

COMSAT
Technical Review

Volume 7 Number 2, Fall 1977

COMSAT TECHNICAL REVIEW

Volume 7 Number 2, Fall 1977

Advisory Board Joseph V. Charyk
William W. Hagerty
John V. Harrington
Sidney Metzger

Editorial Board Pier L. Bargellini, Chairman
Robert D. Briskman
S. J. Campanella
William L. Cook
Jorge C. Fuenzalida
R. W. Kreutel
Norman S. Lomas
Akos G. Revesz
George R. Welti

Editorial Staff Stephen D. Smoke
MANAGING EDITOR
Margaret B. Jacocks
TECHNICAL EDITOR
Edgar Bolen
PRODUCTION
Michael K. Glasby
CIRCULATION

COMSAT TECHNICAL REVIEW is published twice a year by Communications Satellite Corporation (COMSAT). Subscriptions, which include the two issues published within a calendar year, are: one year, \$7 U.S.; two years, \$12; three years, \$15; single copies, \$5; article reprints, \$1. Make checks payable to COMSAT and address to Treasurer's Office, Communications Satellite Corporation, 950 L'Enfant Plaza, S.W., Washington, D.C. 20024, U.S.A.

© COMMUNICATIONS SATELLITE CORPORATION 1977

- 351 MARISAT—A MARITIME SATELLITE COMMUNICATIONS SYSTEM
**D. W. Lipke, D. W. Swearingen, J. F. Parker, E. E. Steinbrecher,
T. O. Calvit AND H. Dodel**
- 393 ECHO CANCELLER WITH ADAPTIVE TRANSVERSAL FILTER UTILIZING
PSEUDO-LOGARITHMIC CODING **O. A. Horna**
- 429 DPCM PREDICTION FOR NTSC COMPOSITE SIGNALS **I. Dinstein**
- 447 SIMULATION ANALYSIS FOR DIFFERENTIALLY COHERENT QUATER-
NARY PSK REGENERATIVE REPEATER **Y. S. Lee**
- 475 INCLUDING SYNCHRONIZATION IN TIME-DOMAIN CHANNEL SIMU-
LATIONS **L. C. Palmer AND S. Lebowitz**
- 527 SINGLY TERMINATED PSEUDO-ELLIPTIC FUNCTION FILTER **M. H.
Chen**
- 543 PHASE CORRECTION OF IMPATT AMPLIFIERS **W. J. Getsinger
AND H. L. Hung**
- 555 HIGH-PERFORMANCE FREQUENCY DOUBLERS FOR THE COMSTAR
BEACON **R. Mott**
- 577 RELIABLE IMPATT DIODES FOR THE COMSTAR CENTIMETER
WAVE BEACON **P. Koskos, P. Fleming, AND J. Reynolds**
- 599 PRELIMINARY RESULTS OF COMSTAR 19/29-GHZ BEACON MEA-
SUREMENTS AT CLARKSBURG, MARYLAND **J. M. Harris AND
G. Hyde**
- 625 CTR NOTES
CHARACTERISTICS OF PCM TELEPHONE VOICE CHANNELS **A. G.
Gatfield 625**
ORBITAL PERFORMANCE OF NTS-2 NICKEL-HYDROGEN BATTERY
J. D. Dunlop AND J. F. Stockel 639
- 649 TRANSLATIONS OF ABSTRACTS
FRENCH 649
SPANISH 655
- 661 ERRATUM

MARISAT—A maritime satellite communications system

D. W. LIPKE, D. W. SWEARINGEN, J. F. PARKER,
E. E. STEINBRECHER, T. O. CALVIT, AND H. DODEL

(Manuscript received May 12, 1977)

Abstract

This paper describes the communications techniques and facilities used in the MARISAT system to provide services to commercial users. The commercial portion of the system has been designed to meet performance requirements for high-quality voice and data communications between shore-based subscribers and sea-going vessels as well as other maritime installations.

The commercial system configuration is described, including the C-band shore stations, the C- and L-band repeaters in the satellites, and the L-band ship terminals which operate automatically. Modulation, access, and signaling techniques are discussed to explain how network control, centralized at the shore stations, provides rapid and automatic circuit connections. Signal characteristics, frequency plans, link budgets, and system communications capacity are also presented.

