)

$$A = CBT_s g'(R)P/2 \tag{19}$$

where C is a constant representing receiver gain, and B is the effective band width:

$$B = \frac{\left[\int_0^\infty |H(f)|^2 \, df\right]^2}{\int_0^\infty |H(f)|^4 \, df} \tag{20}$$

with $|H(f)|^2$ the power gain, and

$$g'(R) = \frac{dg(\beta)}{d\beta}\bigg|_{\beta=R} = \frac{d}{d\beta} e^{-\mu\beta^2/w^2}\bigg|_{\beta=R}$$

$$= -2 \frac{\mu R}{w^2} e^{-\mu R^2/w^2}$$

$$= -2 \frac{\mu R}{w^2} \tag{21}$$

R as used above is the scan radius and w is the antenna beamwidth between half power points.

$$\mu = 4 \ln 2 = 2.773 \tag{22}$$

From Eq. (18) above,

$$\ln r = \ln (1 - hA) \cong -hA \tag{23}$$

substituting in Eq. (17),

$$T = P/hA \tag{24}$$

Thus, expressing $G = h\Delta$ after dropping the negative sign,

$$G \text{ (radiometer)} = \frac{\Delta w^2}{\mu CBT\ T_s R} \qquad (25)$$

For ς acecraft, Ref. 2 (Eq. 101) has

$$A = \frac{k'P\mu R}{2w^2} \qquad (26)$$

where k' is found by calibration of the AGC slope.

By using Eq. (25) and $G = h\Delta$,

$$G \text{ (spacecraft)} = \frac{2\Delta w^2}{k'\mu TR} \qquad (27)$$

k' may be calculated by using the derivation for fast AGC in Ref. 3, where in Eq. (12), the expression to be used is

$$V_{AGC} = \frac{1}{\alpha} \ln v(t) \qquad (28)$$

where $v(t)$ is the normalized voltage gain (Eq. (63), Ref. 2),

$$\alpha = \frac{1}{k'} \qquad (29)$$

with R^2, θ^2, $\phi^2 \ll w^2$ (Eq. (63), Ref. 2) using Eq. (91) of Ref. 2, and Eq. (26) above,

$$E_C = h \frac{\mu RP\, \phi_k}{2\alpha w^2} \qquad (30)$$

α is given in Eq. (3) of Ref. 3, and ϕ_k is the symbol for elevation (cross-elevation) deviation in the kth scan.

$$8.686\, \alpha = \text{slope in dB/volt} = S$$

or

$$k' = 1/\alpha = 8.686/S \qquad (31)$$

Equation (27) above can be rewritten by substituting k' from Eq. (31) above, and $\mu = 2.773$

$$G = 0.0830 S \frac{\Delta w^2}{R} \qquad (32)$$

We will now discuss phase-shift calibration. Equations (8), (9), and (11) above include the phase term Z. The phase shift is needed because

(1) For radio source tracking, filtering in the radio meter causes a phase lag.

(2) There is a filtering effect of AGC, during spacecraft tracking.

(3) Phase lag of the antenna is a function of the scan period, e.g., for a 64-meter antenna (Ref. 2) and

(a) For a 28-s scan, $Z = -30$ deg

(b) For a 58-s scan, $Z = -15$ deg

The required value of Z may be measured by the following means:

(1) Assuming a nominal value of Z, CONSCAN the antenna to achieve boresight.

(2) Open the control loop, and offset the antenna in one axis only, say cross-elevation.

(3) If the nominal Z is correct, the correlation and hence the indicated correction will be nonzero in only the offset axis.

(4) If Z is incorrect (i.e., if correlation is nonzero in both the axes), the correction for Z for the case of cross-elevation offset is:

$$\Delta Z = \tan^{-1} \frac{\text{elevation correction}}{\text{cross-elevation correction}}$$

$$= \tan^{-1} \frac{E_C(k)}{X_C(k)}$$

If the fast Fourier transform (FFT) is used for signal processing, the phase angle of the fundamental output $Y(1)$ is exactly ΔZ.

VII. Signal Source Acquisition Policy

Predicted offset from boresight is expected to be within 10 dB (X-band) initially. As this raises the possibility of large anomalies, a spiral scan for acquisition is recommended.

To attain higher accuracy in initial acquisition, three parameters should be established.

(1) Beamwidth B in which acquisition is highly likely.

(2) Time T necessary to make a decision.

(3) Size of the search region.

The procedure for initial acquisition is as follows (Fig. 3):

(1) Dwell at best guess for T seconds.

(2) Move out $B/2$ in one axis.

(3) Dwell at that position for T seconds.

(4) Traverse spiral until the search is complete (the spiral is discussed below).

(5) Set the scanner at the point of maximum signal level. Alternatively, one may stop when any signal is found; but this may be a side lobe.

The spiral search is done by arguing that for every 2π increase in θ there should be an increase in R by B, i.e.,

$$\frac{dR}{B} = \frac{d\theta}{2\pi} \qquad (33)$$

The coordinate system is shown in Fig. 3a.

If the scan rate is $\dot{\theta}$, at R radius of scan velocity is,

$$\text{velocity} = \bar{R}\,\dot{\theta} \qquad (34)$$

so if in T seconds the beam has moved by an amount B, then

$$(R\dot{\theta})\,T = B \qquad (35)$$

solving Eqs. (33) and (35), we have

$$R(t) = \sqrt{R_0^2 + \frac{B^2 t}{\pi T}} \qquad (36)$$

$$\phi(t) = \frac{2\pi}{B}\,[R(t) - R(0)] \qquad (37)$$

where

$$\theta(0) = 0; \quad R(0) = R_o$$

Another maybe simpler way to a solution is to implement Eq. (35) directly. Every T seconds, command a change in θ at the rate

$$\frac{\Delta\theta}{\Delta t} = \frac{B}{RT} \qquad (38)$$

with $\Delta t = T$,

$$\Delta\theta = \frac{B}{RT}\,T = \frac{B}{R} \qquad (39)$$

Using Eq. (33) now and substituting Eq. (39), we have

$$\Delta R = \frac{B}{2\pi}\,\Delta\theta = \frac{B^2}{2\pi R} \qquad (40)$$

The response to commands every T seconds can be prevented from being jerky by choosing T/N intervals instead of T, with N such that smoothness results. Simultaneously, $\Delta\theta$ and ΔR quantities can be attained in N steps.

VIII. Conclusions

This report is intended to be the final one as regards investigation into CONSCAN and treatment of appropriate algorithms for signal cleanup and limit tests to assure its maximum accuracy and desirability.

The treatment described herein is designed to lead to development of software for CONSCAN antenna operation including K band. Summarizing the recommended techniques to be employed in conjunction with CONSCAN:

(1) For initial acquisition, employ stepped spiral search.

(2) Employ continued and incessant testing of ... to verify:

(a) In-lock.

(b) Reasonable in level.

(c) Within limits.

(3) Use FFT for signal processing if feasible, as it provides more tests and insight into existence of glitches and spurious content of the input signal level.

References

1. Abichandani, K. P., and Ohlson, J. E., "Boresighting Techniques for the Antenna Control Assembly (ACA)," *TDA Progress Report 42-64*, May and June 1981, Jet Propulsion Laboratory, Pasadena, California, August 15, 1981.

2. Ohlson, J. E., and Reid, M. S., *Conical-Scan Tracking with the 64-Meter-Diameter Antenna at Goldstone*, Technical Report 32-1605, Jet Propulsion Laboratory, Pasadena, California, October 1, 1976.

3. Ohlson, J. E.. "Exact Dynamics of Automatic Gain Control," *IEEE Transactions on Communications*, Vol. COM-22, No. 1, pp. 72-75, January 1974.
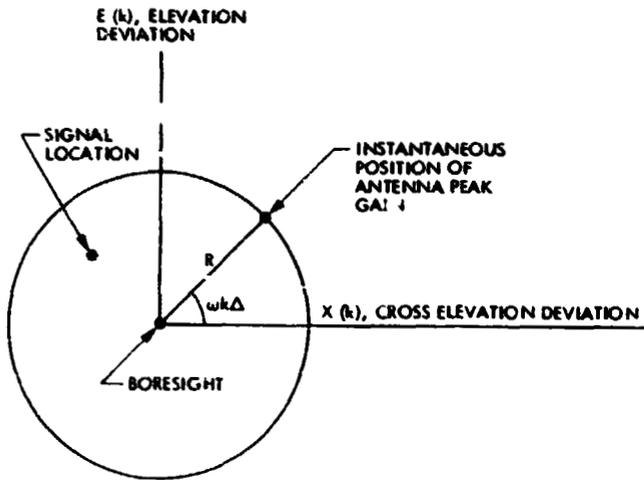
Fig. 1. Scan radius defined
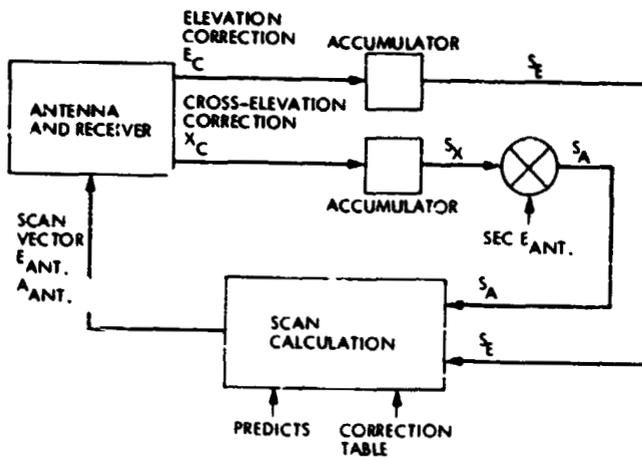
Fig. 2. Block diagram of CONSCAN pointing
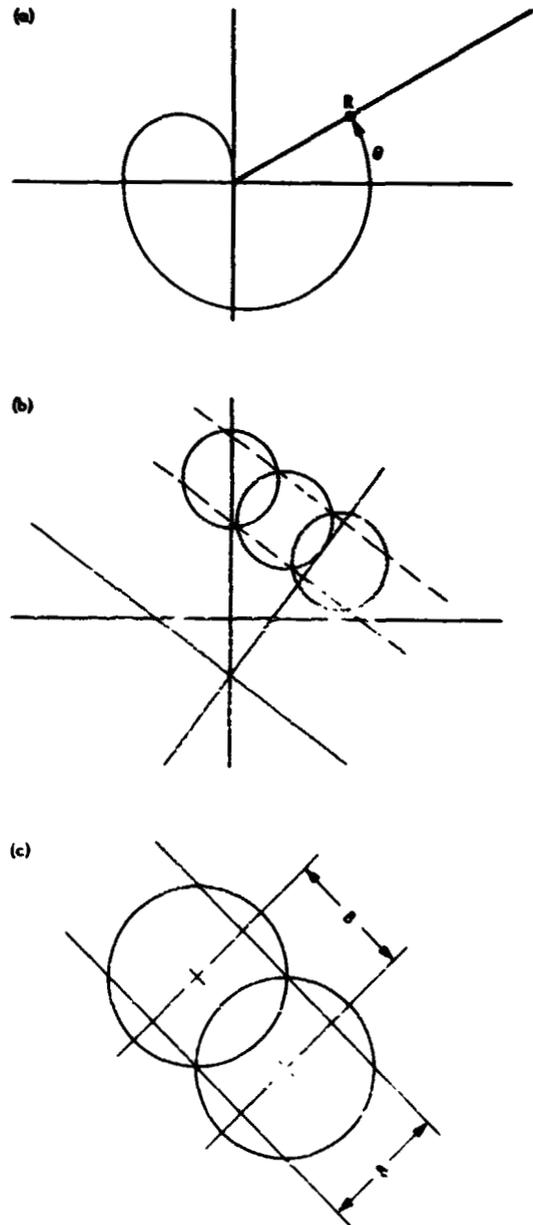
Fig. 3. Spiral acquisition scan geometry: (a) spiral
scanning; (b) N-step scan; (c) scan sweep

GCF Mark IV Development

L. O. Mortensen

DSN Data Systems Section

This article provides an overview of the Mark IV GCF as it is being implemented to support the Network Consolidation Program.

I. Introduction

The Network Consolidation Program (Ref. 1) requires that the Ground Communications Facility (GCF) (Ref. 2) be upgraded in order to support the mission set of the eighties. The key characteristics of the GCF do not change appreciably '- the Mark IV DSN (see Figs. 1-3). The changes made are in the area of increased capacity rather than changing characteristics. Common carrier circuits continue to be the medium for data transfer. The message multiplexing in the Mark IV era differs from the Mark III era in that all multiplexing will be accomplished in a GCF computer under GCF software control vs hardware multiplexing at the NASA Communications (NASCOM) level, similar to the multiplexing currently done in the High-Speed Data Subsystem.

The Signal Processing Center (SPC) assemblies will be controlled and monitored by the SPC Monitor and Control Subsystem while the Central Communications Terminal (CCT) equipments will be monitored and controlled by the GCF Monitor and Control Subsystem. The computers required for the Mark IVA will be composed of the Modcomp II/25 computers currently in use by the GCF; their functions as redefined will require software changes.

II. Digital Communications Subsystem

The Digital Communications Subsystem performs the exchange of digital data between the SPC and JPL using common carrier circuits. Because lease costs of circuits are high, block multiplexing is used to allow sharing of the lines between projects. This subsystem also does data routing either to the proper SPC or the proper user. To maintain high-quality communication, error correction by retransmission will be used in the 56-kb/s duplex line, using an algorithm similar to the current algorithm used with high-speed data. Original Data Record (ODR) data logging will be accomplished at the SPC. At the CCT, the capability for front-end recording will be continued. This subsystem interfaces with the Network Operations Control Center (NOCC) (for which it accomplishes data routing) and with the Mission Control and Computing Center (MCCC), the Remote Mission Operation Center (RMOC), and Project Operation Control Centers (POCCs). This subsystem also routes data to the Data Records Subsystem (see Fig. 4).

The Digital Communications Subsystem is composed of five subassemblies.

A. Area Routing Assembly (ARA)

The ARA at each SPC is composed of two identical computers: one for prime and one for backup. The two computers are to be converted Communication Monitor and Formatter Assembly (CMF) computers (Fig. 5). The ARA assembly will provide regulation and control of data transmission. It will provide routing of data received to the proper link at the SPC. Low-rate data will be multiplexed on the 56-kb/s duplex line. Data will be sent in 4800-bit blocks, though 1200-bit blocks will be available to the GCF users.

B. Station Digital Communications (SDC)

The SDC will consist of data transmission equipment including line interfaces, data sets, modems, NEDs, CBs, digital and analog test equipment, and patch facilities for trouble isolation as well as a front-end line interface between the actual data sets and NED appearances (see Fig. 6).

NASCOM Engineering will provide all data transmission equipment including data sets, modems, line interface circuits, and analog test equipment. The DSN/GCF demarcation point between the DSN/GCF Data Terminal Equipment (DTE) and the NASCOM Data Communications Equipment (DCE) will be at the digital interface to the data set or other such carrier equipment provided by NASCOM.

C. Central Digital Communications (CDC)

The CDC in support of Mark IV-A will consist of data transmission equipment including line interfaces, data sets, modems, NEDs, CBs, digital and analog test equipment and patch facilities for trouble isolation as well as front-end digital line switches between the actual Digital Service Units (DSU) and the NEDs (see Fig. 7).

NASCOM Engineering will provide all data transmission equipment, including data sets, modem, DSUs and common carrier interfaces associated with overseas SPCs. The DSN/GCF interface between the DSN/GCF DTE and the NASCOM DCE will be at the digital (V.35) interface to the NASCOM DCE provided by NASCOM.

As part of the CDC implementation, a new digital line switch will be provided to interface 56 kb/s and/or 224 kb/s lines from each of the SPCs. The switch will provide appropriate interface for 7.2 kb/s data sets as well as WBD (56 or 224 kb/s) line devices.

D. Error Correction and Switching (ECS) Computer

Four Operational ECS computers and one backup will be provided. They will be converted from the present four ECS computers plus one former CMF computer. The prime func-

tions of the ECS computer are those of multiplexing, demultiplexing, error correction, and data routing (see Fig. 8).

E. Network Communications Equipment (NCE) Assembly

One operational NCE computer and one backup will be provided. They will be the present two NCE computers with minor modifications (see Fig. 9). The NCE serves to interface the Network Data Processing Area (NDPA) Real-Time Monitor (RTM) and support computers directly with the Digital Display Processors (DDPs) and with the outside world via wideband data lines (WBDL) to the ECSs. The NCE provides a communication link between the ECSs and the subsystems of the NOCC in the NDPA. It extends the GCF interface to the RTMs and the support processor.

III. Analog Intersite Communications

The GCF microwave presently on line at the Australian and Spanish locations will be decommitted on the completion of the front-end area (FEA) moves to the SPC locations. The radios currently in use between DSS 12 and DSS 11 will be reused (see Fig. 10). The radios at DSS 11 will be decommitted and reinstalled at SPC-10, and the antennas at DSS 12 will be reoriented as required; the DSS 11 antennas will be reinstalled at SPC-10. The multiplex and radio equipment will be configured in standard 8-foot open-channel racks (not cabinets) similar to the existing microwave racks. The additional equipment will be Collins 518-W radios and Collins MX 106 multiplex equipment. The existing path will require additional antennas to connect the additional radios.

The computers at FEA-12, being remote from SPC-10, connect to a LAN interface panel. The interface panel accepts data from the computers and forwards the data via microwave to a companion interface panel at SPC-10. The computer-generated data are reconstituted and forwarded to the SPC LAN. Transmission in the opposite direction is similarly handled. Dual LAN interface panels and microwave channels are used for redundancy.

The LAN interface panel will remote the LAN ports in the SPC to matching LAN ports available to FEA processors. The link will be via microwave and will utilize a 56 kb/s channel, full duplex. The Western Union microwave link (JPL-Goldstone) will be expanded to handle the Mark IV service.

IV. Voice Communication Assembly

The Mark IV voice assembly will be configured from the equipments currently used, namely the tactical intercom,

comm junction module, and station voice switch assemblies. There will be no changes required at the JPL end (GCF-20) (see Fig. 11).

A new communication panel will be developed for use with the new SPC consoles. Because of the advanced age of the Tactical Intercom Assembly (TIC) panels, a new circuit board will be developed to bring the system more in line with current technology.

V. Teletype Communications Assembly

The Mark IV teletype assembly will be made up of the equipments as they are configured at the present time. There are two basic services provided:

(1) 110-baud service using teletype machines with the circuits routed by the NASCOM GSFC teletype (TTY) switch to the WCSC located at GCF-20. (This provides for normal administrative test message service.)

(2) 300-baud service, whereby the station personnel have access to the DSN data base located in the JPL Administrative Computer via an auto dialer located at GCF-20.

VI. Monitor and Control

One operational CCM computer and one backup will be provided. They will be the present two CCM computers, with minor modifications (see Fig. 12).

The GCF monitor and control subsystem is based in the CCM, whose major functions include the collecting, processing,

and displaying of real-time status and performance of the GCF subsystems. The overall monitoring of the CCT will be accomplished via the CCM computer and its associated displays and via the central console. The CCM computer will provide monitoring of the status of all equipment of the GCF.

Computer status reports, which do not deal with configuration but primarily with GCF performance (error rates, traffic flow, etc.), are formatted and forwarded to the CCM as standard GCF data blocks. The CCM will provide displays by means of the Grinnel display converter and the TV switch.

A CCM line printer will be used for DRG IDR gap reporting and summary and post-pass histories. Labels for tapes from the ECS and DRG computers will be printed on label printer terminets. The summary for the data channels and status and alarm messages will be printed on the console terminet.

VII. Data Records Generator

Three operational DRG computers and one backup will be provided. They will be converted from the present DRG lineup. The chief function of the DRG is to provide IDRs that are recorded on magnetic tape. The program of the DRG is the only one that is mission-dependent (see Fig. 13).

The DRG software will check each data stream for correct SPC, spacecraft ID, UDT/DDT, gross data description (GDD) block serial number (BSN), block header time, and error status code. During the pass, the DRG will detect gaps and output real-time gap statistics. These data will normally be output on a CCM display, but may also be printed. At the end of a pass, the DRG outputs a complete IDR report including gap list, via the CCM.

References

1. Yeater, M. L., and Herrman, D. T., "Networks Consolidation Program," *TDA Progress Report 42-65, July and August 1981,* Jet Propulsion Laboratory, Pasadena, Calif., Oct. 15, 1981.