Introduction

In 1976 almost a decade of study and design dedicated to the development of a satellite system to serve mobile maritime users culminated with the implementation of the MARISAT satellite system. Following the MARISAT launches on February 19 and June 9, 1976, which placed satellites in geostationary orbit over the Atlantic and Pacific Oceans, respectively,

full commercial service was initiated on August 15, 1976. A third MARISAT spacecraft, launched on October 14, 1976, is located over the Indian Ocean. This system provides signal quality and channel availability not possible with HF and MF maritime facilities.

The MARISAT system represents a unique departure from the international and domestic satellites used for fixed point to fixed point communications. The design of the MARISAT satellite system reflects two important differences. First, separate repeaters share the prime power of each satellite so that service can be leased to the U.S. Navy at UHF and commercial maritime service can be provided at frequencies in the C- and L-bands. Secondly, in commercial service, the MARISAT system connects shore stations with small, inexpensive shipboard terminals which may continually move within the coverage area of the satellite.

With a large number of operational ship terminals, one or more shore stations assign channels and provide interconnection with the domestic and international telecommunications networks. Operating procedures and computer controlled switching techniques have been developed to perform those functions on an automatic basis as much as possible. This paper describes in detail that part of the MARISAT system which provides commercial services.

System description

The MARISAT communications system has been configured to meet the following objectives:

- a.* high-quality, full-duplex, reliable, real-time voice, data, and teleprinter services;
- b.* interconnection with existing international and domestic telecommunications networks;
- c.* safety and distress services;
- d.* growth in commercial capabilities as user requirements increase;
- e.* operation with multiple shore stations;
- f.* introduction of new services.

To meet these objectives, the MARISAT system incorporates the following features:

- a.* demand-assigned satellite communications capacity;
- b.* single-channel-per-carrier (SCPC) voice grade service;
- c.* a telegraphy transmission technique which combats Doppler shifts, frequency instabilities, and spurious FM noise;

- d.* an automatic, power-efficient, system access/control plan which allows rapid access between a large number of mobile ship terminals and several shore stations;
- e.* network control centralized at shore stations;
- f.* minimization of ship terminal equipment complexity and costs;
- g.* automatic ship terminal operation;
- h.* a carrier frequency plan which minimizes performance degradation due to intermodulation products;
- i.* the application of C.C.I.R. and C.C.I.T.T. recommendations as practicable.

The MARISAT system presently consists of the facilities shown in Figures 1 and 2. Three operational geostationary satellites are located at 15°W, 176.5°E, and 73°E longitude. Although the Indian Ocean satellite at 73°E longitude is primarily a backup for the other two satellites, it can be used for U.S. Navy (UHF) and commercial (L-band) services. Shore stations for commercial L-band services are located at Southbury, Connecticut, and Santa Paula, California. Tracking, telemetry, and command (TT&C) facilities are located at Southbury, Connecticut; Santa Paula, California; and Fucino, Italy; and the system control center is in Washington, D.C. The system also includes ship terminals approved for access. Major system design parameters for these facilities are shown in Table 1. The nominal service coverage (5° elevation angle) of the system is shown in Figure 3.

Each satellite contains three channels in the 250/300-MHz bands for use by the U.S. Navy: one wideband (500-kHz) channel and two narrowband (25-kHz) channels. Any or all of the three channels can be activated or deactivated by ground command. These channels are used by the U.S. Navy through its own ship and shore terminal facilities and are separate from the commercial communications system. Each satellite has two wideband repeaters for L-band service to commercial users [1]. One repeater translates shore-to-ship signals from 6 GHz to 1.5 GHz, while the other translates ship-to-shore signals from 1.6 GHz to 4 GHz.

Sharing satellite power between the Navy and the commercial users is made possible by a unique, 3-level, L-band traveling wave tube amplifier (TWTA). The L-band power level is selected in accordance with the satellite primary power beyond that required for the Navy UHF channels in actual use. Post-launch tests have verified that in-orbit primary power exceeds design specification values; therefore, the L-band tube can operate in the medium-power mode while all UHF channels are activated. With all UHF

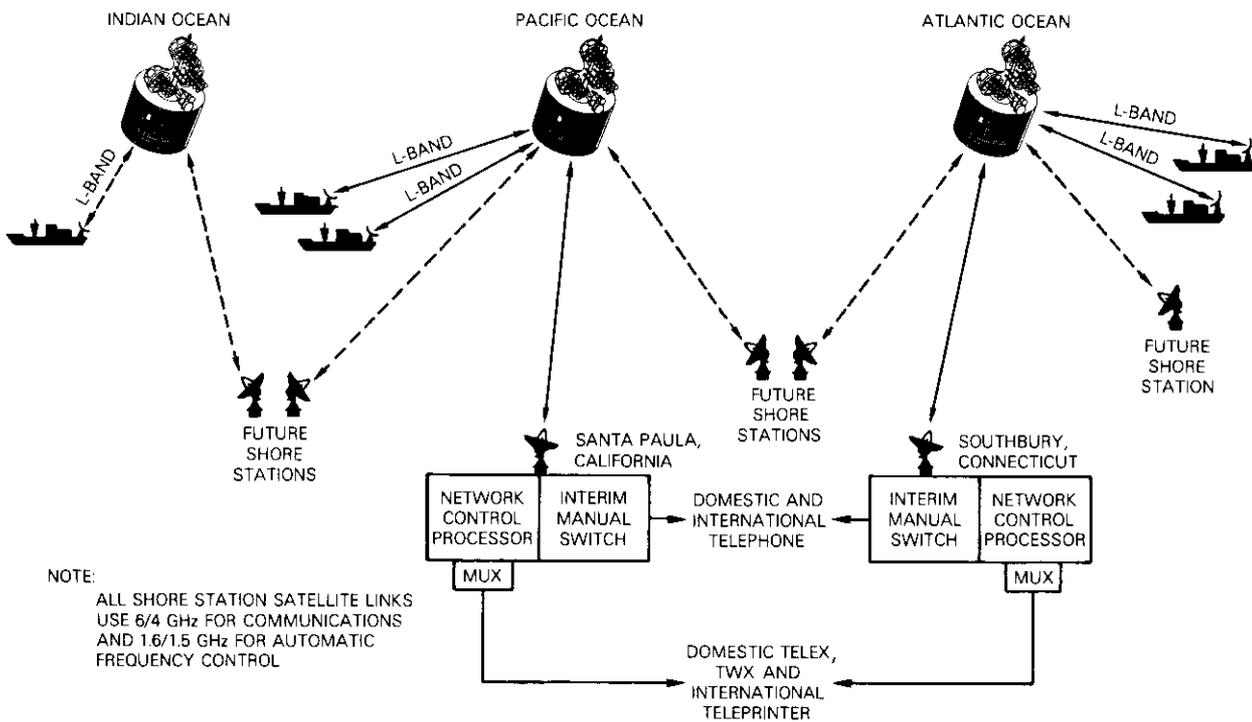


Figure 1. MARISAT Commercial Communications System

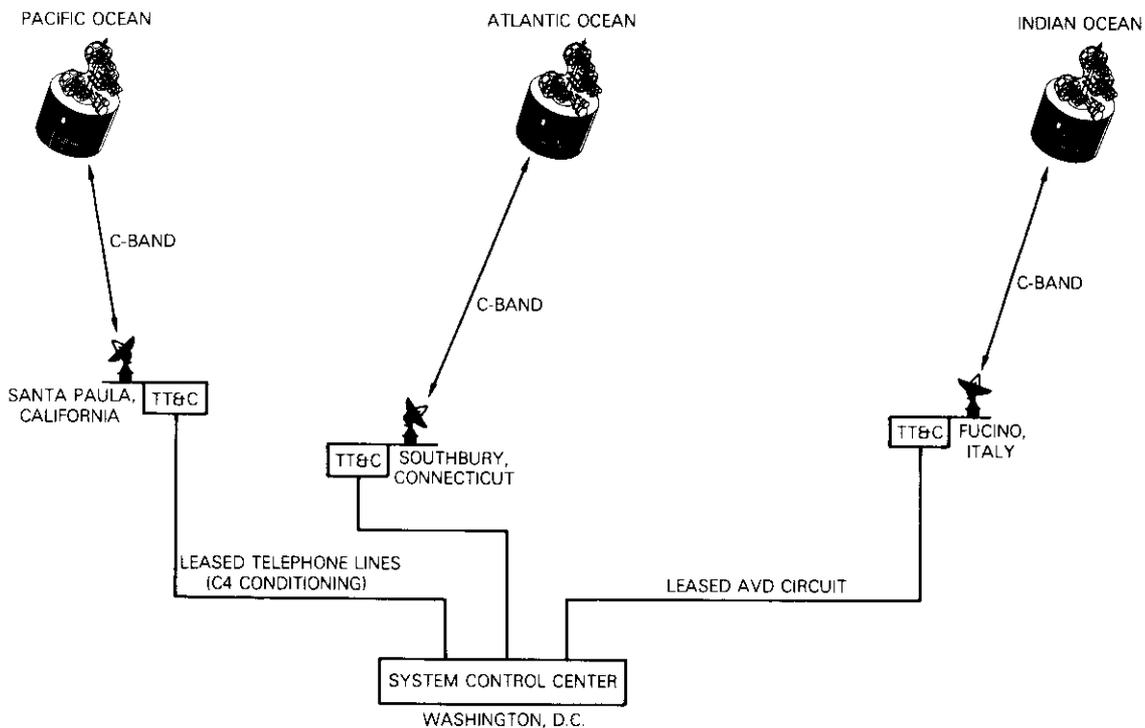


Figure 2. MARISAT TT&C System

TABLE 1. SYSTEM DESIGN PARAMETERS

Satellite Locations	$15^\circ \pm 0.5^\circ\text{W}$, $176.5^\circ \pm 0.5^\circ\text{E}$, $73^\circ \pm 0.5^\circ\text{E}$
Orbit Inclination	$\leq 3^\circ$
Frequency Bands	