2. Evans, R. H., "DSN Ground Communications Facility," *TDA Progress Report 42-65, July and August 1981,* Jet Propulsion Laboratory, Pasadena, Calif., Oct. 15, 1981.
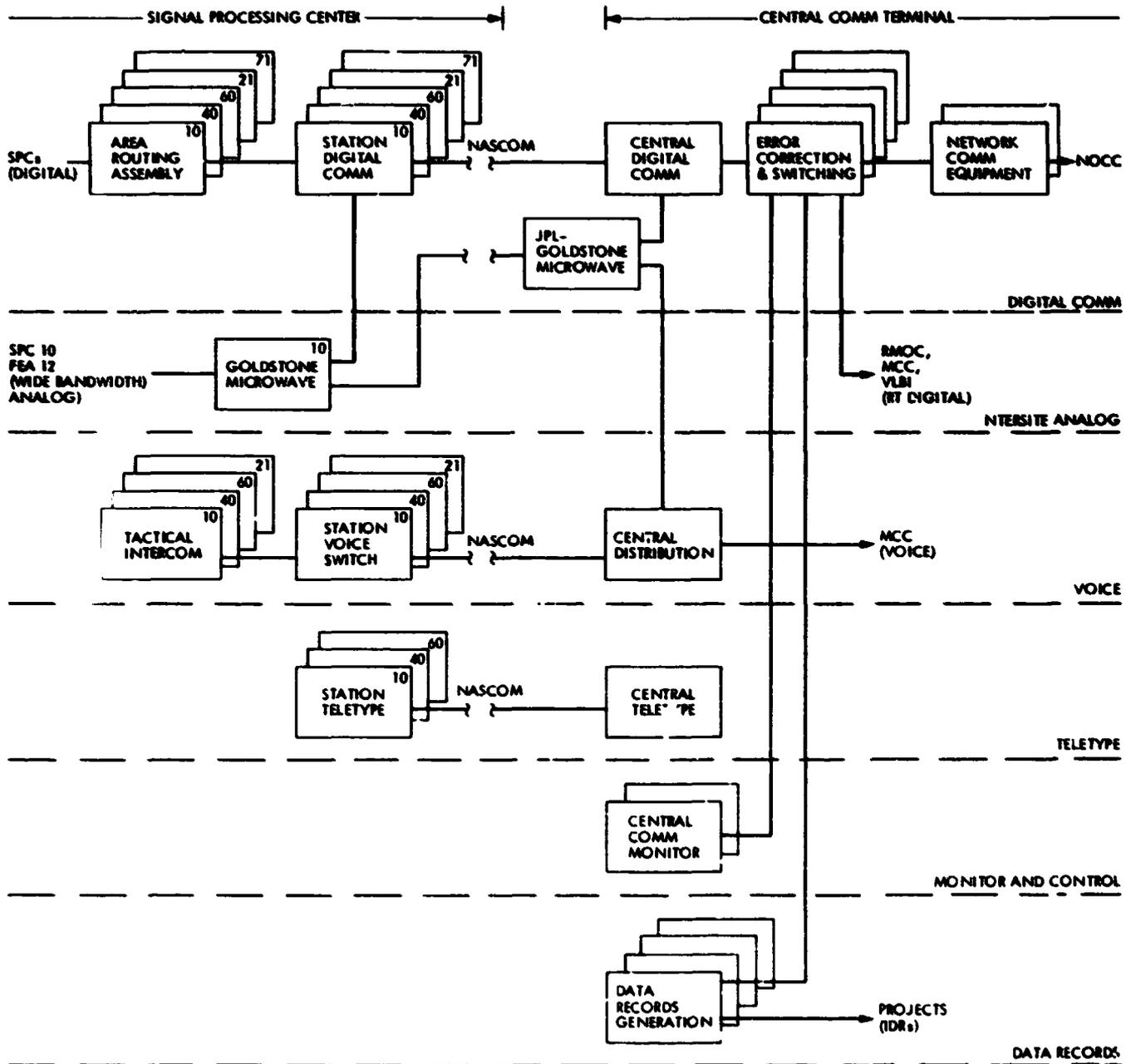
Fig. 1. Mark IV-A era GCF subsystems
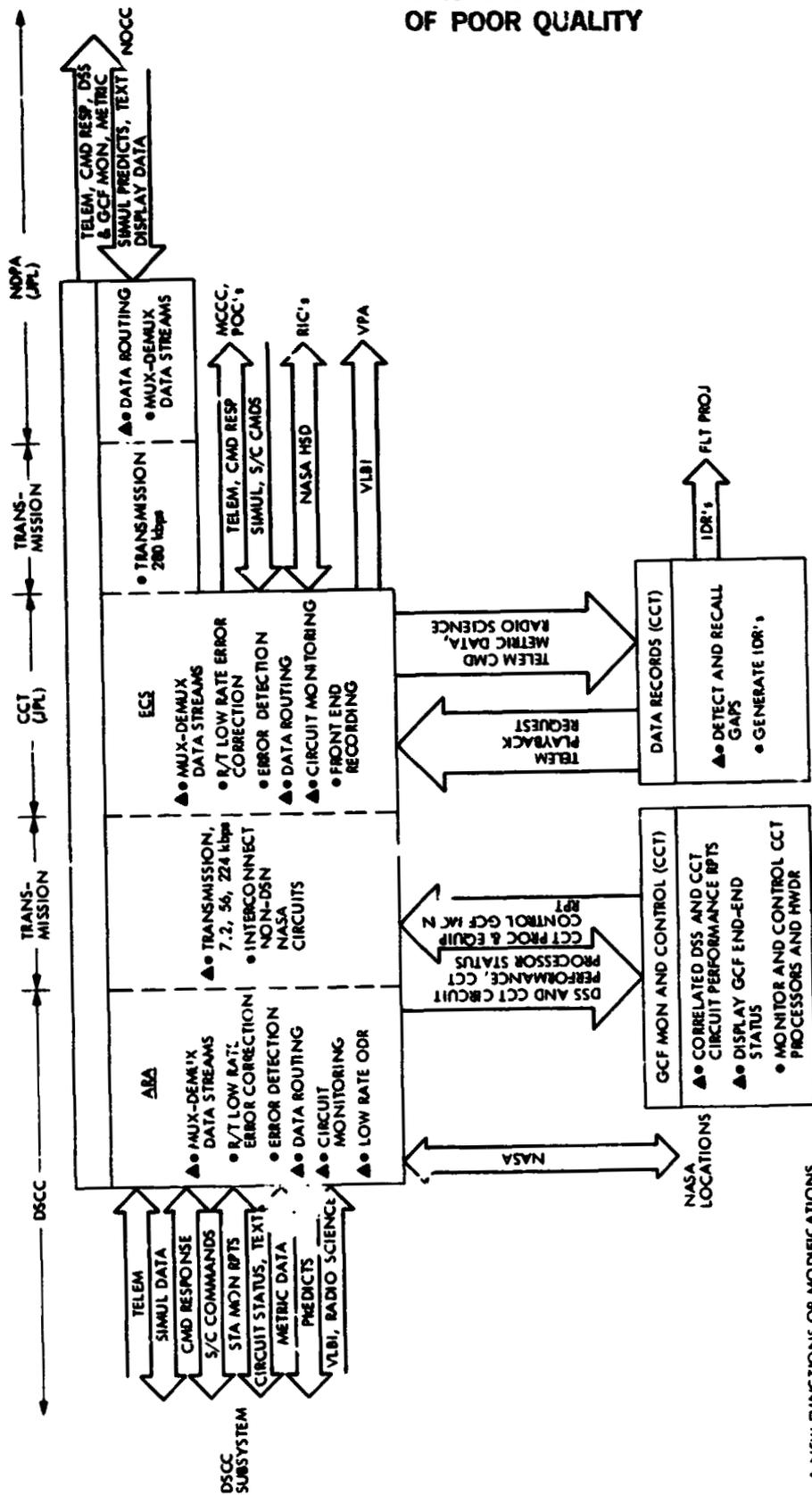
Fig. 2. Mark IV-A end-to-end digital communications, data record, monitoring, and network communications equipment functions and interfaces

△ NEW FUNCTIONS OR MODIFICATIONS

165
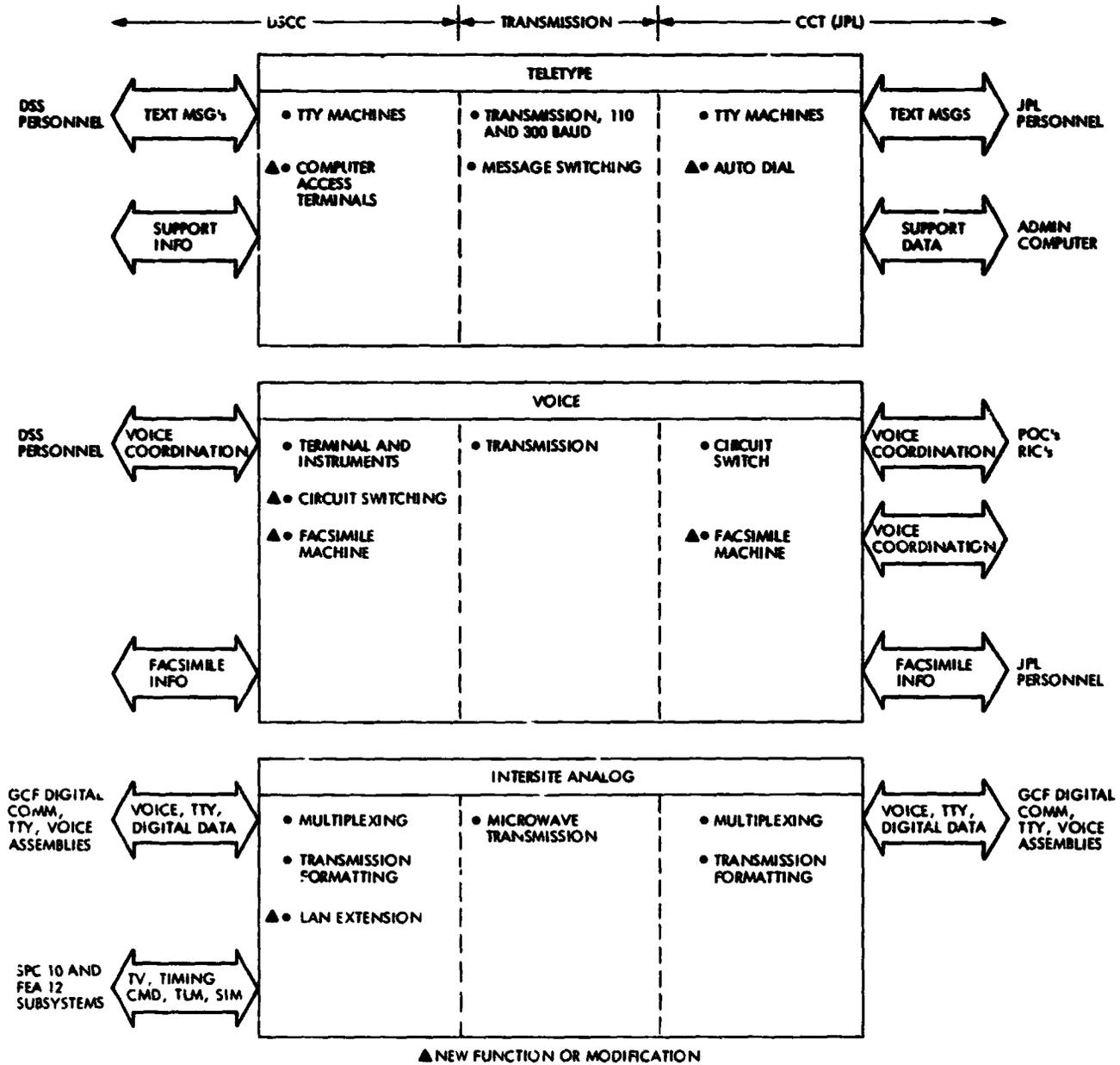
Fig. 3. Mark IV-A end-to end teletype, voice, and intersite analog functions and interfaces
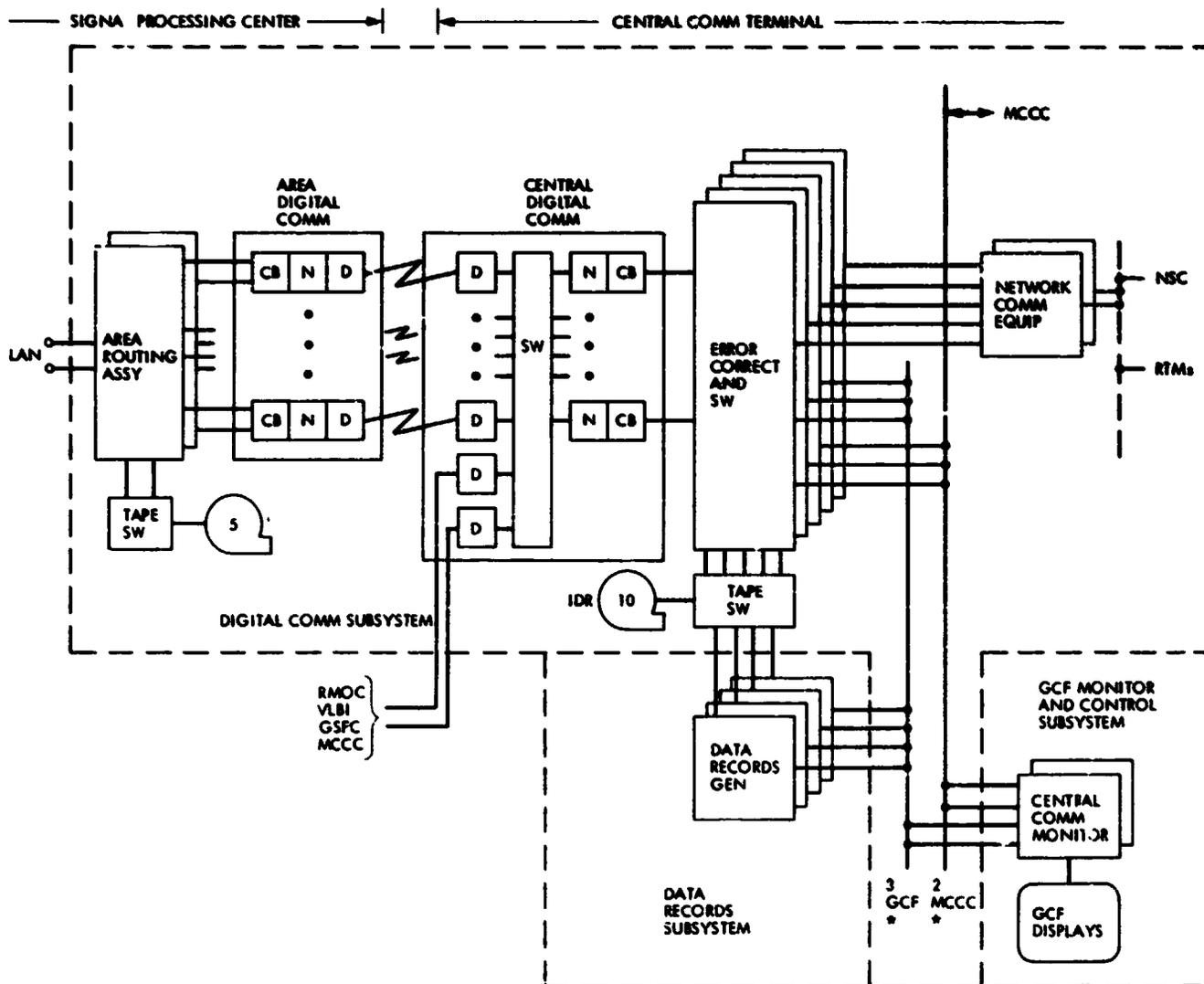
Fig. 4. Digital communications subsystem overview of GCF digital configuration
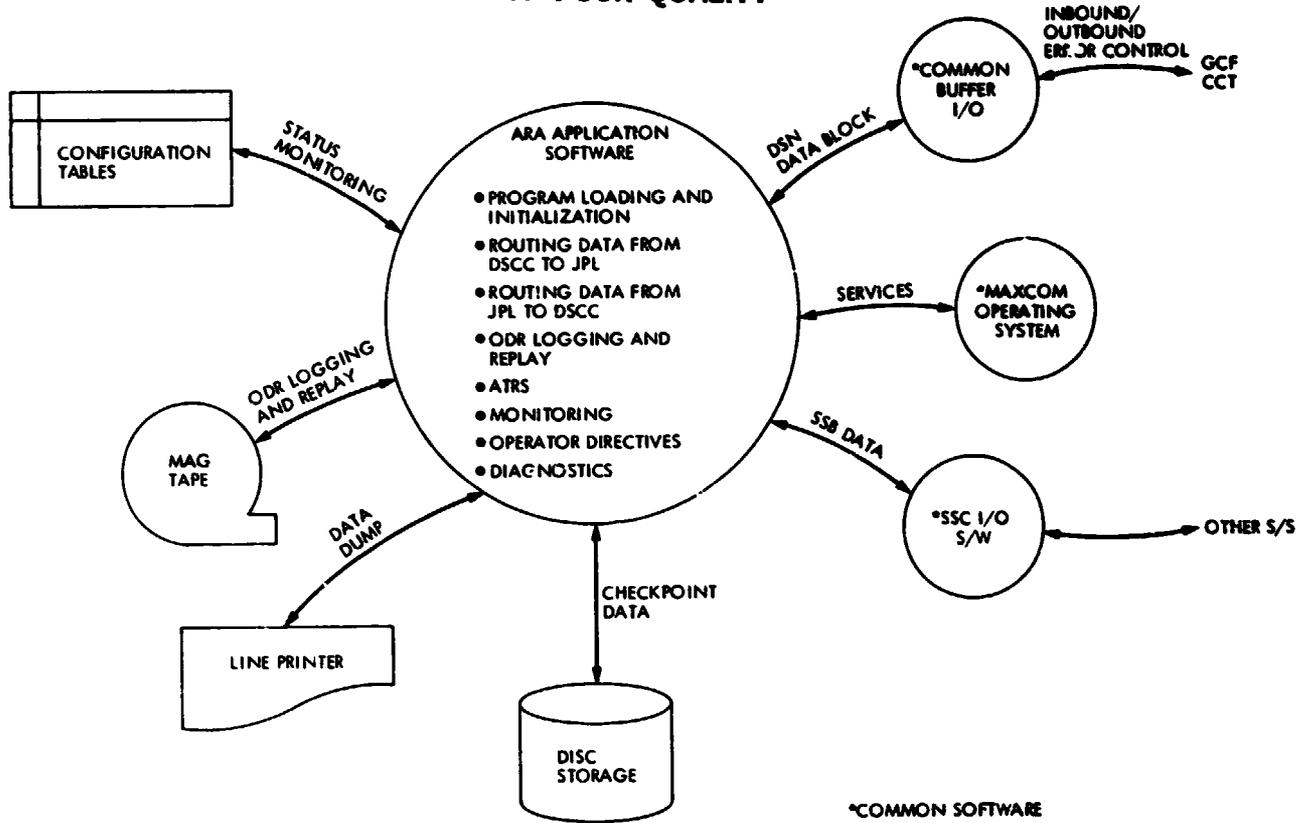
Fig. 5. ARA software functions and interfaces
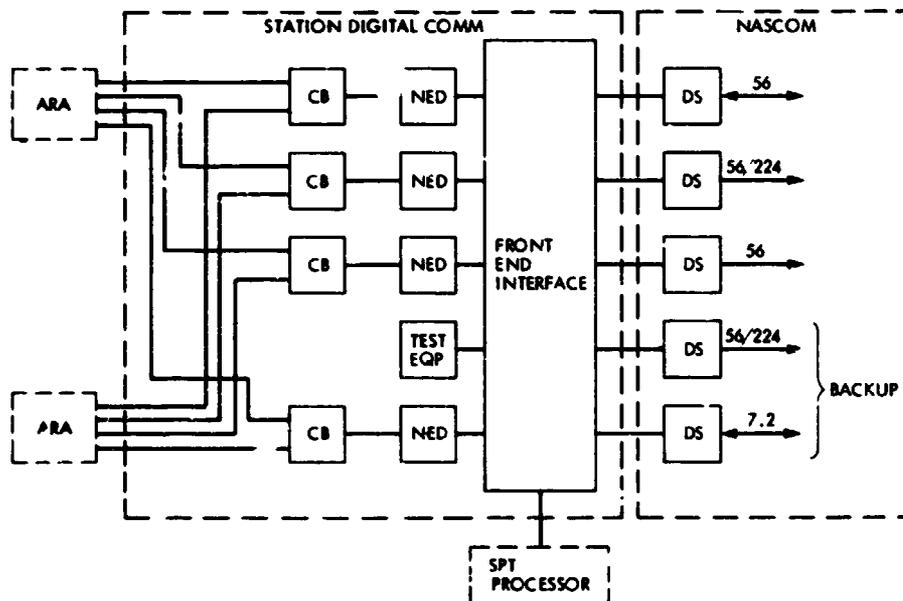
Fig. 6. Data transmission, station digital communications
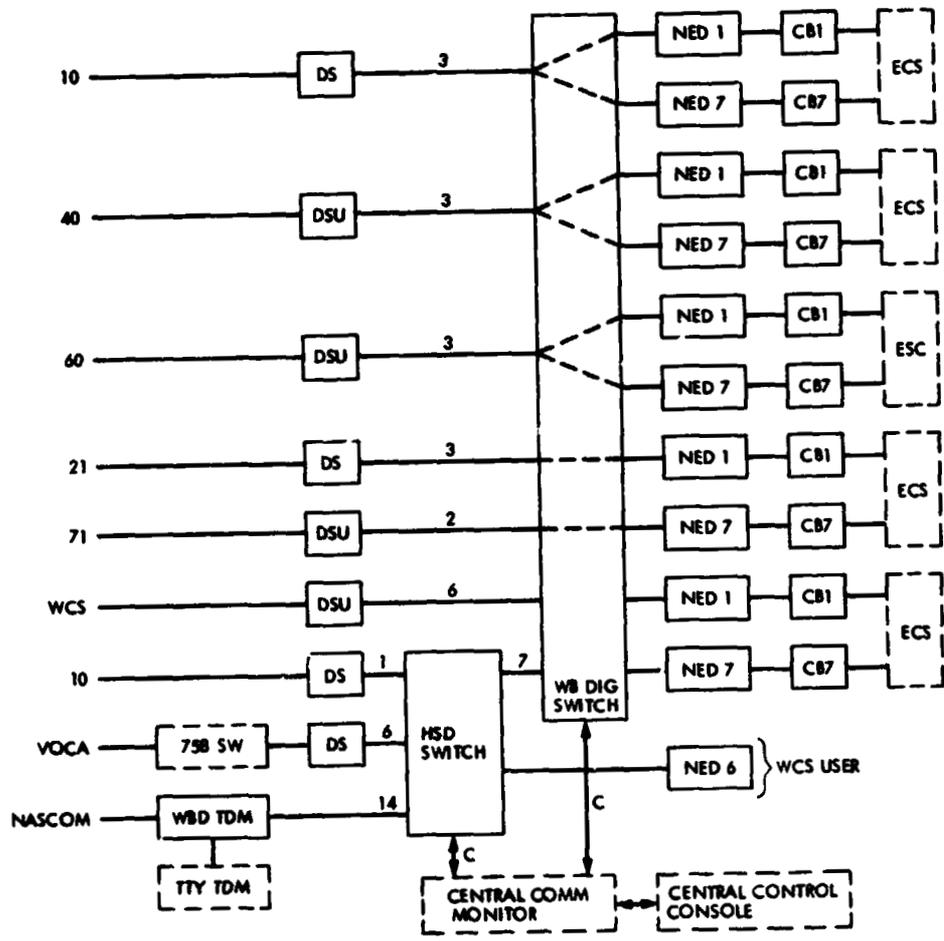
Fig. 7. Data transmission, central digital communications
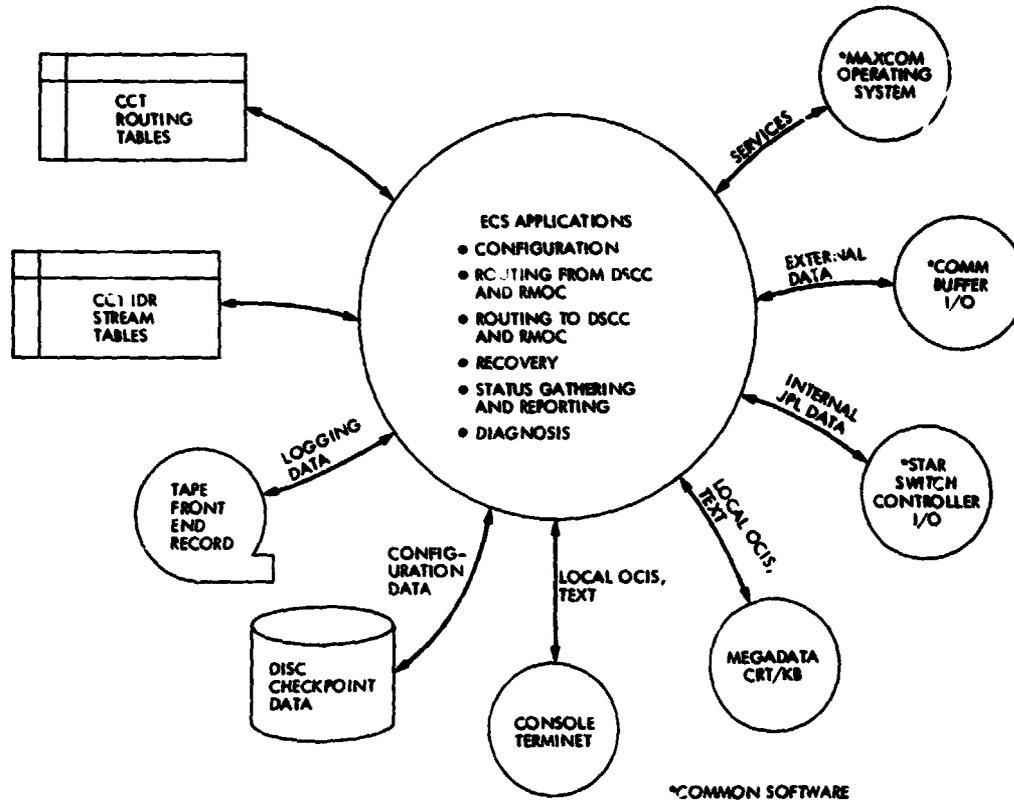
Fig. 8. ECS software functions and interfaces
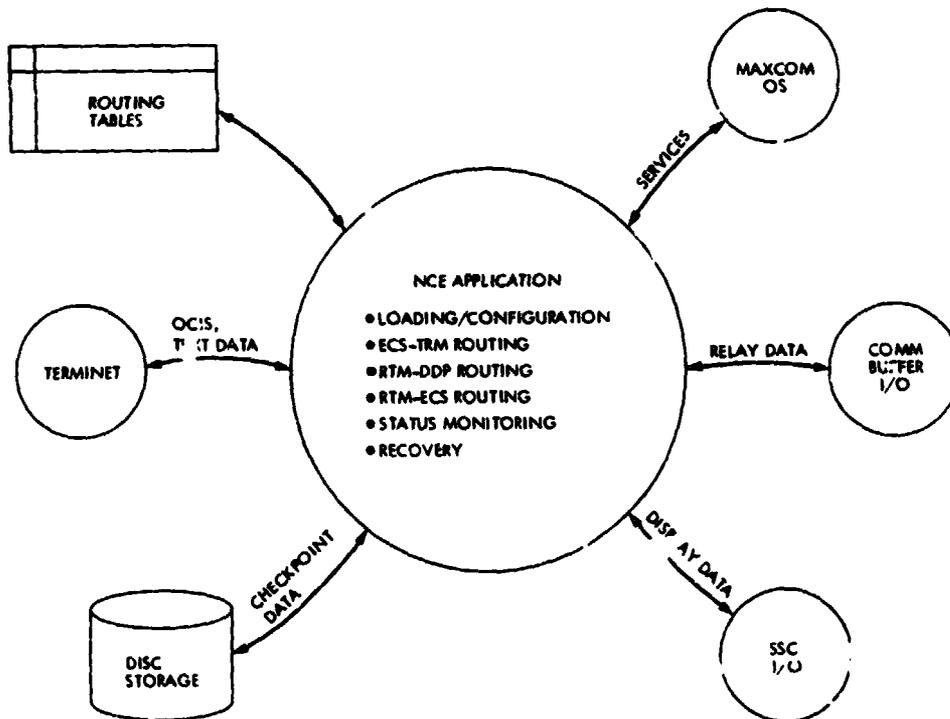
Fig. 9. NCE software functions and interfaces
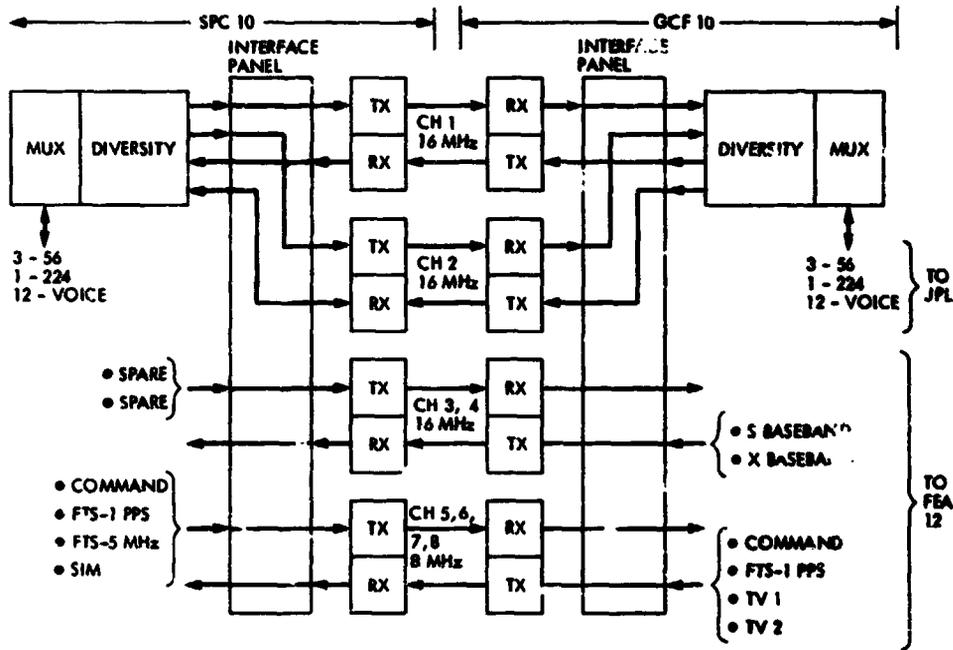
Fig. 10. Intersite microwave configuration, general design
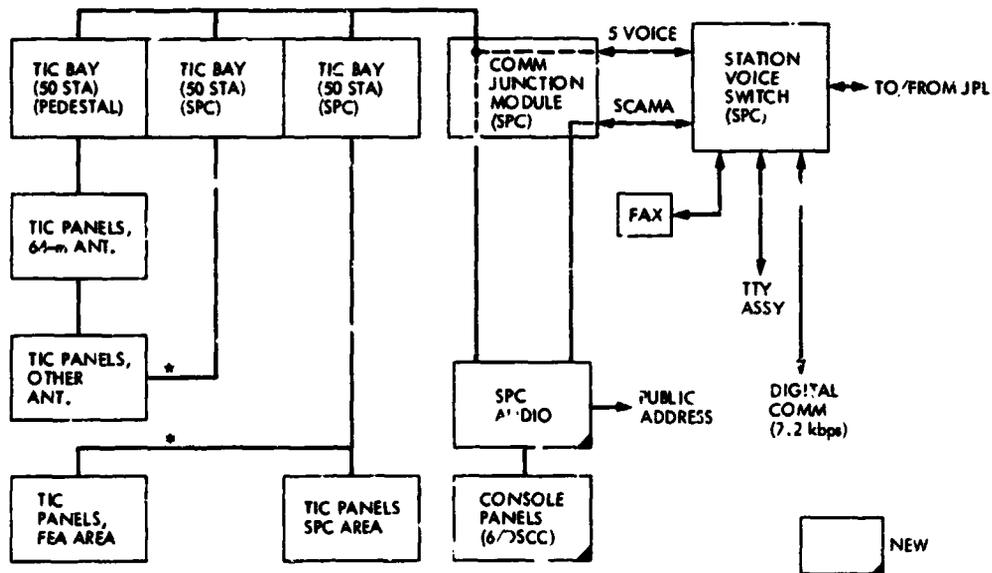
* WILL BE MICROWAVED IN THE CASE OF FEA-12