Shore/Satellite (C-band)	
Up-Link	6.420–6.424 GHz
Down-Link	4.195–4.199 GHz
Ship/Satellite (L-band)	
Up-Link	1.6385–1.6425 GHz
Down-Link	1.537–1.541 GHz
Satellite G/T (beam edge)	
C-band	–19.5 dB/K
L-band	–17 dB/K
Satellite e.i.r.p. (beam edge)	
C-Band	18.8 dBW (saturated output)
L-Band	
Low-Power Mode	20.0 dBW
Medium-Power Mode	26.0 dBW
High-Power Mode	29.5 dBW
Ship Terminal G/T (L-band)	≥ -4.0 dB/K
Ship Terminal e.i.r.p. (L-band)	37 ± 1 dBW (all transmissions)
Shore Station G/T (C-band)	32.2 dB/K
Shore Station e.i.r.p. (C-band)	
Low-Power Mode	72 dBW voice, 69 dBW TDM
Medium-Power Mode	66 dBW voice, 63 dBW TDM
High-Power Mode	62 dBW voice, 59 dBW TDM

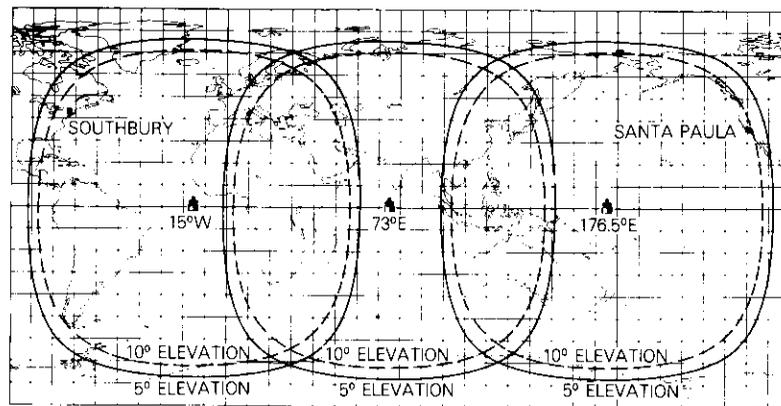


Figure 3. MARISAT Coverage

channels deactivated, the L-band transmitter can operate at full power. The satellite C-band transmitter (satellite-to-shore) always operates at full power.