+ SCAMA SERVICE AVAILABLE AT OVERSEAS CONSOLES ONLY

* MAY REQUIRE AMPLIFICATION/POWER FOR PANELS IN DISTANT ANTENNAS

Fig. 11. DSCC voice, general design
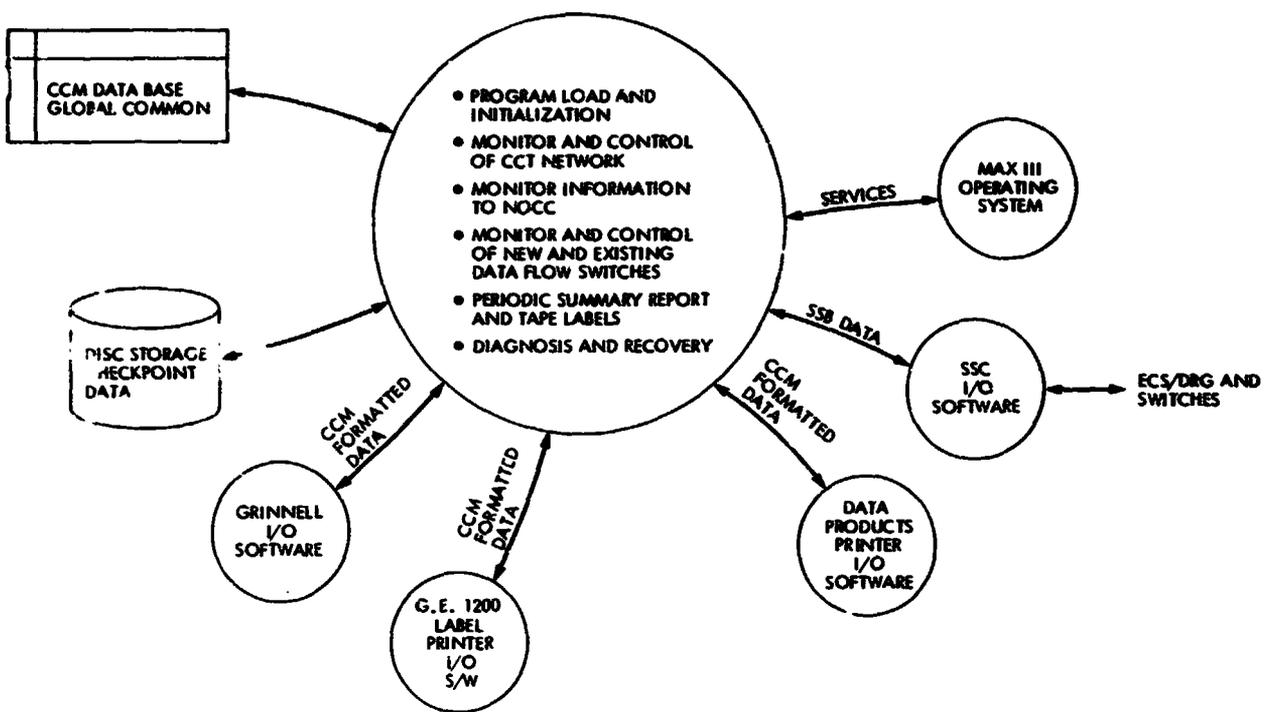
Fig. 12. CCM software functions and interfaces
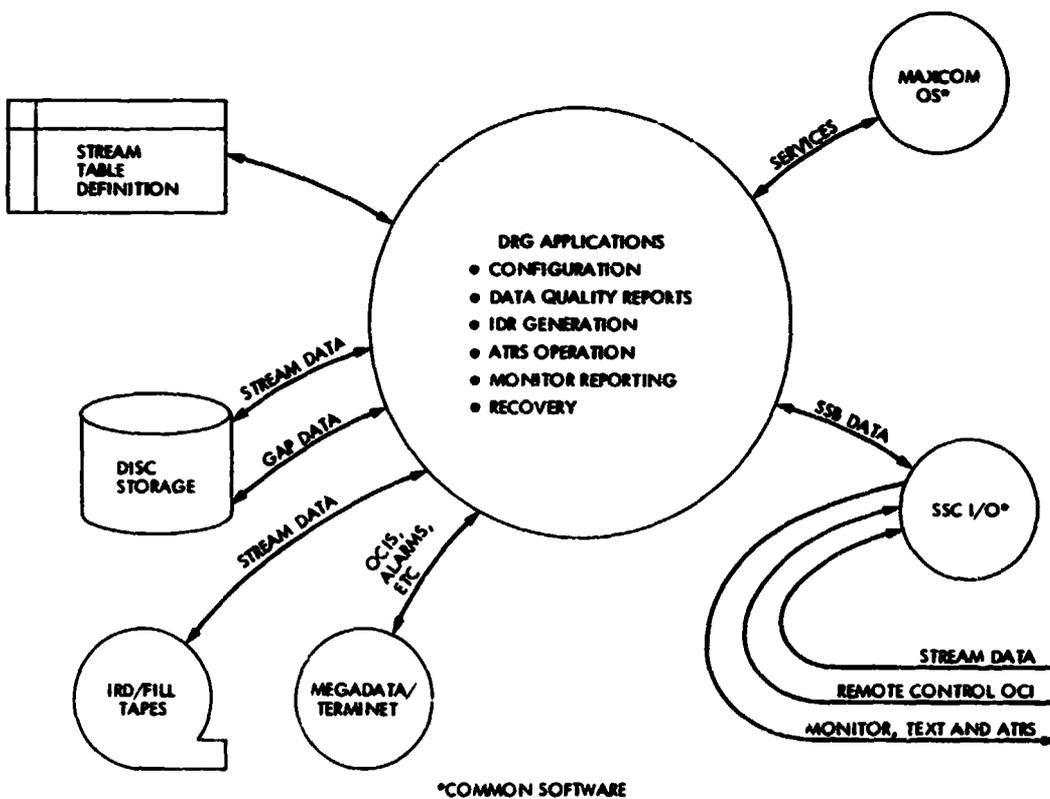
*COMMON SOFTWARE

Fig. 13. DRG software functions and interfaces

C-3

An Introduction to the New Productivity Information Management System (PIMS)

R. Hull

Telecommunications Science and Engineering Division

This report describes the Productivity Information Management System (PIMS), which is being developed at JPL. The main objective of this computerized system is to enable management scientists to interactively explore data concerning DSN operations, maintenance and repairs, in order to develop and verify models for management planning. Thus, PIMS will provide users with a powerful set of tools for iteratively manipulating data sets in a wide variety of ways. Most current database systems are designed to support a narrow range of predetermined types of queries. Thus, the design of PIMS includes unique, state-of-the-art features. The initial version of PIMS will be a useful but small-scale pilot system. This report (1) discusses the motivation for developing PIMS, (2) describes the various data sets which will be integrated by PIMS, (3) sketches the overall design of PIMS, and (4) describes how PIMS will be used. A survey of relevant databases concerning DSN operations at Goldstone is also included.

I. Introduction

The operation of the Deep Space Network (DSN) costs roughly $59 million a year.[1] The annual cost to JPL of operating the facility at Goldstone alone is roughly $15 million, and involves some 215 contractor manyears (Ref. 1). In this period of financial limitations, even incremental improvements in DSN efficiency yield important savings for JPL; and selections between alternative implementations and policies can have substantial financial implications. These facts underline

the utility and necessity of reliable, easily accessible information about DSN operations and expenditures. This report outlines the proposed development of a new computerized system, directed at providing convenient access to some of this information. Although only in the pilot stage, this system will provide very flexible access to an integrated, moderately detailed view of DSN productivity and costs.

Several highly successful computerized information management systems have been implemented at JPL over the past few years to facilitate various aspects of operating the DSN (e.g., the Equipment Database and the Engineering Change Management Databases of Section 377, and the DSN Scheduling Database of Section 371). Each of these systems was developed to accomplish a specific range of tasks concerning a fairly

---

[1] This includes the operation of the three deep space complexes and the Network Operations Control Center and the logistical and sustaining engineering costs. Long-range planning for advanced equipment and future projects costs an additional $35 million.

narrow aspect of DSN activity. In most instances, these databases were developed independently of the others. For these reasons, it is hard to use these systems to develop an integrated yet detailed picture of DSN operations and expenditures. Furthermore, it is difficult and cumbersome to obtain information of an ad hoc, nonroutine nature from these databases.

To illustrate this point in dramatic but clearly oversimplified terms, we present a brief analogy. Suppose that we were given (1) an alphabetical listing of all JPL employees and their office phone numbers, (2) a listing by section number of all the employees in each section, and (3) a listing, ordered by increasing telephone number, of the charges accrued against each phone during a given month. It is easy to imagine how such different lists could be generated by different groups for different purposes. Suppose now that a listing of the total of charges accrued against the phones in each section were desired. Clearly, it would be cumbersome and time-consuming to provide such a listing. Speaking broadly, management scientists are interested in all kinds of ad hoc, nonroutine aggregate summaries such as this, and the existing DSN databases simply cannot provide them in a rapid, convenient manner.

A second obstacle to developing an integrated view of DSN operations is that important information concerning expenditures is recorded in a variety of different ways. In fact, some of it is not at present recorded electronically. For example, information concerning most of the activity of the Maintenance and Integration Unit at Goldstone is recorded and stored on paper. Furthermore, some information relevant to determining operating costs (e.g., how many manhours are spent in transit between stations) has not, until recently, been recorded at all.

As a first step in resolving this problem, a group working in Division 330 (Telecommunications Science and Engineering) is currently developing the Productivity Information Management System (PIMS). The system will integrate data concerning various aspects of DSN operations. A distinguishing feature of PIMS will be the highly flexible accessing capabilities; users will be given tools for interactively manipulating and analyzing data sets in any way they choose. The initial implementation of PIMS is narrow in scope and will focus entirely on DSN operations at Goldstone. The system is viewed primarily as a pilot rather than as a full-blown, general-purpose tool. However, the system should prove useful to management scientists for the purpose of developing and testing management models, and to both managers and management scientists for the purpose of analyzing current DSN operations and choosing among various future alternatives. The experience gained in implementing and using this pilot version of PIMS can later be used to guide futher expansion of PIMS, and possibly to provide impetus for the development of a considerably more comprehensive information management system for the DSN.

The aim of the current report is to discuss the motivation for developing PIMS, to describe the various data sets which will be integrated by the initial version of PIMS, to briefly sketch the overall design of PIMS, and to indicate how PIMS will be used. In Section II the motivation behind PIMS is discussed in more detail, and the long-range direction of PIMS is considered. Section III presents an overview of the general capabilities of the initial version of PIMS, and Section IV describes the overall design of this initial version. In Section V we describe two examples of how PIMS will be used to integrate data and derive certain types of information. And in Section VI we conclude by indicating the current status of the effort to implement PIMS and discussing some possible extensions of PIMS. Two appendices are included. The first presents brief descriptions of the major databases currently maintained concerning DSN operations at Goldstone. The second contains copies of several of the forms used to collect data for those databases, and also examples of the outputs of some of them.

II. Motivation for PIMS

This section presents an overview of the motivations behind and objectives of PIMS. As noted in the Introduction, the expense of operating the DSN is considerable. Thus, implementation of efficient operational policies can lead to substantial savings and cost reductions. In this period of financial limitations and reductions, such savings gain even greater significance.

An important tool in developing these cost-saving policies is studying the current operation of the DSN. Indeed, as indicated in Appendix A, a wealth of data concerning DSN operations is being recorded each week, and much of it is publicized through periodic reports or is directly accessible. However, the different data sets are recorded and maintained for different purposes, and their overall characters reflect these differences. For example, the Barstow Production Control (BPC) database (A4 in Appendix A) stores the manhours expended on component repairs, retaining detailed manhour information for each component repaired. On the other hand, the Manpower Utilization Reports (A6), which record manhours expended in the actual operation of the antenna stations, list only the weekly totals of manhours expended in various categories. Thus, solely because of the nature of the actual data stored, it is inherently difficult to integrate information from the different data sets in a meaningful, useful manner. As a result, it is difficult to analyze this information from the perspective of reducing overall operational costs.

A second, distinct obstacle to integrating DSN operations data is that the various databases are stored in different formats and use different overall methodologies. On the one hand, some of the databases such as the Scheduling and BPC Databases, use sophisticated data storage and access routines written largely in assembly language. Others, however, such as the Manpower Utilization Reports and Transfer Agreement Status Database (A2) are really file management systems using simply formatted records. Data access for these is generally performed through the generation of a complete report rather than through response to a specific inquiry. And at the extreme, the Maintenance and Integration Work Orders (A5) are not stored electronically at all, but rather retained in their original hand-written form.

The PIMS effort is intended to be the first step in overcoming these two obstacles, in order that management scientists and managers can easily obtain integrated data concerning DSN operations. The uses of such integrated data abound. For example, the productivity, efficiency and expense of a wide variety of different activities could be determined and compared. Expenditures could be categorized in a variety of different ways to emphasize different aspects of DSN operations. (For instance, the total operational costs -- including original investment, operations, maintenance, and component repair costs -- of different subsystems or assemblies could be calculated and adjusted for relative usage rates, etc.). Comparisons could be made between the stations, and between past and present operational policies. Finally, the numbers computed from this integrated data could be used as the basis for a variety of cost-reducing statistical studies.

PIMS will also provide a second, distinctive type of access to integrated DSN operations data. Specifically, PIMS users will be given very flexible tools for iteratively manipulating the data in the systems. As a result, users will be able to interactively formulate queries which are based on the results of previous queries. Thus, it will be possible to interactively "explore" the data, and thereby discover anomalies or patterns of interest.

To illustrate the usefulness of such flexible access to integrated data in more concrete terms, we briefly mention the Remer and Lorden study conducted in the late 1970's (Ref. 2) The study analyzes data concerning the operation of DSS 13 in an automated mode during the latter half of 1978 in order to determine whether that automated mode resulted in cost savings. Discussions with the authors of this study indicate that obtaining the raw data underlying their analysis was difficult and time-consuming, and that the depth of the study was restricted as a result. It is anticipated that the initiation and maintenance of PIMS should partially alleviate such difficulties in future studies.

As currently envisioned, PIMS will be a narrowly focused pilot system with three primary objectives. These are (1) to provide access to integrated data concerning a limited portion of DSN operations, (2) to demonstrate the utility of a PIMS-like system, and (3) to provide practical experience in the use of such a system. Thus, while PIMS will address itself to only a portion of DSN operations, it is expected to provide a firm basis for designing a more comprehensive PIMS-like system in the future.

III. Overview of PIMS

We now discuss the overall capabilities of the (initial version of the) Productivity Information Management System. In broadest terms, PIMS will provide interactive access to data concerning the manhours expended at Goldstone by three different types of personnel (operations, maintenance and integration, and repair), and to data concerning "end user hours." PIMS users will be able to make direct queries to the database, and can also create and manipulate subsets of the data in a wide variety of ways. These access mechanisms will make it possible to (1) derive specific information, and (2) generate tables listing averages and totals for virtually any categorization of manpower expenditures. After this capability is fully developed, a mechanism for displaying these tables in a simple, easy to understand format may also be incorporated, as well as various statistical routines. Finally, a capability for investigating causal relationships may be added to the system.

A central theme in the design of the initial version of PIMS is to provide a simple, convenient user interface which allows users to perform virtually any manipulation on the underlying data sets, but which insulates users from the actual implementation details. In this manner, PIMS provides a powerful but easy to understand tool for performing virtually any data retrieval. To provide this capability, a major portion of PIMS is devoted to performing the routine and tedious detail work required in data processing as users specify various operations in a simple and abstract manner. Indeed, a major component of the preliminary version of PIMS is concerned entirely with such data management, and is completely hidden from the user's view. Specifically, this component performs the initial processing of raw data, which involves transforming data stored in a variety of different formats and locations into data all having uniform format.

A final general characteristic of the preliminary version of PIMS is the modularity of its design. This modularity will make modifications and expansions of PIMS capabilities a relatively easy and painless task. In view of the role of PIMS as a pilot, and also the possibility that the raw data available to PIMS may change over time, this is a particularly important feature.

IV. Overall Design of PIMS

In this section we consider the overall design of the initial version of PIMS as currently being implemented.

The global design of PIMS is shown in Fig. 1. The system has two primary modules, one for data input and one for data output. The data input module uses the raw data to generate files which contain records of a certain kind, called "event records." The data output module supports interactive access to these files of event records.

Event records are intended to store information concerning individual "events" involving the expenditure of manhours. Examples of events include the performance of a single preventive maintenance task, the repair of a single component, and the operation of a station during a tracking pass. Various parameters concerning events are stored in event records. For example, these include the type of activity, the number of manhours expended, the end-user benefited (if any), the subsystem and assembly involved (if applicable), and the preventative maintenance number (if applicable).

Referring to Fig. 1, we now describe in turn each of the modules and components of PIMS.

A. Raw Data

As mentioned in the previous section, PIMS will initially integrate data concerning manhours of three categories of personnel: (1) operations personnel, (2) maintenance and integration personnel, and (3) repairs personnel. Data concerning operations manhours will be drawn from the Weekly Histories compiled by the Data Processing Unit at Goldstone (in Appendix A, see A1) and the Manpower Utilization Reports (A6). Data concerning maintenance and integration manhours will be drawn from the Maintenance and Integration Work Orders (A5). Finally, the recently implemented Productivity Database (A4) will be used to obtain data on repairs manhours.

B. Data Input Module

The sole function of the data input module will be to transform the raw data into files of event records. This module will consist of three submodules, one each for the three types of personnel data used. Each of these submodules will have the capability of reading and processing raw data from the appropriate data sources. Thus the PIMS event files can be updated as new raw data accumulates.

C. Event Files

In its initial version, PIMS will maintain nine files of event records, three for each of DSS 11, 12 and 14, these being devoted to "operation" events, "maintenance" events and "repair" events. Data is separated according to station primarily to enhance efficiency — much of the raw data can most easily be processed one station at a time, and the separation will prevent the stored data files from becoming unreasonably large. Also, many data access requests are expected to distinguish between the stations, so data processing time will be saved.

To understand why data is separated according to personnel category, we note that although many event parameters (such as manhours and day-of-year) are shared by events of each category, other event parameters (such as end-user hours for operations events, or turnaround time for repairs events) are unique to a given category. Thus the separation of events permits more efficient storage of the data. It should be noted, however, that sets of events of different types can be readily combined by PIMS users (see below).

D. Data Output Module

The function of the data output module is to provide convenient interactive access to the event files. As currently envisioned, this module will provide a menu-oriented interface to users. Thus when the system is on, users will be presented a "menu" of possible commands to choose from. As a result, the system will guide users through the correct steps of a data accessing procedure, and hence be very · ccessible to novice users.

The commands which PIMS users can give via the data output module will give users the capability of directly manipulating files of events. Specifically, users will be able to create new files, select specific subfiles according to given parameter values (e.g., select all events with manhour value between 3 and 5 hours), sort files, and merge files (possibly containing different types of events). Also, capabilities to print out the contents of these files, and to calculate simple numerical summaries of them (e.g., list the total manhours expended, broken down by week and subsystem type) will be available. To accomplish this the data output module will provide users with a small set of "atomic" file manipulation commands which can be applied repeatedly to obtain desired files and results.

E. Reference Files

The final major component of PIMS is the set of files maintained for reference purposes. These will include, for example, a portion of the Transfer Agreement Status Database (Appendix A, see A2) which lists the numbers, three-letter acronyms, and brief descriptions of DSN subsystems and assemblies.

Another example is the listing of preventive maintenance numbers and their short verbal descriptions. Because the data in these files are modified occasionally, they are given a fairly independent status in PIMS. This will ensure that their modification can be incorporated in a simple straightforward manner.

V. Some Examples

In this section we briefly illustrate some of the capabilities that PIMS will have by describing three representative examples. Together they indicate the primary capabilities of PIMS as currently being developed; other capabilities will probably be added after the system is operational.

A. Table Generation

A basic capability of PIMS will be to generate tables summarizing information concerning various aspects of Goldstone operations. For example, suppose that a table is desired which lists, for the period June 1 to August 1, 1981, the total number of manhours expended per week, broken into categories of tracking, preventive maintenance, corrective maintenance, and repairs. To obtain such a table, the following sequence of steps could be performed. First, since only the periods June 1 to August 1, 1981, are desired, nine working files could be formed, each consisting of the relevant portion of one of the permanent event files (see Section IV-C). Now these files could be merged into one, and sorted by day of year. Next, the resulting file could be partitioned into one-week blocks and, within each block, sorted according to work category. (This would have the effect, within each block, of placing all events concerning a given work category physically next to each other.) Having arranged the file in this manner, a routine can now be executed which calculates, for each week, the number of manhours expended within each work category. Finally, a table printing routine can be called to print the results on paper or display them on the screen.

All of the procedures described above will be implemented in a very flexible fashion in PIMS. Thus a table can be constructed that lists total manhours broken into virtually any categories. Other parameters can also be totalled (e.g., end-user hours or downtime), and other types of aggregate functions will be available (e.g., average instead of total).

B. Comparison of Productivity

A second application of PIMS will be to compare corresponding aspects of different parts of Goldstone activities. For instance, suppose that a comparison, between the three Goldstone stations, of the ratio of the manhours expended on preventive vs corrective maintenance is desired. To obtain this,

a procedure similar to that used for generating tables can be applied. Specifically, the user can first create files, one for each station, which contain all events involving preventive or corrective maintenance. Next, these files can be used to determine, for each station, the number of manhours expended on the two categories of maintenance. The desired ratios are then easily calculated.

Since PIMS is capable of categorizing data in a large number of ways, it will be useful in making many different types of comparisons.

C. Iterative Manipulation of Files

Another basic feature of PIMS is that users will be able to manipulate working files in an iterative fashion. For example, suppose that the user created the table described in Section V-A, and noticed that repairs costs were considerably higher than the other costs. The user may at that point wonder whether this imbalance was peculiar to a given subsystem or occurred in all of them. Using PIMS, the user can sort the working file already obtained by subsystem, and then list for each subsystem the number of manhours expended in each of the specified work categories. If interested, the user may then refine the data further, listing manhour expenditures categorized by assemblies within one or more subsystems.

It is clear that this kind of iterative, ad hoc file manipulation capability will provide users a means to literally "explore" the data in any way they please.

VI. Concluding Remarks

We conclude by describing the current status of the PIMS effort, mentioning some possible future directions for it.

At present, the overall design of the initial version of PIMS is essentially complete. The routine for processing raw data concerning operations personnel (Section IV-A) has been implemented and debugged. Also, the module which performs basic manipulations of event files (Section IV-D) is essentially completed, and the module for handling the reference files (Section IV-E) is under development. The remaining modules include those for inputting maintenance and repair data and for driving the menu-driven user interface.

Once the system is in operation, it is expected that, based on their experiences, users will determine that certain capabilities should be added to PIMS. For example, certain new data sets, such as Discrepancy Report information (Appendix A, see A7), may be desired. Also, more complicated statistical capabilities may be desired. Finally, new ways of representing

the data, e.g., using plotted curves to indicate one parameter as a function of another, could be added.

More generally, if PIMS proves to be a useful tool for managers and management scientists at JPL, the project may lead to a more substantial effort to provide access to integrated DSN operations data. Given a firm commitment from management, a more ambitious database management system might be devised to perform the same functions as PIMS, except in a much more sophisticated and complete manner. Indeed, the PIMS effort may indicate the desirability of incorporating, at a fundamental level, a PIMS-like capability into the data management component of the Network Consolidation Project.

Bibliography

1. Fahnestock. R. J., personal communication.

2. Remer, D. S., and Lorden, G., "Deep Space Station (DSS 13) Automation Demonstration," *The TDA Progress Report 42-57, March and April 1980*, Jet Propulsion Laboratory, Pasadena, Calif. pp. 103-119, June 15, 1980.
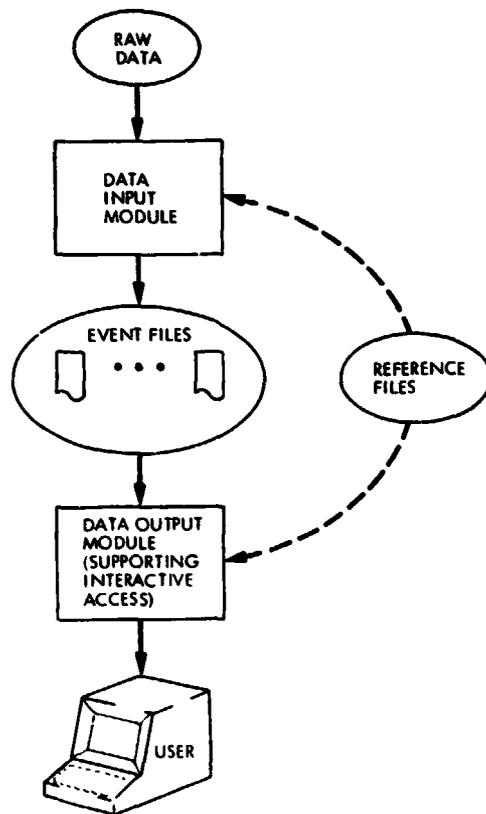
Fig. 1. Overall design of PIMS

Appendix A

Survey of DSN Operations Databases and Data Reports

The sheer size and complexity of the DSN has necessitated the development of several highly successful computerized information systems which are used to support its operation. In this appendix we briefly survey some of the more significant of these databases, and also mention a couple of related data sets (including an important database which has not been computerized to date). Table A1 provides a brief summary of our discussion. As noted above, PIMS will focus primarily on data in three of these databases (namely A1, A4 and A5) and make reference to some of the others (notably A2 and A6).

A1. The DSN Scheduling Database

The DSN Scheduling Database is maintained and used by Section 371 to schedule, on a week-to-week basis, the tracking activities of all of the DSN antenna stations. Roughly 500 to 700 events are scheduled for any given week, and events can be scheduled up to 53 weeks in advance. Once fixed, the actual schedule is used by the various DSN stations to plan, on a minute-to-minute basis, specific station activities (both those called for on the schedule and others such as certain preventative maintenance tasks). After a week has passed, Section 371 modifies the we k's schedule to reflect the actual events of the week, and archives it as a weekly "history." These weekly histories form the basis of "DSN Utilization Reports," which summarize DSN activities, categorized by antenna size and end users. Independently, the Data Processing Unit of Bendix (in Barstow) updates the Goldstone portion of the weekly schedule and archives its own weekly "histories." These are used by the Data Processing Unit as the basis of "Station Utilization Reports," which are subsequently distributed by Section 371. These "Station Utilization Reports" list, for each station, the number of station operating hours (SOH) and end-user hours (EUH) devoted to each of the DSN "end-users" in the given week (see Appendix B for a sample report).

A2. Transfer Agreement Status Database (890-61)

The Transfer Agreement Status Datab is concerned with recording information concerning engineering responsibility, from the station and facility level down to the subsystem and assembly level, of the DSN. For each station, subsystem and assembly it lists the subsystem engineer, the cognizant development engineer, the cognizant operations engineer and the cognizant sustaining engineer (if applicable), and the current status of these responsibility assignments (e.g., transfer planned, transfer complete). This database is stored at JPL's Information Processing Center (IPC) on the Univac 1100/81, and is maintained by Section 355. It is updated as needed to reflect assignments and transfer status changes. A variety of accessing modes to this database, each generating a report of a certain kind, is available.

An important aspect of the Transfer Agreement Status Database is that it provides information cross-referencing various naming conventions which have arisen for describing parts of the DSN. Each station can be viewed as consisting of roughly 25 "subsystems," and each subsystem is broken into roughly 10 "assemblies". Some subsystems are common to more than one station while others are unique to a specific station. Generally, a given subsystem or assembly can be identified in each of the following three ways: (1) the name of the subsystem or assembly (e.g., "34M. Ant Mechanical S/S" or "Electronic Control Assembly"), (2) the three-letter acronym, which sometimes extends to six letters (e.g., ANT or SVO), and (3) the two- to four-digit two-level hierarchical identifier (e.g., 46.00 or 46.02; here the first (two) digit(s) refer to the subsystem while the latter digit(s) refer to the assembly within the subsystem). The Transfer Agreement Status report provides a correlation between these three modes of identification and specifies which subsystem and assemblies are relevant to a given station.

A3. DSN Equipment Database

The Equipment Database is a large-scale database management system implemented on the Univac 1100/81 and maintained by Section 377. It holds data concerning all of the actual physical components comprising the DSN. The database is used primarily for inventory control and component tracking and also to support ad hoc operations performance analyses and answer various ad hoc queries. The database currently holds some 170,000 records and grows at a rate of roughly 1000 records per week. It lists, for each DSN component, a variety of information, including its unique identifier (called the "DSN control number" or "conaudit number"), description, manufacturer identification number, current location, and information concerning its repair history. The database can be accessed through a versatile interactive command language which supports m altikey retrieval and totally flexible format specification. The database can also be interfaced directly by computer software.

A4. Barstow Production Control Database

Quite recently, a new computerized database system has been implemented to monitor and control the flow of DSN components through the repair facilities at Goldstone. This Barstow Production Control (BPC) Database is in the final stages of implementation in Section 377, and will be maintained (and possibly enhanced) by that section. The database will store very detailed records concerning each instance where a DSN component was repaired, including various dates, what specific activities were performed (including procedure numbers, if applicable), which DSN test equipment was used, and how many manhours were spent on various ac ivities (Appendix B shows an example of a "Service Report," which can list all of the information stored in one record of this database).

The database is housed on the Univac 1100/81, currently holds roughly 15,000 records, and is growing at a rate of approximately 1000 records per month. In its current implementation the database provides sophisticated data compres sion, and also indexing to provide fast access according to cer tain data fields. Data is put into the system via computer termials, and records are updated as various stages of repair, testing or calibration are completed. The system can be accessed by a flexible, command-oriented query language, both online and in batch mode. At present the output of this database is formatted identically to the Service Report.

The BPC Database replaces in part another comprehensive database, the Failure Database, which was maintained by Section 377 until April 1981. That database was used to main tain records concerning component failures, and stored a fail ure history for each component. The database was partially integrated into the Equipment Database, and data could be accessed through a number of data fields. Whereas the Failure Database was used to monitor repair activity and store a comprehensive history of each substantial component, the BPC Database at present only monitors individual instances of component repair.

A5. Maintenance and Integration Work Orders

An important database concerning DSN Goldstone opera tions which is not currently computerized concerns the activity of the Maintenance and Integration (M&I) Unit (of the Bendix Corporation) at Goldstone. This unit is responsible for performing a large class of routine preventive maintenance activities, trouble-shooting station problems, removing and replacing components, and performing some of the work gen erated by Engineering Change Orders. This activity is moni tored and directed via "Work Orders" (see Appendix B), wi i i.

are used (1) to specify that a given task is to be performed, and (2) to record the work that was performed, including manhours expended (and beginning recently, the amount of time used in transportation). Although the current system used by the M&I unit to record its activity is certainly ade quate, it is clear that a computerized system would enhance the ability to obtain interrelated data and overview info tion, thereby improving the unit's overall performance.

A6. Manpower Utilization Reports

The Manpower Utilization Reports are generated on a weekly basis by the Tracking Operations and Data Processing Units (of the Bendix Corporation) at Goldstone and Barstow (respectively). These reports give a weekly summary of the operations personnel and M&I personnel manhours expended at Goldstone, broken into roughly 20 categories (see Appen dix B). Note the distinction between Manpower Utilization Reports, which list manhours expended, and the DSN and Station Utilization Reports (discussed in A1 above), which list station operating hours and end-user hours.