Each shore station is equipped with a 12.8-meter (42-ft) parabolic antenna which transmits to and receives from the MARISAT satellites in the 6- and 4-GHz bands, respectively. Each antenna has a full backup facility shared with other programs. The shore station antenna also transmits and receives L-band signals to monitor transmissions, to compensate for frequency translation errors, and to check satellite performance. Other facilities provide the TT&C signal conditioning and formatting required to connect the station to the System Control Center in Washington, D.C. Total equipment redundancy is provided at each shore station.

Commercial teleprinter and telephony traffic between shore subscribers and ships is routed through network control equipment at each shore station to automatically establish connections through the satellite. The two U.S. shore stations interconnect with U.S. domestic and international networks to provide the following types of end-to-end user service:

a. Teleprinter. The teleprinter at the ship terminal resembles any other machine in either the U.S. domestic or international telex network. Call setup and clear down procedures as well as signaling between ship terminals and gateway switch facilities are handled automatically at the shore stations.

b. Telephone. Voice services between terrestrial subscribers and ship terminal operators are established with the assistance of operators at the shore stations. However, the system design incorporates features to permit automatic ship-to-shore voice communications.

c. Medium-Speed Data. The voice channel can transmit data rates up to 2,400 bit/s for interconnection with the public network. Rates higher than 2,400 bit/s are possible with appropriate modulation methods, signal parameters, and operational procedures selected to suit specific applications.

d. Facsimile. Both analog and digital facsimile can be transmitted using the voice channel.

In addition to the routine communications services, the MARISAT system incorporates unique features which permit the following capabilities:

- a.* ship-to-ship calls via a double satellite hop;
- b.* emergency calls handled on a priority basis;

c. broadcast calls to all ships in a network, to ships of a particular national grouping or fleet, or to ships in a specific ocean area.

The MARISAT space segment and certain portions of the shore station facilities are owned by the MARISAT Joint Venture, which consists of four communications companies: COMSAT GENERAL Corporation (86.29 percent), RCA Global Communications, Inc. (8 percent), Western Union International, Inc. (3.41 percent), and ITT World Communications Inc. (2.3 percent). COMSAT GENERAL Corporation serves as manager of the system. Ship terminals, which are not part of the jointly owned facilities, are granted access to the system if they comply with the technical requirements of the MARISAT Joint Venture.

Communications system

The MARISAT communications system [2] provides voice grade channels, telegraph channels, and signaling channels. The voice grade channels are allocated on a single-channel-per-carrier basis using frequency modulation (FM). The telegraph channels are time-division multiplexed (TDM) in the shore-to-ship direction and time-division multiple-accessed (TDMA) in the ship-to-shore direction; in each case, 22 telegraph channels share the same carrier. Channel assignment and signaling information is integrated into the telegraph TDM carrier. For ship-to-shore signaling, the mobile terminals share a single frequency on an as-needed basis for channel requests. Service requests from the mobile terminals to the shore stations are in the form of short bursts on this "request channel" frequency.

The frequencies for voice and telegraphy in the shore-to-ship and ship-to-shore directions are paired and have a frequency separation of 101.5 MHz at L-band. When a voice grade call is set up, a pair of frequencies is selected by the shore station and assigned to the particular mobile terminal. After the voice call is completed, the frequency pair is released and made available for use in another call.

The use of frequency pairs for telegraph calls is different than that for voice calls. The shore-to-ship TDM carrier is modulated by a bit stream of 1,200 bit/s using phase shift keying (PSK). The ship-to-shore TDMA frequency is occupied by gated bursts transmitted in a predetermined sequence by up to 22 different mobile terminals. Each burst is PSK modulated at 4,800 bit/s and contains up to 12 characters of message text. The time reference for the mobile terminal transmissions is derived from the continuously received TDM signal. Therefore, synchronization among

terminals is not required. Sufficient time is allocated between bursts to accommodate propagation delay differences between terminals at the center and edge of the satellite coverage area. Voice grade circuits are used for conversational speech, facsimile, and data transmission at rates up to 2,400 bit/s. Channel transmission characteristics are shown in Table 2.