The Manpower Utilization Reports are assembled from the "shift reports" kept at each station (at Goldstone) and by M&I (see Appendix B for a blank copy of Shift Report) and stored in the Sycor minicomputer maintained in Goldstone by the Data Processing Unit in Barstow. Quarterly and annual summaries of this information may also be generated upon request.

A7. Discrepancy Report Database

The Discrepancy Report Database is maintained by Sec tion 371 to monitor "discrepancies." A discrepancy is defined to be an instance in which an end-user was scheduled to receive telemetry data, and received either degraded data or no data. Thus, discrepancies are initially ' nerated" when an end-user reports such data degradation or loss. Once a dis crepancy has occurred, it is considered "open" until the cause of the discrepancy has been located (and if applicable, remedied).

The Discrepancy Report Database is maintained in an IBM 3032 computer managed by the Administrative Comput ing Service (ACS) at Caltech. The database is used to store and update records concerning open discrepancy reports, and has records of past discrepancy reports going back to 1975. It is also used to determine operational "Mean Time Between Failures," system trends and distribution of problems by hardware, software and procedural anomalies. Finally, special software routines can be used to answer ad hoc queries made to the database.

A8. Engineering Change Management (ECM) Database

The ECM database is used to monitor the implementation status of Engineering Change Orders (ECO) in the DSN. These data include a description of the change, its application and various milestones/status reports during the development, shipment, and facility installation and testing of it. Periodic management reports are generated and some ad hoc queries are supported in predefined formats. The system was originally implemented in 1976, is housed in the IPC and is managed by Section 377. It should also be noted that before approval, information concerning Engineering Change Requests (ECRs) is maintained in AODC word processors.

A9. System for Resources Management (SRM)

Although not dedicated solely to DSN operations, we briefly describe the System for Resources Management. The SRM provides the backbone for accounting activities at JPL. It is used to monitor all JPL income and expenditures, to coordinate future expenditures against future income, and to record past income and expenditures. The SRM is capable of formatting and summarizing accounting information in a variety of ways, producing reports such as the Resources Status Report (RSR), and the SRM planning summaries. Also, interfaces between the SRM and the WADSUM (see A10) have been developed. At this time, the SRM is not an interactive system — updates and modifications to the data must be done in batch mode, and only after a special processing of

the entire data set is performed can the (newly revised) data be printed out.

The SRM is housed in the ACS IBM 3032. As well as containing data for the current year, it stores detailed records for the past 4 years and archived records for earlier years. Also, it contains information concerning projected expenditures for the next 5 years. The database is quite large; it holds roughly 200,000 records for a current year, and at present the primary history file holds another 320,000 records.

A10. The Work Authorization Document Summary System (WADSUM)

This system was developed to fulfill reporting requirements regarding planned TDA resources allocations. The database, resident on the IPC Univac 1100/81, contains headcount and expenditure data for each account in the TDA program. The database is composed of 1500 records, and may be accessed and updated either interactively or in batch mode. Numerous sorting and report writing capabilities are supported.

Various interfaces exist between the WADSUM and the SRM. For example, a WAD performance report, generated monthly on the ACS IBM 3032, details the discrepancies between the "planned" manpower and funding levels in the WADSUM database, and corresponding accrued "actu·' recorded in the SRM data bases.

The WADSUM system was developed and is maintained under the cognizance of the TDA Program Control Office.

Table A-1. Database summary

	Name	Section	Original implementation	Residence	Primary objectives	Secondary usage(s)	Accessibility
A1	DSN Scheduling	371	1965	IPC	Facilitate scheduling of DSN tracking activities	Basis for station operations. Basis for various "utilization" reports and for weekly "histories"	Schedules generated at various intervals. Other access almost exclusively restricted to direct scheduling activity
A2	Transfer Agreement Status Database (890-61)	355	1975	IPC	Store current engineering responsibility assignments for DSN	Correlation of subsystem and assembly nomenclatures	Several alternative report generating capabilities
A3	DSN Equipment Database	377	1975	IPC	Inventory control of DSN equipment	Operations performance analysis. Ad hoc questions and analysis	Versatile interactive command language for item specification and listing. Direct software interface available
A4	Barstow Production Control Database	377	1981	IPC	Monitor and control flow of DSN components through DSN repair facilities		Command-oriented query language; output to (hard copy) "service report" and to terminals
A5	Maintenance and Integration Work Orders	M&I Unit at Goldstone		Goldstone (hard copy)	Monitor and control M&I activity		Hard copies stored in sequence at Goldstone. Subsets of data independently maintained by cognizant engineers
A6	Manpower Utilization Reports	Tracking operations and data processing unit at Goldstone and Barstow	1978	Goldstone (Sycor mini-computer)	Report summaries of manpower expended by operations and M&I personnel		Reports generated weekly
A7	Discrepancy Report Database	3.	1966	IBM 3032 at Caltech	Monitor discrepancies		Ad hoc queries answered using special software routines
A8	Engineering Change Management (ECM) Database	377	1976	IPC	Monitor engineering changes to DSN		Interactive access to facilitate processing and monitoring of engineering change orders. Summary report generation
A9	System Resources Management (SRM) Database	632	1969	IBM 3032 at Caltech	Schedule and monitor JPL financial accounts		Report generation; ad hoc queries; interface to WADSUM
A10	Work Authorization Document Summary System (WADSUM)	410	1972	IPC	Fulfill NASA reporting requirements regarding planned resource allocations		Interactive access; interface to SRM; report generating capabilities

Appendix B

Examples of Schedules and Reports

```
**********************************************************************************

                        DSS-14 STATION UTILIZATION REPORT
                                     FOR
                              WEEK  5, 1981


I. DSN USER SUPPORT                                    SOH    EUH    PER
   A. 1. SPACECRAFT TRACKING
            PIONEER-7                                  6.17   4.83   3.67%
            PIONEER-9                                  5.25   4.50   3.13%
            PIONEER-10                                14.17  12.17   8.43%
            PIONEER-11                                 7.92   6.67   4.71%
            PIONEER-12                                16.42  13.42   9.77%
            VOYAGER-2                                 13.67  10.67   8.13%
      2. PROJECT RELATED SUPPORT
            VOYAGER                                    6.00   3.00   3.57%
      3. DSN PROJECT PREPARATION
            DSN                                        2.75   2.75   1.64%
   B. RADIO SCIENCE
            OSS                                       11.67  10.92   6.94%
   C. ADVANCED SYSTEMS
   D. SPECIAL
                                                     ------- ------- -------
      SUB-TOTAL                                       84.00  68.93  50.00%


II. FACILITY ACTIVITIES
    A. MAINTENANCE
       1. PREVENTIVE                                  25.75          15.33%
       2. CORRECTIVE(DOWNTIME)                        41.75          24.85%
       3. CORRECTIVE(NO DOWNTIME)                      0.00           0.00%
    B. PERSONNEL TRAINING                              1.08            .64%
    C. DSN ENGINEERING
       1. ENGINEERING SUPPORT                         15.42           9.18%
       2. DEVELOPMENT OR TESTING                       0.00           0.00%
       3. MINOR MODS                                   0.00           0.00%
III. OTHER ACTIVITIES
     A. MAJOR MODIFICATIONS                            0.00    .      0.00%
     B. HOST COUNTRY RADIO SCIENCE                     0.00           0.00%
     C. MISCELLANEOUS                                  0.00           0.00%
                                                     ------- ------- -------
     TOTAL HOURS                                     168.00  68.93 100.00%


**********************************************************************************
```

Fig. B-1. Station Utilization Report (Data Processing Unit, Goldstone)

Fig. B-2. (Portion of) Weekly DSN Operations Schedule (Section 371)

SHIFT REPORT

DSS _____   Date: _____ 197_

Crew Supervisor _____ Crew _____ DOY _____ From _____ to _____ (GMT) Shift _____

	Absent						Assigned Station Personnel					
Sick	**UPB**	**Vac**	**Loan**	**School**	**Comp**	**Other**	**Crew**	**Other**	**Total Present**	**Man-Hours**	**Station Opns Hours**	**End User Hours**

DSN USER

Support _____ Mission & Spacecraft No. _____ Pass No. _____ Cat. _____ __.__ __.__ __.__

Support _____ Mission & Spacecraft No. _____ Pass No. _____ Cat. _____ __.__ __.__ __.__

Support _____ Mission & Spacecraft No. _____ Pass No. _____ Cat. _____ __.__ __.__ __.__

Support _____ Mission & Spacecraft No. _____ Pass No. _____ Cat. _____ __.__ __.__ __.__

FACILITY MAINTENANCE & DEVELOPMENT

Maintenance Preventive _____ 2A1 __.__ __.__

Maintenance Corrective (Down Time) _____ 2A2 __.__ __.__

Maintenance Corrective (No Down Time) _____ 2A3 __.__ __.__

Training 2B1 Name	Type				Description	Instructor		
	TM	**TI**	**LPD**	**Other**				
							__.__	__.__
							__.__	__.__
							__.__	__.__
							__.__	__.__
							__.__	__.__
							__.__	__.__
							__.__	__.__
							__.__	__.__
							__.__	__.__
							__.__	__.__
							__.__	__.__

Engineering Support _____ 2C1 __.__ __.__

Development/Testing _____ 2C2 __.__ __.__

Minor Modification _____ 2C3 __.__ __.__

Major Reconfiguration _____ 3A1 __.__ __.__

Miscellaneous _____ 3C1 __.__ __.__

(SOD Initials _____) Totals __.__ __.__ __.__

RECORD OF EVENTS

GMT	EVENT

Fig. B-3. Shift Report for antenna stations (Tracking Operations Unit, Goldstone)

WEEKLY MANPOWER UTILIZATION REPORT

GOSCC TRACKING STATIONS

WEEK ENDING 22 NOV 81

PROJ OR TASK	WORK CODE	DSS-11 NUMBER OF PERSONNEL=14			DSS-12 NUMBER OF PERSONNEL=14			DSS-14 NUMBER OF PERSONNEL=14		
		MAN HOURS THIS WEEK	% TOTAL MAN HRS THIS WEEK	% TOTAL MAN HRS YEAR TO DATE	MAN HOURS THIS WEEK	% TOTAL MAN HRS THIS WEEK	% TOTAL MAN HRS YEAR TO DATE	MAN HOURS THIS WEEK	% TOTAL MAN HRS THIS WEEK	% TOTAL MAN HRS YEAR TO DATE
HELIOS-1	1A1	134.8	24.4%	14.6%			2.7%			.2%
HELIOS-2	1A1			< .1%			.2%			.1%
PIONEER-10	1A1				133.5	25.1%	1.1%			9.4%
PIONEER-11	1A1			13.9%	30.6	5.8%	2.9%			3.2%
PIONEER-12	1A1	175.3	31.8%	16.3%	15.2	2.9%	5.3%			7.4%
VOYAGER-1	1A1			4.1%	82.6	15.5%	8.6%			1.6%
VOYAGER-2	1A1			2.6%	108.4	20.4%	21.8%			14.0%
VIKING-1	1A1			< .1%			.1%			.2%
PROJ REL SUPP	1A2	29.5	5.3%	3.3%	20.0	3.8%	7.3%	92.0	16.3%	13.4%
DSN PROJ PREP	1A3	11.0	2.0%	2.0%	15.0	2.8%	1.7%	4.0	.7%	1.6%
RADIO SCIENCE	1B1-4	16.6	3.0%	1.3%						2.0%
ADVANCED SYS	1C1,2									2.5%
OFFICE OF APP	1D1			.9%						.2%
PREVENT MAINT	2A1	128.6	23.3%	17.2%	87.1	16.4%	19.5%	72.0	12.8%	9.6%
CORR MAINT	2A2			.6%			.4%	338.0	59.9%	11
CORR MAINT	2A3						.3%			.4%
PERS TRAINING	2B1	24.2	4.4%	15.6%	23.6	4.4%	8.7%	30.0	5.3%	5.5%
ENG SUPPORT	2C1			.2%			.2%	16.0	2.7%	.5%
ENG SUPPORT	2C2						.8%			2.3%
MINOR MOD	2C3						.1%			.4%
MAJOR MOD	3A1						7.7%			1.7%
MISC	3C1			.1%			< .1%			.6%
ADMIN				.8%			1.6%			3.3%
ABSENCE		32.0	5.8%	5.9%	16.0	3.0%	7.1%	12.0	2.1%	6.9%
FIRST AID TNG				.5%			1.1%			.1%
EYE EXAMS				< .1%			< .1%			.1%
SAFETY				< .1%			.1%			< .1%
OTHER PROJ										
PIONEER-6	1A1									.7%
PIONEER-7	1A1									.4%
PIONEER-8	1A1									.2%
PIONEER-9	1A1						.5%			.5%
DSN 1SEE	1A1						.2%			
STATION TOTALS		552.0			532.0			564.0		
M&I TOTALS		211.5			403.0			888.5		
OVERALL TOTAL (HRS)		763.5		25,302.5	935.0		28,453.1	1452.5		42,404.1
		DSS-11			DSS-12			DSS-14		

Fig. B-4. Manpower Utilization Report (Data Processing Unit, Goldstone)

WORK ORDER

DSS _____ PRIORITY _____ SUBSYSTEM _____		NUMBER **44721**	

ORIGINATOR	EXT
DATE	ORIGINAL DATE
TECHNICIAN(S)	
WA REVIEW ☐	P.M. No.

Emergency	☐	EER No.
Unscheduled	☐	DR No.
Training	☐	ECO No.
Other	☐	
Scheduled	☐	DOY
From (GMT)		To (GMT)

Description of Trouble or Service Requested:

_____

_____

_____

_____

Instructions: _____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Work Performed _____

	TIME
Start	
Finish	
No. of Persons	
Total Manhours	

CLEAN-UP	☐
Configuration	☐
Check-Out	☐

INITIALS	
S.O.D.	
Crew Supv.	
M & I Engr.	
M & I Supv.	

Follow Up Action Required _____

_____

Completed by	Date

ORIGINAL

Fig. B-6. Work Order (Maintenance and Integration Unit, Goldstone)

SERVICE REPORT

DSN CONTROL NO   CAT   PRIORITY   TYPE   NO

FROM FAC   DATE IN   FOR DMC   CMF   SHOP

TRAVELER   SHP NO   COMPLEX MEMO NO   DUE OUT

REQ NO

INCOMING INSTRUCTIONS

IN INSPECTION

SERVICE REQUIRED

OIL SAMPLE   ANALYSIS   STORAGE LOCATION   OPERATIONAL CHECK   REPAIR   CALIBRATE

MODIFY PER ECO

SERVICE OVERHAUL IN ACCORDANCE WITH

RE TM/DOC

SERVICE PERFORMED   IP INSPECTOR

EQUIPMENT IS NOT SERVICEABLE

SEND TO VENDOR FOR   REP   MOD   EXC

NO REPAIR OR CAL REQUIRED, EQUIPMENT TESTED GOOD

SUSPENDED WORK FOR   STORAGE LOCATION

	ON	RESUMED WORK		ON	RESUMED WORK
PARTS			DOCUMENTATION		
TEST EQUIPMENT			DMC ENG SUPPORT		
RECRYSTALIZATION			JPL ENG SUPPORT		
			OTHER		

REQ. NO.   REQ. DUE DATE

REPAIRED   SERVICED/OVERHAULED IN ACCORDANCE WITH

MODIFIED PER ECO

ANALYZED

CALIBRATED

CALIBRATION DATA

CONDITION RECEIVED   IN TOL   OUT OF TOL   FAILED

ADJUSTED   YES   NO   FULL SPEC.   LIMITED SPEC.

PROCEDURE NO

DSN CONTROL NOS OF TEST EQUIPMENT USED DURING CALIBRATION ACTIVITIES

STD CAL INT   RECOMMENDED INT   NEXT DUE DATE

DATE SERVICE ACTION COMPLETE   TECHNICIAN   FN INSPECTOR

SERVICE TIME   FINAL DISPOSITION

RETAINED AT   DMC   CMF   STORAGE LOCATION

HOURS   SENT TO ORIG FAC

TST ANL   SENT TO DLF

REPAIR   SENT TO ____ FOR DYNAMIC TEST

MOD   SENT TO

CAL   SENT TO VENDOR   REP   MOD   EXC   WARRANTY   YES   NO

SCRAPPED

TOTAL   TRAVELER NO   SHIPPER NO   DATE

COMMENTS

CHANGED (MODEL PART NO etc)

DATA ENTRY COMPLETED   DATE   INI

JPL 6311/DSN (6/81)

Fig. B-6. Service Report of Barstow Production Control Database (Section 377)

Top Down Implementation Plan for System Performance Test Software

G. N. Jacobson and A. Spinak

Operations Sustaining Engineering Section

This article describes the Top Down Implementation Plan used by the Operations Sustaining Engineering Section for the development of System Performance Test software during the Mark IV-A era. The plan is based upon the identification of the hierarchical relationship of the individual elements of the software design, the development of a sequence of functionally oriented demonstrable steps, the allocation of subroutines to the specific step where they are first required, and objective status reporting. The results are meaningful determination of milestones, improved managerial visibility, better project control, and ultimately a successful software development.

I. Introduction

The Mark IV-A era represents significant changes in the operating environment and the hardware configuration within the DSN. System Performance Test (SPT) software will be needed to test and verify the operational integrity of the DSN during the Mark IV-A implementation.

The SPT software package being developed by the Operations Sustaining Engineering Section will consist of a test executive and a set of seven applications tasks. Basically, the executive distributes input data to the applications, provides resource allocation services, and performs common processing functions such as dumping, display generation, and test procedure reading. The application tasks are each designed to test a particular system. Tracking, Telemetry, Command, Monitor and Control, Very Long Baseline Interferometry, Radio Science, and Frequency and Timing will all be supported by SPT software for the Mark IV-A configuration.

The SPT software will reside in the System Performance Test Assembly (SPTA). The SPTA, which also serves as the backup Complex Monitor and Control (CMC) computer, will be a Modcomp Classic 7345. The operating system will be a modified version of the MAX IV Operating System supplied by the computer manufacturer.

The software development effort for the Mark IV-A Project, summarized above, represents the largest project undertaken by the SPT Software Support Group. To achieve the technical, budgetary, and scheduling goals of the project, a top down implementation plan has been created that always develops and maintains the SPT system in a continually cycling, demonstrable fashion. It is the intent of this paper to describe the history, justification, and components of the SPT implementation plan.

II. Background

The success of large software development efforts has improved throughout the industry. These improvements are largely attributable to the application of a technological wave of new approaches that have been loosely referred to as

structured programming. More explicitly, the new technologies include replacement of flowcharts via usage of design languages, elimination of the GOTO by confinement to a small complete set of logical constructs, increased emphasis and formalization of the role of the programming support librarian, increased emphasis on reviews via usage of either structured walk-throughs or inspection teams, and a reorganization of programming personnel into the chief programmer team formation. Unfortunately, despite the considerable progress that has been made, many projects still fail to meet their schedule, have cost overruns, and the end product never quite operates as reliably as intended. In any event, even for supposedly successful projects, the cost of software is still too high.

A major reason for these continuing software difficulties and continued high costs, despite advances in technique, is that the impact of the aforementioned technical advances is limited when constrained by the effects of traditional management techniques. All of the previously mentioned structured programming techniques deal with the programmer and programming. None deal directly with the issues of planning and managing a large-scale software development. The industry is generally using the same planning and managing approaches it has always used, and these have frequently proven to be unsuccessful. The result is that the manager continues to have little visibility and little effective control over the developing system. If the manager had a mechanism that permitted him to arrive at a meaningful implementation plan; permitted him to objectively assess the project's status as it developed; provided him clear visibility of the development activity; considered cost, schedule, manpower and the chosen design, then the manager would be in a position to truly manage the project and lead it to a successful conclusion at minimal cost.

The SPT top down implementation plan fills this gap. The SPT plan is to management what structured programming is to the programmer. As with structured programming, which is complemented by the SPT plan, the SPT plan improves visibility, meaningfulness and orderliness. It allows the manager to start the project off on the right path, closely monitor the software development as it progresses, and ultimately to bring the project to the desired successful conclusion.

III. Implementation Plan Selection Criteria

The top down method appears to have been first espoused by Dr. H. D. Mills of IBM. Though this discussion, and other discussions in the computing literature, advocated top down

implementation, little advice was presented on how to plan a top down implementation. Stating that a computer program should be implemented in a top down sequence is insufficient for a large software development. Due to the complexity of its hierarchical structure, literally thousands of top down implementation sequences may be possible. Thus, selecting the proper top down implementation sequence becomes a very significant issue.

For example, one could implement all the subroutines at a particular level for the entire system, followed by all the subroutines at the next level across the system, and so on, until finally the bottom level subroutines are implemented. There are those in the industry who advocate this sequence. This would be top down, but in our opinion, represents an inferior implementation sequence. This is because bottom level subroutines usually are required to provide a demonstration of a complete operational system function. Thus, for most of the system's development, very few operational functions would be demonstrable. However, early functional demonstrability is one of the main benefits that should be achieved from top down implementation. Alternatively, many top down sequences might conflict with expected equipment delivery dates.

Thus, selecting the most appropriate top down implementation sequence is an important issue. The SPT plan is using a comprehensive methodology which has been developed for creating and maintaining an optimal top down implementation plan. This technique provides broad benefits to the ensuing software development.

IV. Top Down Concepts

Top down design refers to a method of designing a computer program wherein higher level or calling segments are designed before lower level or called segments. This does not mean that all segments at one level must be designed, or named, before creating the design, or name, of any segments at the next level. It means that if one were to consider the system's subroutine hierarchy as a treelike structure, then along each branch of the tree, subroutines are defined and chosen for design, starting from the top of the hierarchy and working down.

Top down implementation refers to the development of a computer program in a downward hierarchical sequence along each branch of the program's subroutine hierarchy. Design, documentation, coding, integration and testing usually are

concurrently performed on different portions of the developing system. In a top down sequence, these are performed along each branch simultaneou .y under development.

V. Preparation of the SPT Top Down Implementation Plan

A viable software implementation plan can only be prepared after a sufficient quantity of system analysis activities have occurred and before the detailed implementation has begun. The plan is then used to launch the implementation phase for a large-scale software development. In operation, the SPT approach is based upon the utilization and interplay of three documents. They are the Subroutine Hierarchy, the Network of Demonstrable Functions (NDF), and the Software Status Report.

In a nutshell, the Subroutine Hierarchy represents a design abstract for the computer software. The Network of Demonstrable Functions represents a functional abstract of the operational system. The Software Status Report relates the software design to the NDF system functions for the purpose of scheduling the software and maintaining the status of software development. The main point of this nutshell description is that attentive preparation of these documents results in a meaningful schedule that allows management to have real visibility in areas such as the software's true status, cost to completion, and time to completion.

VI. Subroutine Hierarchy

The Subroutine Hierarchy is a high-level representation of the structure of the hierarchical design of the computer program. It readily conveys a high-level image of the design being represented, showing all of the parts constituting the design, their hierarchical relationship to each other, their categories and, to a degree, their functions. All of the segments must represent small subroutines, perhaps averaging 25 to 50 higher-order language statements.

The Subroutine Hierarchy evolves as the design and the software evolve. Initially, when the implementation plan is first prepared, the Subroutine Hierarchy represents the intended design structure of the computer program. At completion of the software development, it represents the actual structure of the computer program At all intermediate stages, it is kept current and represents the currently projected structure of the program.

For a large computer program, with perhaps one thousand or more subroutines, the subroutines may be treated in a statistical manner for the purposes of making estimates,

schedules and plans This is one of the principles of the SPT approach. Namely, by partitioning a large computer program into its elemental pieces (subroutines), the effects of isolated misjudgement (e.g., size or complexity) relative to any individual subroutine tend to average out over the total program development, and do not affect the overall outcome. The effect of frequent misjudgement of the same characteristic of many subroutines (e.g., development time) tends to become quickly apparent and serves as a reliable indicator of development trends and ultimate results (if not corrected).

The Subroutine Hierarchy enables the software designers to conveniently conceptualize about the program and its parts, and to visualize the hierarchical organization of the program. It communicates in an overall conceptual manner the structure of the design. It is the essential design element, representing the components of the program's design for the purpose of planning and tracking the implementation of that design. It thus becomes the primary determinant in estimates of cost and memory size for the computer program.

Figure 1 shows a portion of the subroutine hierarchy for the SPT Project. Due to the large number of subroutines, the hierarchical structure is represented in a horizontal rather than vertical (or treelike) manner. Varying hierarchical levels are represented by varying levels of indentation. The hierarchy identifies both the symbolic name and descriptive name of the subroutine. It identifies the particular step in the SPT implementation plan where the subroutine will be first required. (The next section will elaborate on the definition of steps.) Only the first occurrence of a subroutine in the tree is expanded to the bottom level. Subsequent occurrences of any subroutine use a reference number to identify the line number of the first occurrence. If the subroutine itself invokes other subroutines, an asterisk is used to indicate the full expansion can be found at that first occurrence.

VII. Network of Demonstrable Functions

The structure of the software design and the identification of the constituent subroutines have been described as part of preparing the Subroutine Hierarchy. No discussion has yet occurred relative to the development sequence of these subroutines, nor relative to the individual milestones that will be scheduled and tracked during the development. This is where the NDF comes into play.

The NDF is that part of the SPT implementation plan that identifies the individual increments, or steps, and the sequence of development for those steps. The steps are scheduled, developed, tracked, integrated, tested and eventually internally accepted. In other words, on the surface it is a "Pert-like network." Beneath the surface there are a number of aspects of

the NDF that must be explained before its value can be fully grasped.

First of all, the NDF must be created by personnel that have an in-depth functional understanding of the application, its requirements, and the expected operational characteristics of the system. The personnel assigned to the NDF task will have acquired the necessary knowledge as a result of their prior system analysis activities. If they do not have this knowledge, they must first acquire it before they can hope to create a meaningful, detailed, functionally oriented plan of demonstrable steps.

Secondly, the NDF steps are oriented towards functions from the user's standpoint, not from the programmer's standpoint. For example, "output directive/menu index" is a typical step. This is as opposed to "build test configuration table," which would occur internally within the computer and not provide the user direct observation of the step having occurred. On the other hand, a step such as "print test configuration table" could be demonstrated to the user. Successful demonstration of this would imply successful construction of the test configuration table.

This leads us into the third important aspect of the NDF. To the maximum extent possible, steps of the NDF should be readily demonstrable to an observer who is not a programmer. Those few steps that are not readily demonstrable to such an observer must, nevertheless, still be demonstrable. This demonstrability is the only basis upon which an objective determination can be made as to the completion of the step.

The principle of demonstrability leads us to a fourth important aspect of the NDF. The development sequence of demonstrable steps must correspond to a natural functional sequence of increasing functional capability. To put it another way, from the user's operational standpoint, it must be a sequence which demonstrates "first things first."

A fifth important aspect of the NDF is that the steps must each add on to an already cycling system. Each new step must be directly integrated into the cycling system, producing a continuously increasing functional capability that is always demonstrable. Steps required to demonstrate a new step must be implemented and integrated prior to the integration of the new step. In terms of subroutines, this means that for a particular step, those subroutines that are required for invoking a particular segment of that step must be implemented as part of that step or as part of a prior step. In order words, the design must be implemented in a top down sequence along each branch of the subroutine hierarchy. Lower level sub-

routines that are not required for demonstration of the particular step are to be left as stubs until a step requiring those subroutines is undertaken.

One final aspect of importance is that the steps must fit together to comprise functional paths of the system. In actuality, to create the NDF, the functional paths are defined first, and the paths are then broken down into a sequence of steps. Each path corresponds to a relatively independent (but not necessarily totally independent) major function of the operational system.

As an example, the Telemetry Path of the SPT NDF is shown in Fig. 2. The Telemetry Path itself consists of a main path, a long loop path, and an Automatic Total Recall Subsystem (ATRS) path. For ease of reference, each step is given a number, preceded by a path identifier. Increments of 10 are used to allow insertion of additional steps, should this prove necessary because of changing requirements or priorities. Dashed lines between paths imply dependencies; with respect to Fig. 2, implementation of the ATRS path is dependent upon completion of the capability to accept directives.

VIII. Software Status Report

The Software Status Report ties the Subroutine Hierarchy and the NDF together by relating the design elements to the demonstrable steps. This is the fundamental point from which the value of the Software Status Report, and even the SPT Implementation methodology, is derived. The Software Status Report meaningfully relates the design to demonstrable functions and the corresponding schedules.

In concept, the document is very simple. For each step from the NDF, the corresponding required subroutines from the hierarchy are listed. Each subroutine is allocated to a single step, the first step from the NDF that requires the particular subroutine. Consequently, the subroutines listed under a particular step are just those subroutines still required for demonstration of the particular step's function. Other subroutines may also be required for the step, but they would not be listed with the step if some prior step already required the subroutines. For status tracking purposes, columns are provided where design, code, documentation, test size and other status fields can be checked off for each subroutine. These fields will be recorded as complete or will contain the date set for completion. No attempt is made to allow percentage estimates of completion by the programmer. Statistical data provided in the Software Status Report is based on treatment of individual segments as statistical equivalents. In addition, the Software Status Report includes a description of each step, i.e., the function that is being demonstrated by

the particular step. The report also identifies the qualifications or limitations, if any, that apply to the step's demonstration and the requirements that the step fulfills.

Whereas in concept the Software Status Report is very straightforward, creation of the report requires a thorough functional understanding of the system and of the corresponding design as represented by the Subroutine Hierarchy. Only with such knowledge as a base could the programming staff hope to allocate specific subroutines to each NDF step.

Thus, the Software Status Report contains all of the steps from the NDF and all of the assigned subroutines from the hierarchy, along with the development status for each subroutine and step. With automated support, highly objective status reports are easily generated from this data base. Technical and administrative management are provided accurate visibility into the status of the total software development.

Figure 3 shows the Software Status Report for a typical step from the SPT Implementation Plan. Fig. 4 provides a brief description of each of the fields on the report. Fig. 5 contains a Management Summary for one of the paths on the SPT NDF.

IX. Summary

The SPT approach to top down implementation planning is based on certain premises. One of these premises is that by minimizing or eliminating large unknowns, management has the best chance of accomplishing the project's goals. If there is some large functional area for which management has little basis, other than someone's intuition, for expecting the implementation to take say six months, with a particular size staff, as opposed to say three years, then the project is in a precarious position. A large nebulous function which has only been quantified at its total level by intuition, even though based on experience, is a dangerous unknown. The obvious way to get better control of a big unknown is by reducing it to many small pieces, some of which may be small unknowns. To put it in other terms, analyzing the task and breaking it down into many smaller pieces eliminates the risk of the large unknowns. There may still be some unknowns or surprises, but the potential absolute effect of a misjudgement relative to a small task is going to be inherently smaller than for a misjudgement associated with the much larger original task. An important additional aspect is that in the process of decomposing the

original function, understanding occurs, and for the most part, comprehension replaces intuition.

Also, by decomposing a system into a large number of small pieces, a point is reached where the individual pieces can be treated for planning purposes as statistically equivalent. At the management level, the differences in size or complexity of individual small subroutines is of minimal importance. As subroutines are implemented, actual data should be used to update the estimated statistical characteristics of the average subroutine. For example, suppose an original memory allocation of 128K is made for 2000 subroutines. This averages 64 memory cells per subroutine. Suppose, after 200 subroutines have been implemented, 15000 cells have been used. This would show an actual average of 75 cells per subroutine with the trend total being 150K for all 2000 subroutines. Thus, with only 10% of the segments implemented, a reliable danger signal has been raised, and the signal includes the magnitude of the forecasted overrun. With such an early warning, management still has time to take some appropriate effective action to act upon the issue before it becomes an actual problem.

The key basis for planning and tracking the implementation is assigning the implementation of each subroutine to a single step. This is where it all comes together. But it must be done with a great deal of care and precision. The correlation between the Subroutine Hierarchy and the Software Status Report must be accurate. When design or plan changes occur, and they will, changes must be made to both documents. Both of these documents should be looked upon as evolving documents, but they must evolve concurrently and in parallel.

Why does this "single step" premise form the basis to the SPT approach to implementation planning? Because everything is accounted for. Each subroutine appears for implementation on only one step — the first step that requires the subroutine. The effort required for each step can be considered to be a function of the number of subroutines in the step. The programming implementation budget can be spread over the steps in proportion to the number of segments in each step. Then, if you are on schedule, you are on budget. Subroutines don't appear redundantly (on more than one step) to confuse the bookkeeping. Everything balances and all subroutines are able to be tracked. A full decomposition of the system into subroutines and a careful and complete assignment of those subroutines to a series of well-defined, demonstrable steps is of fundamental importance to successful usage of the SPT methodology.

SYMBOLIC NAME 1 2 3 4 5 6 7 8 9 *	DESCRIPTIVE NAME	CLASS /STEP	LINE NUM	REFER NUM
. . . . . CSTMAC	ACTIVE MODE PREDICTOR	C90	287	
. . . . . . CDBCOM	DOUBLE INTEGER COMPARE		288	۱7
. . . . . . CSHDGB	HSD BIT ACQUIRE		289	۵
. . . . . . CDBADD	DOUBLE INTEGER ADD		290	
. . . . . . CSTMEV	EVENT PREDICTOR	C50	291	
. . . . CDBADD	DOUBLE INTEGER ADD		292	277
. . . . CSHDGB	HSD BIT ACQUIRE		293	195
. . . . CDBSUB	DOUBLE INTEGER SUBTRACT		294	284
. . . . CDBCOM	DOUBLE INTEGER COMPARE		295	197
. . . . . CSRCRP	RECALL QUEUE RESPONSE PROC.	C90	296	
. . . . . CSCHEK	STATUS CHECK		297	258*
. . . . . CSSUPD	SUSPEND RESPONSE PROC.	C130	298	
. . . . . CSTMSU	MODEL RE-ADJUST	C130	299	
. . . . . . CDBCOM	DOUBLE INTEGER COMPARE		300	197
. . . . . . CDBADD	DOUBLE INTEGER ADD		301	277
. . . . . . CDBSUB	DOUBLE INTEGER SUBTRACT		302	284
. . . . . CSCCLO	CONTROL CENTER RESPONSE PROC.	C60	303	
. . . . . CSHDGB	HSD BIT ACQUIRE		304	195
. . . . . CSCHEK	STATUS CHECK		305	258*
. . . . . CSMOUP	MODE CHANGE RESPONSE PROC.	C100	306	
. . . . . CSHDGB	HSD BIT ACQUIRE		307	195
. . . . . CHXASC	HEX TO ASCII CONV.		308	217
. . . . . GEFMSG	DISPLAY EVENT MESSAGE		309	31
. . . . . CSCHEK	STATUS CHECK		310	258*
. . . . . CSTIRM	RADIATION TIME TEST		311	281*
. . . . . CSTWOV	WINDOW OVERRIDE RESPONSE PROC.	C120	312	

Fig. 1. Subroutine Hierarchy
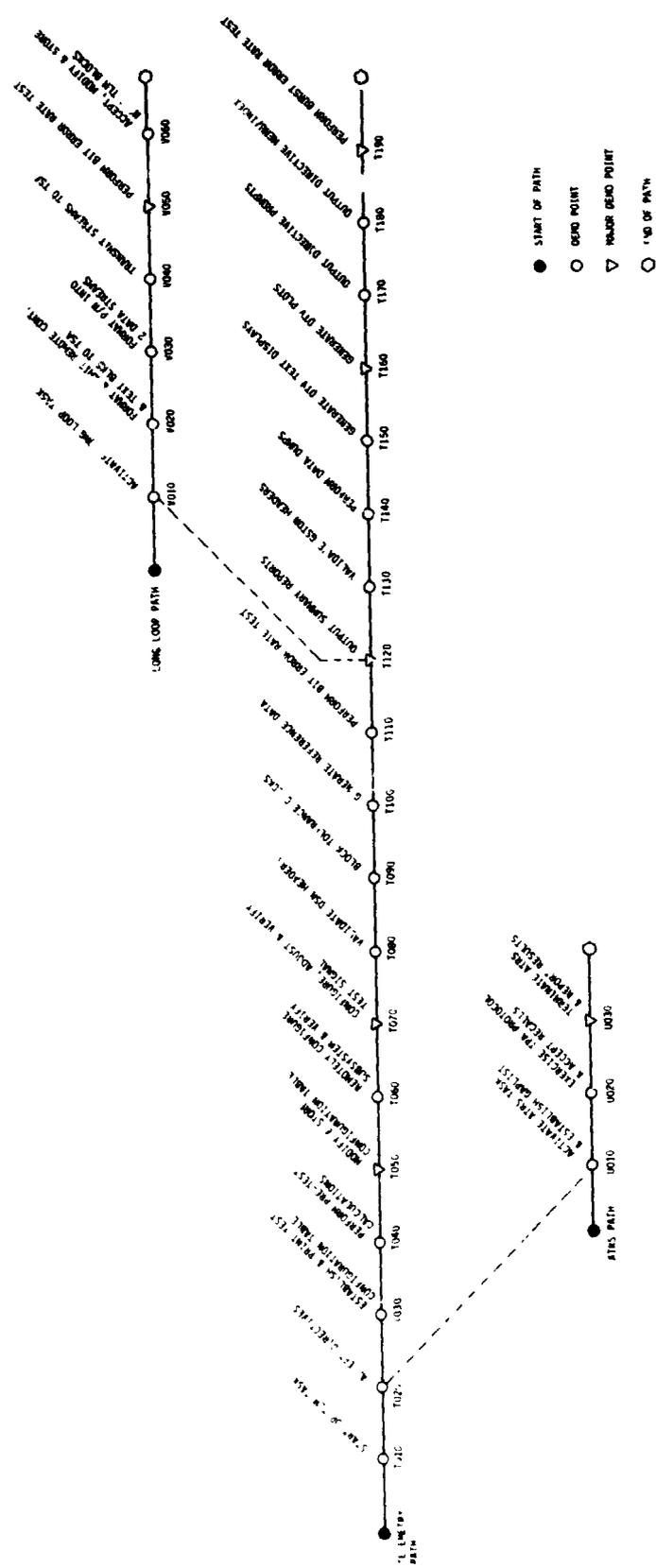
Fig. 2. Network of demonstrable functions—telemetry path

STEP· 010   Transfer Operator Entry to SYMBIONT

DESCRIPTION:

This step provides an operator entered directive to the Symbiont.

INTENT:

This step demonstrates that an operator entered directive is
placed in an SSB and entered in the Symbiont Queue. This step
simulates the eventual LMC, LAN Handler, FIDM to Symbiont
interface. This also provides a basic buffer and queue manage-
ment mechanism.

SOURCE:

AUTHOR: A. Spinak

DUE DATE: 03/02/82

COORDINATOR: T. GREER

STATUS: DATE   TYPE

DSGN RVU: 02/09/82   TEAM

TEST: 03/11/82   ACCEPTED

ANOMALIES: 0

NOTES:

Most or all of this step's software is in the nature of a tempo-
rary workaround or Environmental Interface Module (EIM). To
prove, the SSB and the Symbiont Queue should be checked.
No queue boundary conditions will be demonstrated.

SEGMENTS	DESCRIPTION	CL	---- DATE	DESIGN PERSON	-- ST	---- DATE	CODE PERSON	-- ST	QA	CMP	LINES
SIOINP	OPERATOR 2 SPT SIMULATION	I	02/01/82	MJB	•	02/11/82	MJB	*		*	21
SIOPOM	PARSE OPERATOR MESSAGE	I	02/02/82	MJB	*	02/11/82	MJB	*		*	21
SIOPAK	PACK CD SSB	I	02/03/82	TO	•	02/11/82	TO	*		*	30
SIOCBB	LOAD TEST COMM BUFF BLOCK	S									0
SIOSSB	LOAD TEST SSB BLOCK	S									0
SIOLNK	CHANGE LINK ID	S									0
ORQENQ	PLACE NODE ON QUEUE	I	02/02/82	TCG	*	02/16/82	MJB	*		*	51
ORQINI	INIT FREE Q NODE POOL	I	02/05/82	TCG	•	02/16/82	TCG	*		*	14
OPOPQU	POP FREE QUEUE NODE	!	02/05/82	MJB	•	02/16/82	TO	*		*	15
GBIQUE	I/F TO QUEUE ENTRY	I	02/02/82	TO	*	02/12/82	TO	*		*	35
GQINIT	I/F TO INIT QUEUES	I	02/04/82	TO	•	02/12/82	TO	*		*	14
OPSHQU	PUSH FREE QUEUE NODE	I	02/05/82	MJB	•	02/12/82	MJB	*		*	16
SPINIT	SPT INITIALIZATION	I	02/10/82	TCG	•	02/15/82	TCG	*		*	42
GBINIT	I/F TO INITIALIZE BUFFERS	I	02/22/82	MJB	*	03/05/82	TO	*		*	58
GBPOOL	I/F TO RELEASE BUFFER	I	02/22/82	MJB	•	03/05/82	MJB	*		*	55
GETBUF	I/F TO GET BUFFER	I	02/22/82	MJB	•	03/05/82	MJB	*		*	56

Fig. 3. Software status report—step 010

197

STEP: NUMBER AND NAME OF THE STEP

DESCRIPTION:

DESCRIPTION OF THE SYSTEM FUNCTIONS AND CAPABILITIES IMPLEMENTED IN THIS STEP AND A CONCISE SUMMARIZATION OF THE DEMONSTRATIONS TO BE TESTED' (UP TO 3 LINES)

INTENT:

EXPLANATORY INFORMATION THAT IS USEFUL IN INTERPRETATION AND COMPREHENSION OF THE STEP. (UP TO 5 LINES)

SOURCE.   REFERENCE TO THE REQUIREMENTS MET BY THIS STEP.

AUTHOR: ORIGINATOR OF THIS STEP INPUT.

DUE DATE: DATE WHEN THE STEP WILL BE COMPLETED AND READY FOR ACCEPTANCE.

COORDINATOR: COORDINATOR OF THE IMPLEMENTATION OF THE STEP AT THE DETAIL LEVEL.

STATUS	DATE	TYPE
DSGN RVU	DESIGN REVIEW DATE	TYPE OF REVIEW (TEAM, CDE OR PROG)
	TEST STATUS DATE	TYPE OF TESTING (CHECKOUT OR ACCEPTED)
		(IF TYPE IS BLANK, DATE IS PLANNED DATE)
		(ELSE DATE IS ACTUAL DATE)

ANOMALIES.   NUMBER OF OUTSTANDING ANOMALIES

NOTES

NOTES WHICH WOULD BE HELPFUL IN EXPLAINING THE STEP AND ITS IMPLEMENTATION. (UP TO 5 LINES)

THE SEGMENT COLUMNS ARE AS FOLLOWS:

STEP NO	-	STEP NUMBER
SEGMENTS		SEGMENT NAMES
DESCRIPTION	-	SEGMENT DESCRIPTION
CL		CLASSIFICATION CODE (I-IMPLEMENT, S-STUB, U-UNDEFINED)
DESIGN DATE	-	PLANNED DATE OF DESIGN COMPLETION
DESIGN PERSON	-	INITIALS OF PERSON RESPONSIBLE FOR THE SEGMENT'S DESIGN
DESIGN ST		DESIGN STATUS (* IF COMPLETED)
CODE DATE		PLANNED DATE OF CODE COMPLETION
CODE PERSON	-	INITIALS OF PERSON RESPONSIBLE FOR CODING OF THE SEGMENT
CODE ST		CODE STATUS (* IF COMPLETED OR IF REJECTED)
QA		QA STATUS CODE (* IF ACCEPTED)
CMP		COMPILATION CODE (* FOR CLEAN COMPILE)
LINES		NUMBER OF LINES IN ACCEPTED SEGMENT

Fig. 4. Software status report description

STEP NUM	STEP NAME	C	UNIQ	CUM	DESGN	CODE	CMP	TEST	ACEPT	DUE DATE	LINES
C01	Initialize Command Task		16	16	0	0	0	0	0		0
C02	Generate Command File		9	25	0	0	0	0	0		0
C03	Accept & Check Incoming Block		12	37	0	0	0	0	0		0
C04	Process Control & Status Block		8	45	0	0	0	0	0		0
C05	P ss Event Block		9	54	0	0	0	0	0		0
C06	Transmit Block to CPA		12	66	0	0	0	0	0		0
C07	Transmit File to CPA		16	82	0	0	0	0	0		0
C08	Recall File Directory		2	84	0	0	0	0	0		0
C09	Attach File to Queue		7	91	0	0	0	0	0		0
C10	Initiate Command Radiation		2	93	0	0	0	0	0		0
C11	Modify Standards-and-Limits Table		18	111	0	0	0	0	0		0
C12	Transmit Additional Directives		15	126	0	0	0	0	0		0
C13	Suspend Abort, Resume Radiation		9	135	0	0	0	0	0		0
C14	Verify Command Bits		11	146	0	0	0	0	0		0
TOTALS			146	146	0	0	0	0	0		0
PERCENTS				100	0	0	0	0	0		0
PROJECTION											

Fig. 5. Software status report management summary

Rates of Solar Angles for Two-Axis Concentrators

C. S. Yung and F. L. Lansing

DSN Engineering Section

Determination of the sun's position by the azimuth and elevation angles and its rate of change at any time of day are essential for designing 2-axis tracking mechanisms of solar concentrators. A study of the sun's angles and their rates is presented at four selected months of the year (March, June, September and December) and for seven selected latitudes (0, ±30, ±60, ±90) covering both the northern and southern hemispheres.

I. Introduction

Solar thermal-electric power systems are being studied for the Deep Space Network Energy Conservation Project to evaluate their technical and economical feasibility. The engineering design of high temperature, point-focusing two-axis tracking solar concentrators in an unattended mode of operation requires, among the many design specifications, a knowledge of the sun angles and angle rates to be incorporated in the automatic control loop parameters. To satisfy this need for different network facilities throughout the world, this study was initiated to determine the sun's position vector, rate of movement in the sky at any given location and at any time of year. Both northern and southern latitudes were investigated for their possible differences at different seasons.

II. Solar Angles

Because the earth's equatorial plane is tilted at an angle of 23.5 degrees to the orbital plane as shown in Fig. 1, the solar declination angle δ, which is the angle between the earth-sun line and the equatorial plane, varies throughout the year. This variation causes the changing seasons with their unequal periods of daylight and darkness. The geometric configurations of the position vector of the sun at any time of day relative to the earth center and relative to a horizontal plane passing by point P on earth, are presented in Figs. 2 and 3.

From Fig. 2, the latitude angle of location in radians L is defined as the angle between the location position vector OP and the equatorial plane (north latitudes are taken positive). The solar-azimuth angle in radians ϕ as shown in Fig. 3 is defined as the angle between the true south direction and the projection of the sun's position vector on the horizontal plane. The south direction is taken as the reference line for both northern and southern latitudes. The sign convention for the angle ϕ is that east of the south is taken positive, and west of the south is taken negative. The hour angle H, which is an indication of the local solar time, is changing with a rate of 15 deg/hr (360 deg/24 hours). The sign for H is taken positive in the morning, and negative in the afternoon, and zero at solar noon. The solar elevation, sometimes called the altitude angle β, and the solar zenith angle z are also shown in Fig. 3 as the angles the solar position vector makes with the horizontal plane and the vertical line, respectively.

Since the declination angle δ does not vary significantly from one day to the next within each month (Refs. 1, 2), a one-day representation of a given month was considered satisfactory. The variation of solar declination throughout the year from a maximum of about 23.5 deg on June 21 to a minimum of -23.5 deg on December 21, gave us a reason to select the twenty-first day of each month to be the representative day of the month. By treating the solar declination tabular data (in Ref. 2) as a periodic function with one com-

plete cycle per year, the following analytic expression was obtained by least squares curve fitting:

$$\delta = (\pi/180)(0.2833 - 23.188 \cos P - 0.15 \cos 2P$$

$$- 0.211 \sin P + 0.1155 \sin 2P) \tag{1}$$

where

$$P = \pi M/6 \tag{2}$$

and M is a month index ($M = 1, 2, \ldots, 12$).

The time of day, described by the hour angle H in radians, is also written as

$$H = \pi(12 - T)/12 \tag{3}$$

where T is the hourly time index in hours ($T = 1, 2, 3, \ldots, 24$). Note $T = 12$ means a solar noon time.

The direction cosines of a unit vector along the location position vector are found from the geometry of OP in Fig. 2 as: $\cos L \cos H$, $- \cos L \sin H$ and $\sin L$, with the X, Y and Z axes, respectively. Note that the equatorial plane lies on the X - Y plane, and point P' is the projection of point P (representing the location on earth) on the equatorial plane. The direction cosines of the sun's radiation unit vector can be also obtained from Fig. 2 as: $\cos \delta$, 0 and $\sin \delta$ with the X, Y and Z axes, respectively. The scalar product of the above two unit vectors will determine the cosine of the zenith angle z (or the sine of the elevation angle β):

$$\cos z = \sin \beta = \cos L \cos \delta \cos H + \sin L \sin \delta \tag{4}$$

On the other hand, a unit vector along the direction of true south ($\overline{PS}$ in Fig. 2) will have the direction cosines: $\sin L \cos H$, $-\sin L \sin H$ and $- \cos L$ with the X, Y and Z axes, respectively. Accordingly, the angle ω, from Fig. 3, which is the angle between the sun's vector and the south direction is obtained by scalar product as

$$\cos \omega = \cos \beta \cos \phi = \cos \delta \sin L \cos H - \cos L \sin \delta$$

or

$$\cos \phi = (\cos \delta \sin L \cos H - \cos L \sin \delta)/\cos \beta \tag{5}$$

An additional relationship for the azimuth angle ϕ, can be obtained also by multiplying scalarly the true east vector and the sun's position vector as shown in Fig. 3, to yield

$$\sin \phi = \cos \delta \sin H/\cos \beta \tag{5a}$$

At solar noon (i.e., $H = 0$), the azimuth angle ϕ is always zero at all locations throughout the year. Also, the angle β is always equal to $(90 - L + \delta)$ deg for both northern and southern latitudes (from Eq. 4). Equations (4) or (5) are subject to one constraint: the argument of the right-hand side should be within ± 1.

The hour angle limit H^*, which determines the hour angle at either sunrise or sunset as measured from solar noon, is given by equating the elevation angle β to zero. Using Eq. (4), the angle H^* (which is half the solar day) is written as:

$$H^* = \cos^{-1}(- \tan L \tan \delta) \tag{6}$$

where the term $(- \tan L \tan \delta)$ in Eq. (6) should not exceed 1, and the minimum value should not be less than -1. Hence, H^* (in radians) can vary between zero and π. If H^* equals to zero, the location on earth will be in complete darkness for the entire day. If H^* equals to π, the solar day will be 24 hours and the location will receive continuous sunlight for 24 hours. The above limiting conditions can only occur at higher latitude angles in the northern hemisphere (or lower angles in the southern hemisphere) than + 66.5 deg, which is a marginal latitude given by Eq. (6) at $\delta = 23.5$ deg. For example, the city of Bettles, Alaska, USA (with 66.5 deg north latitude) has twenty-four hours of sunlight in June and receives no sunlight in December. This fact could be explained by using Eq. (6) since the declination angle is +23.5 deg on June 21 and -23.5 deg on December 21 for all latitudes.

The azimuth angle at sunrise or sunset ϕ^* can be obtained, also, from combining Eqs. (5) and (6) at any latitude or declination angles where β is set equal to zero.

$$\cos \phi^* = - \sin \delta/\cos L \tag{7}$$

Special cases arise for Eqs. (4) through (6). First is the case during the equinox (March 21 and September 21) when the declination angle is zero for all latitudes. The elevation and azimuth angles, for these two months, become:

$$\left. \begin{aligned} \sin \beta &= \cos L \cos H \\ \cos \phi &= \sin L \cos H/\cos \beta \\ \sin \phi &= \sin H/\cos \beta \\ H^* &= \pi/2 \end{aligned} \right\} \quad \text{at } \delta = 0 \tag{8}$$

The above indicates that the day length (from sunrise to sunset) is always 12 hours during the equinox.

Second is the case of a location at the equator. The solar angles at different declination and hour angles are reduced to:

$$\left.\begin{array}{l} \sin \beta = \cos \delta \cos H \\ \cos \phi = -\sin \delta/\cos \beta \\ H^* = \pi/2 \end{array}\right\} \quad \text{at } L = 0 \quad (9)$$

Eq. (9) shows that the day length at the equator is always 12 hours.

Third is the case at the north pole ($L = +90$ deg). The solar angles at different declination and hour angles are reduced to:

$$\left.\begin{array}{ll} \beta = \delta & \text{if } \delta \geqslant 0 \\ \phi = H & \text{if } \delta \geqslant 0 \\ H^* = \pi & \text{if } \delta \geqslant 0 \\ \text{No sun} & \text{if } \delta < 0 \end{array}\right\} \quad \text{at } L = +90 \text{ deg} \quad (10)$$

The fourth special case is at the south pole ($L = -90$ deg) where the solar angles become:

$$\left.\begin{array}{ll} \beta = -\delta & \text{if } \delta \leqslant 0 \\ \phi = (\pi - H) & \text{if } \delta \leqslant 0 \\ H^* = \pi & \text{if } \delta \leqslant 0 \\ \text{No sun} & \text{if } \delta > 0 \end{array}\right\} \quad \text{at } L = -90 \text{ deg} \quad (11)$$

III. Angle Rates

The time rate of change of solar angles can be determined by differentiating Eqs. (3), (4) and (5) with respect to time for the hour angle, elevation angle and azimuth angle, respectively. The units of the angle rate are expressed in degrees per second. The elevation angle rate $\dot{\beta}$ (in deg/sec), at a given δ, L, can be expressed from Eqs. (3) and (4) as

$$\dot{\beta} = \frac{d\beta}{dt} = \frac{0.00417 \cos L \cos \delta \sin H}{\cos \beta} \quad (\beta \neq 0 \text{ des}) \quad (12)$$

Similarly, the azimuth angle rate $\dot{\phi}$ (in deg/sec), at given δ and L, can be expressed as

$$\dot{\phi} = \frac{d\phi}{dt} = 0.00417$$

$$\left\{\frac{\sin^2 H \cos L \sin \beta \cos^2 \delta - \cos H \cos^2 \beta \cos \delta}{\cos^3 \beta \cos \phi}\right\} \quad (13)$$

where β and $|\phi|$ must not equal to 90 deg.

Equations (12) and (13) are valid for any H, L, or δ angles. However, they are reduced to other forms for some special cases presented as follows:

A. Noon Time

At noon time ($H = 0$), the rate $\dot{\beta}$ will be zero for any latitude and any month; which means that β either reaches a maximum point or remains constant. The rate $\dot{\phi}$ is also obtained from Eq. (13) at noon time where:

$$\left.\begin{array}{l} \beta = 90 - (L - \delta) \\ \phi = 0 \\ \dot{\beta} = 0 \\ \dot{\phi} = -0.00417 \cos \delta/\sin (L - \delta) \end{array}\right\} \quad H = 0 \quad (14)$$

B. Sunrise or Sunset Time

The angle rates at sunrise or sunset times (denoted by an asterisk) are obtained from Eqs. (6), (7), (12) and (13) at any latitude or declination angles where

$$\left.\begin{array}{l} \beta^* = 0 \\ \phi^* = \cos^{-1} (-\sin \delta/\cos L) \\ \dot{\beta}^* = 0.00417 \sqrt{\cos (L + \delta) \cos (L - \delta)} \\ \dot{\phi}^* = -0.00417 \sin L \end{array}\right\} \quad \text{at } H^* \quad (15)$$

C. At the Equator

At the equator (where $L = 0$), the angles are obtained from Eq. (9). The angle rates are obtained from Eqs. (12) and (13) at any hour angle H as:

$$\dot{\beta} = 0.00417 \cos\delta \sin H / \sqrt{1 - \cos^2\delta \cos^2 H}$$

$$\dot{\phi} = 0.00417 \sin\delta \cos\delta \cos H / (1 - \cos^2\delta \cos^2 H) \quad \left.\right\} L = 0$$

$$(16)$$

At noon time, for a location at the equator, the rates $\dot{\beta}$ and $\dot{\phi}$ from Eq. (16) are reduced to:

$$\dot{\beta} = 0 \qquad \left.\right\} L = 0$$
$$\dot{\phi} = 0.00417 \cos\delta \qquad H = 0 \tag{17}$$

Eq. (17) is identical to Eq. (14) when setting $H = 0$.

Also, at sunrise (or sunset), for a location at the equator ($H^* = \pi/2$), the rates $\dot{\beta}$ and $\dot{\phi}$ from Eq. (15) or Eq. (16) are

$$\dot{\beta} = 0.00417 \cos\delta \qquad \left.\right\} L = 0$$
$$\dot{\phi} = 0 \qquad H = H^* \tag{18}$$

Equation (16) for the equator is plotted as shown in Fig. (6) at different H and δ angles. Peak values for $\dot{\beta}$ are evidenced at sunrise or sunset time in March and September ($\delta = 0$) as 0.00417 deg/sec. On the other hand, the peak values for $\dot{\phi}$ can reach infinity as $\delta \to 0$. As δ changes between ±23.5 deg, the rate $\dot{\phi}$ changes between ±0.0096 deg/sec.

D. At the North Pole

The angles β, ϕ, H^* at the north pole, from Eq. (10), are substituted in Eqs. (12) and (13) where $L = 90$ and $\delta \geq 0$, hence,

$$\dot{\beta} = 0 \qquad \left.\right\}$$
$$\dot{\phi} = -0.00417 \qquad L = +90 \text{ deg} \tag{19}$$

Eq. (19) is applicable for any hour-of-day variation at any month where $\delta \geq 0$.

E. At the South Pole

The angles β, ϕ, H^* at the south pole, from Eq. (11), are substituted in Eqs. (12) and (13), where $L = -90$, and $\delta \leq 0$. Hence,

$$\dot{\beta} = 0 \qquad \left.\right\}$$
$$\dot{\phi} = +0.00417 \qquad L = -90 \text{ deg} \tag{20}$$

Eq. (20) is applicable for any hour-of-day variation at any month where $\delta \leq 0$.

IV. Discussion of Results

A. Solar Angles

The computations of the solar angles at some selected hours of day, months of year and latitudes were compared against the tabulated data given in Ref. 2. A good agreement was found in all cases. Further, the rates of solar angle variations were determined at seven different latitudes covering northern and southern hemispheres including the equator. In the northern hemisphere, latitudes +30, +60, and +90 deg were selected. In the southern hemisphere, latitudes -30, -60, and -90 were selected.

For four selected months throughout the year (March, June, September and December), the solar angles were plotted as shown in Figs. 4 and 5. Figure 4 shows the solar elevation angles. If $\delta = 0$, the solar elevation angle reaches its peak (90 deg) at solar noon, and the sun's position vector travels in a vertical plane containing the east-west directions. Figure 4 also gives the solar elevation angles for the remaining six latitudes: ±30, ±60 and ±90 deg. Figure 5 shows the comparison of the solar azimuth angle for the seven selected latitudes. The profile of the solar altitude angles, however, is similar in both northern and southern locations of equal latitude. The two hemispheres experience an opposite climatic effect; the "summer" season, including the months of June, July and August in the northern hemisphere, is the "winter" season in the southern hemisphere.

At the north pole, the sun is barely visible at the horizon on March 21 and September 21 as evidenced from Eq. (10). In June, the sun maintains a constant +23.5 deg with the horizontal plane all day long. On the contrary, the south pole experiences 24 hours of daylight in December ($\delta = -23.5$) as evidenced from Eq. (11).

The azimuth angle is always zero (or 180°) at solar noon, and at sunset it is obtained from Eq. (7). In northern latitudes, the trace of the azimuth angle is mostly located in the southern quadrants of the horizontal plane; hence solar concentrators should be oriented facing south. The opposite situation exists in the southern hemisphere; i.e., solar concentrators in the southern hemisphere should be oriented facing north.

B. Angle Rates

Figures 6 through 10 show the time rate of change of solar angles ($\dot{\beta}$ and $\dot{\phi}$) at the selected latitudes. Figure 6 shows the angle rates at the equator as obtained from Eq. (16). The rate of elevation angle in March and September ($\delta = 0$) is constant at 15 deg per hour) throughout the day except at solar noon

where the rate is zero. The maximum rate of azimuth angle $\dot{\phi}$ is 0.0096 deg/sec at solar noon in June.

Figures 7 and 8 present the angle rates in the northern hemisphere where L = 30 and 60, respectively. In general, the angle rates in the northern hemisphere decrease as the latitudes increase. In the north pole, the rate of elevation angle is zero, while the azimuth angle travels at a constant rate of -15 deg/hr (-0.00417 deg/sec).

Figures 9 and 10 present the angle rates for the southern hemisphere where L = -30 and -60 degrees respectively. In all cases studied, with the exception of the north and south poles, the rate of elevation angle $\dot{\beta}$ decreases to zero at noon time from its maximum value at sunrise (or sunset) while the rate of azimuth angle $\dot{\phi}$ increases from its low sunset value to a peak value at solar noon.

The development of the above angle and angle rate analytical expressions for any month, hour of day, and latitude represents an initial study which provides the solar concentrator designer with a quantitative determination of the limiting sun's position and angle rates for an accurate automatic tracking mechanism.

References

1. Threlkeld, J. L., *Thermal Environmental Engineering*, Prentice-Hall, 1962.

2. ASHRAE, *Handbook of Fundamentals*, American Society of Heating, Refrigeration and Air Conditioning Engineers, 1972.
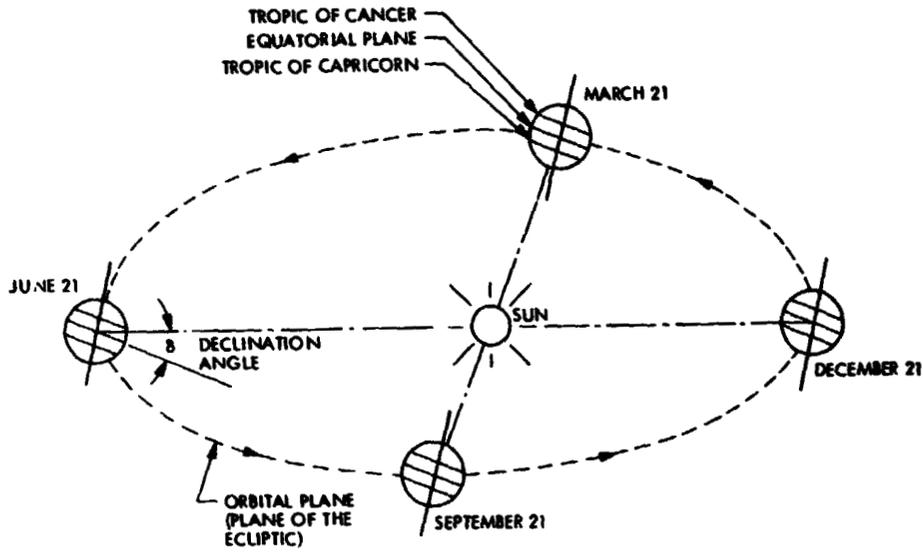
Fig. 1. The earth's motion around the sun
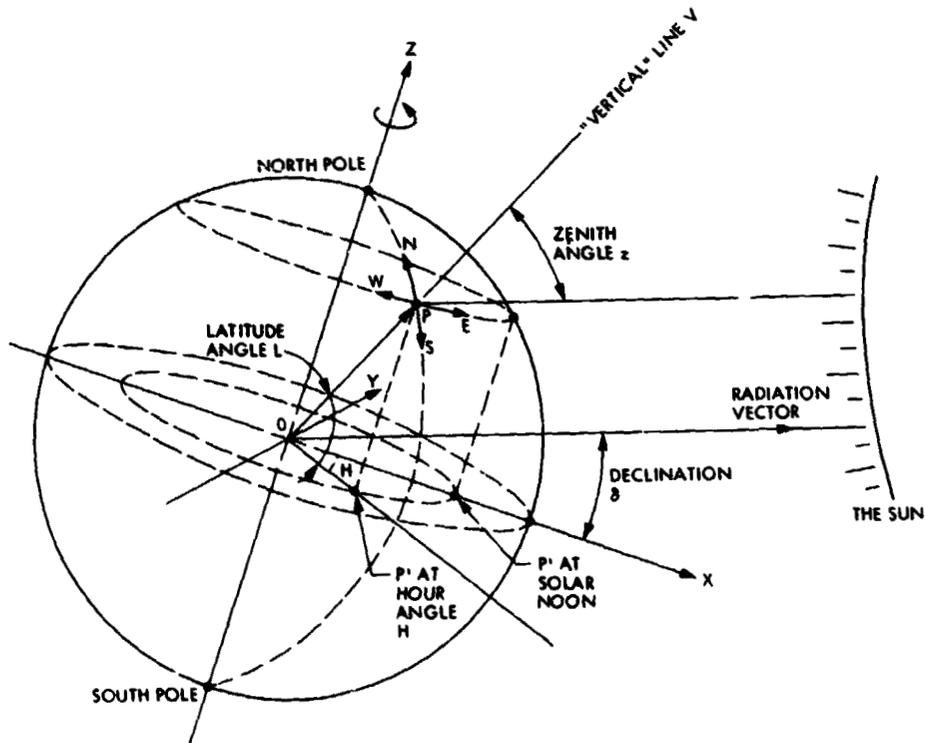
Fig. 2. Earth-sun angles
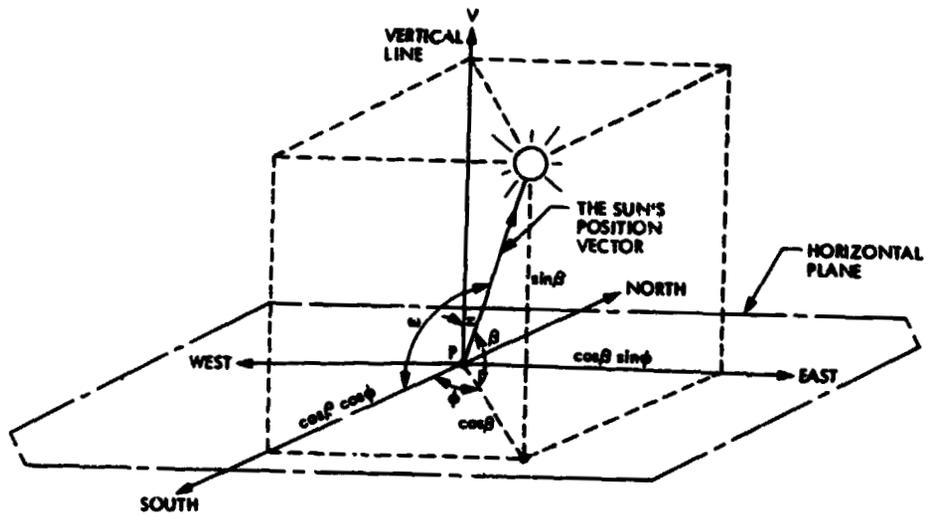
Fig. 3. Direction cosines of the sun's position vector viewed from a point on earth
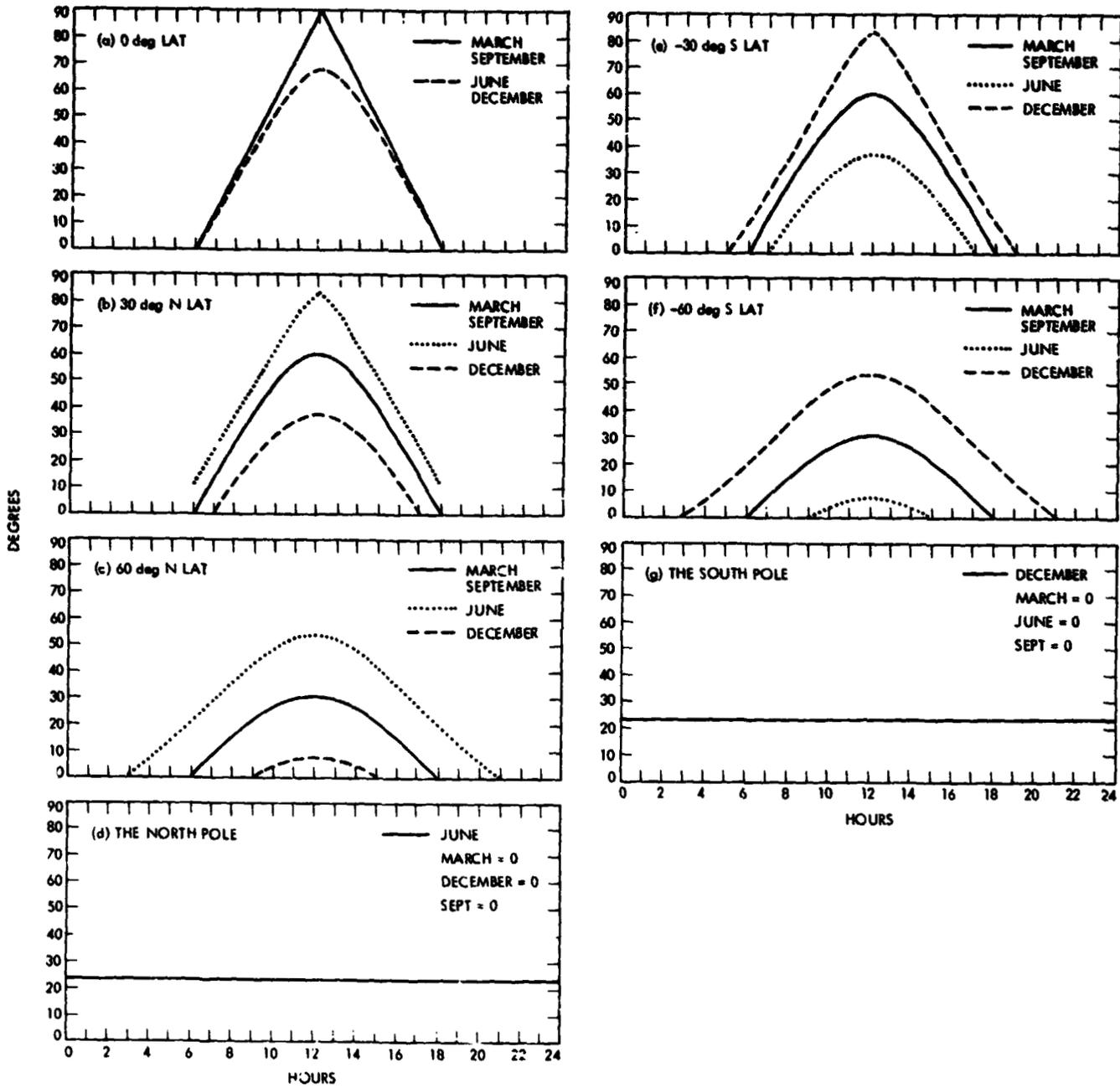
Fig. 4. Comparison of solar elevation angles for seven different latitudes
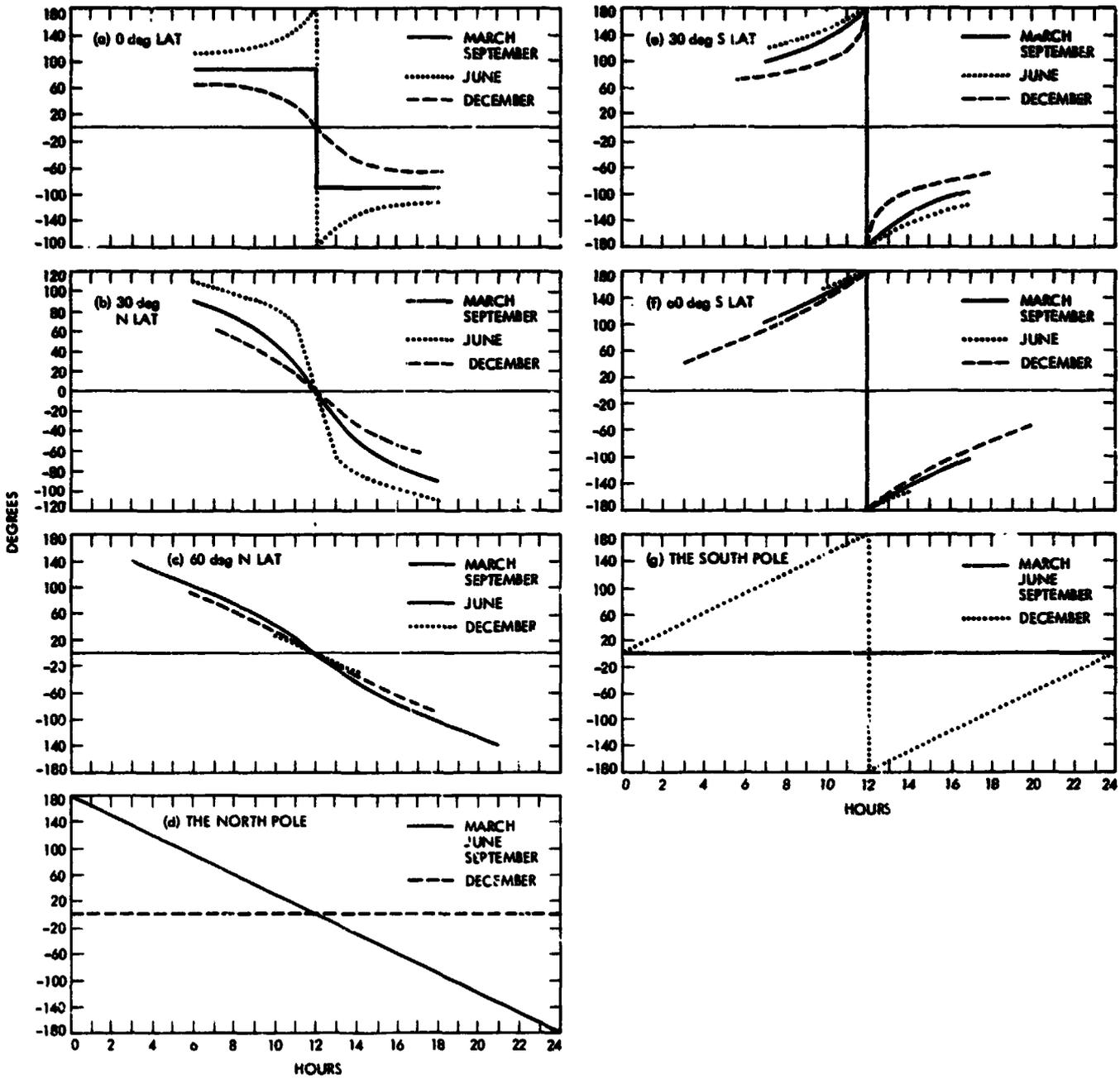
Fig. 5. Comparison of solar azimuth angles for seven different latitudes
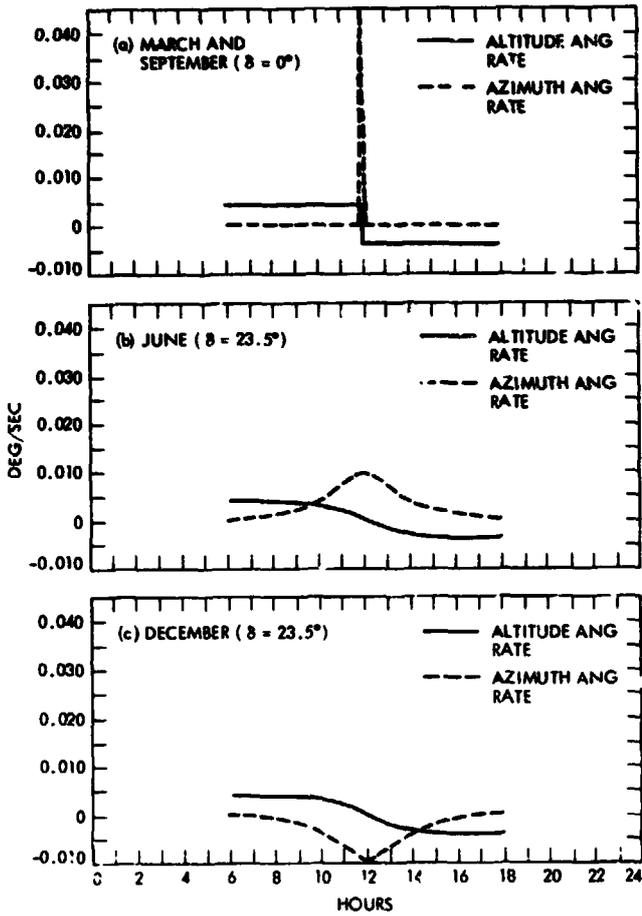
Fig. 6. Rate of solar angles for the equator
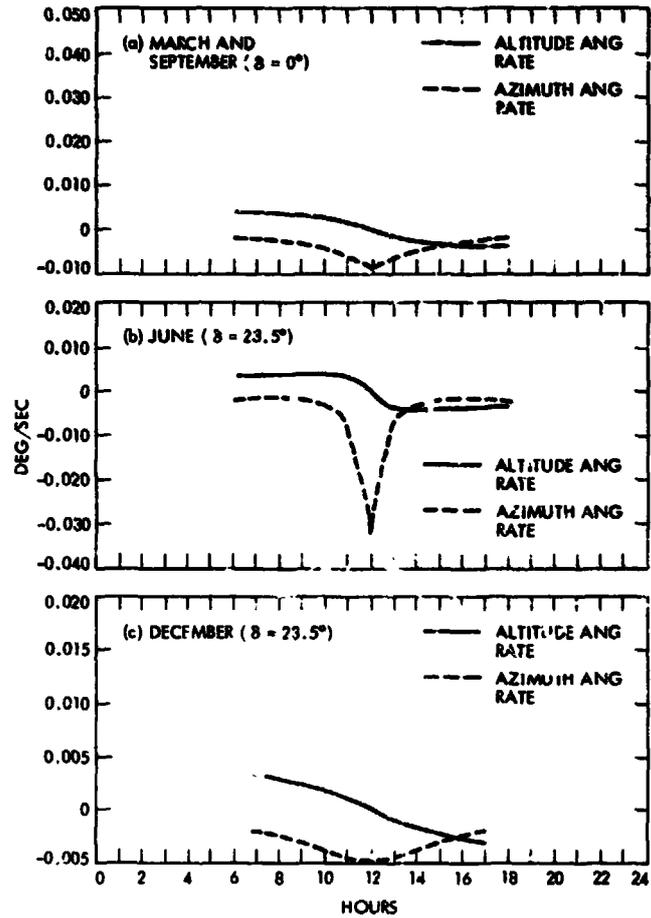
Fig. 7. Rates of solar angles for 30.00 deg north latitude
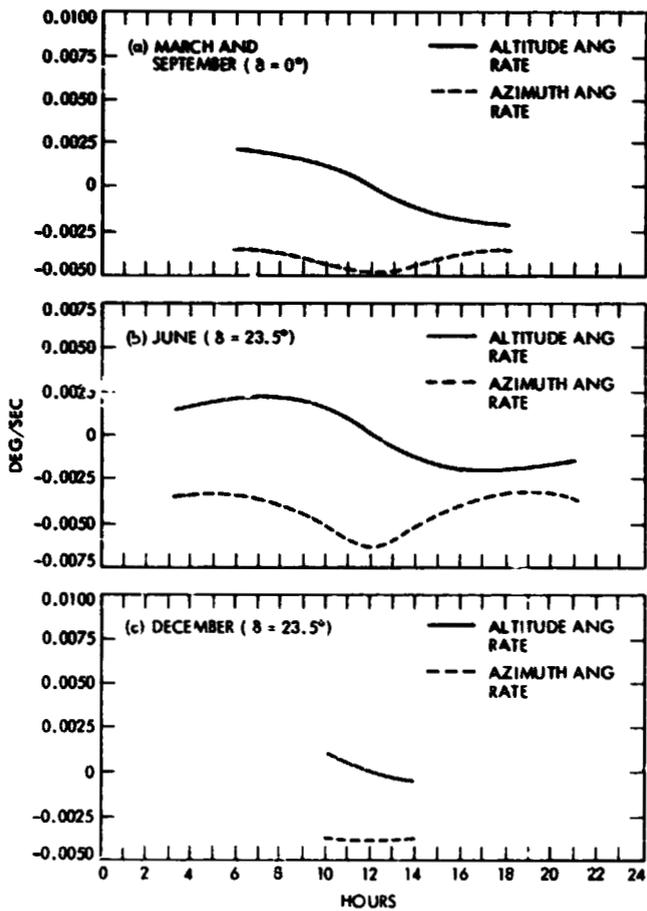
Fig. 8. Rates of solar angles for 60.00 deg north latitude
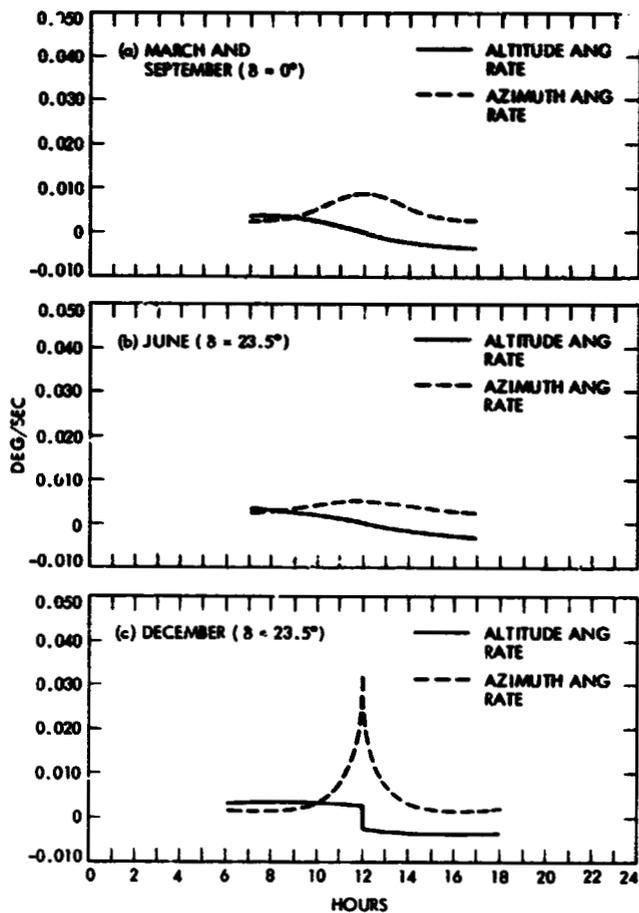
Fig. 9. Rates of solar angles for 30.00 deg south latitude
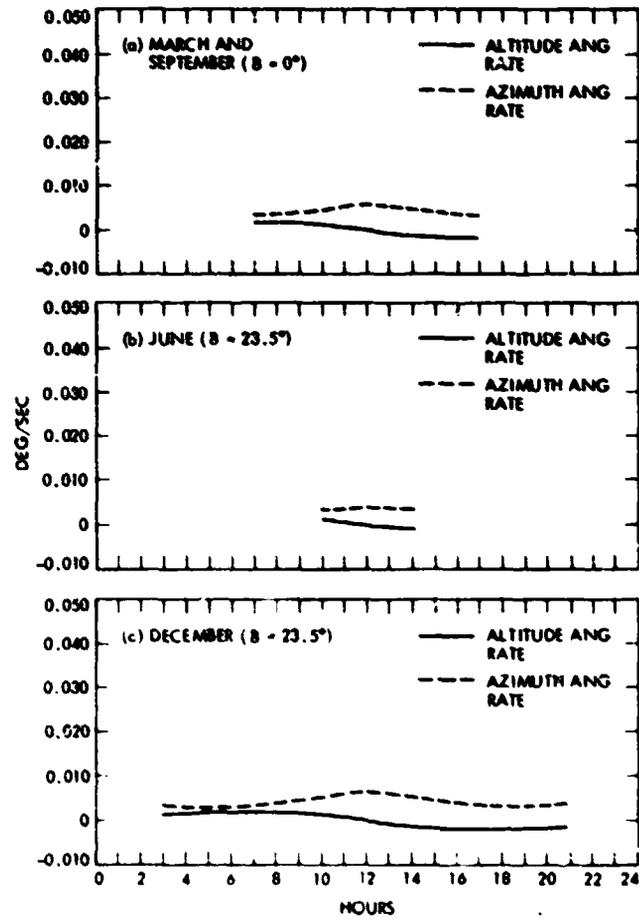
Fig. 10. Rates of solar angles for 60.00 deg south latitude