TABLE 2. CHANNEL TRANSMISSION CHARACTERISTICS

Telegraphy	
Bit-Error Rate ^a	$<1 \times 10^{-6}$
Character Error Rate ^a	No errors for at least 95% of all calls; $\leq 6 \times 10^{-5}$ for any single call
Synchronous Margin on Incoming Signal ^b	$\geq 40\%$
Minimum Acceptable Character Length ^b	145 5/6 ms
Synchronous Start-Stop Distortion ^c	$< 5\%$
Modulation Rate ^c	50.0 ± 0.25 baud
Minimum Stop Pulse Length ^c	25 ms
Voice ^a	
Speech-to-Weighted Noise ^d	
Ship-to-Shore Link	≥ 30 dB
Shore-to-Ship (unfaded) Link	≥ 25 dB
Data/Facsimile ^a (voice channel without companders)	
Data Tone-to-Unweighted Noise ^e	
Ship-to-Shore	> 24 dB
Shore-to-Ship	> 22 dB

^a With satellite at elevation angle $\geq 5^\circ$.

^b Incoming to shore station interface from the terrestrial network.

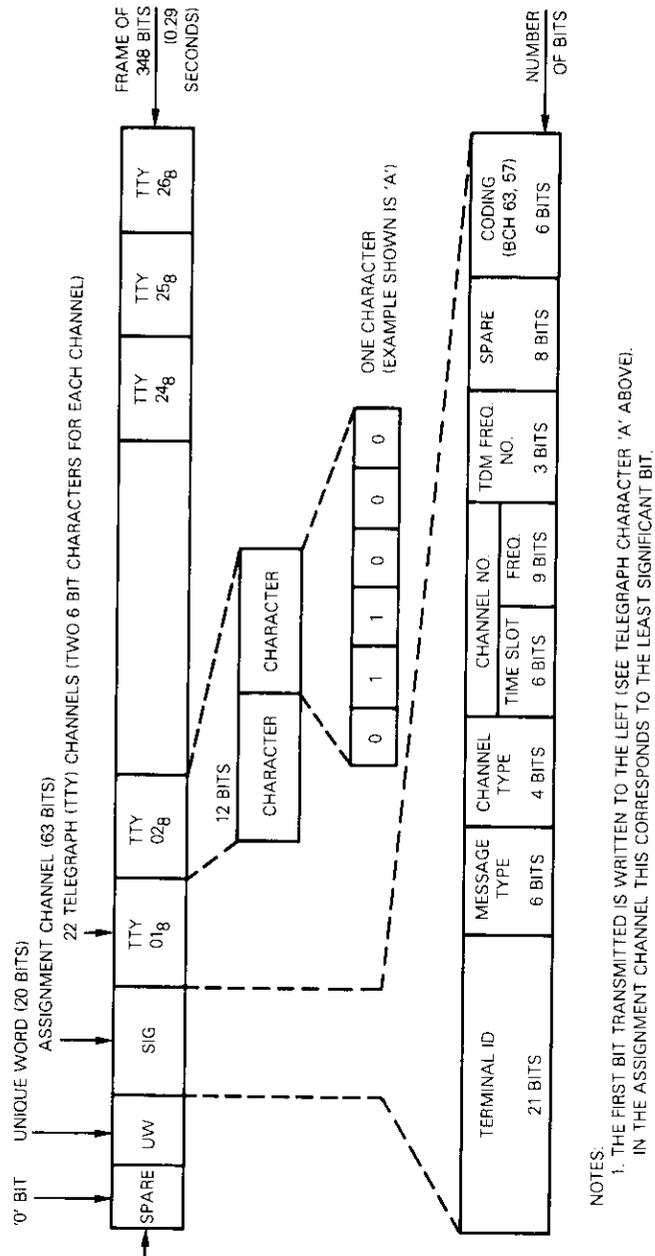
^c Outbound from shore station interface towards the terrestrial network.

^d Subjective equivalent ratio with median talker.

^e Data tone 6 dB below full deviation.

Signal characteristics

Each shore station in the MARISAT system transmits at least one dedicated TDM carrier. The TDM frame format is shown in Figure 4, and the signal characteristics are summarized in Table 3. Each shore-to-ship TDM carrier has a paired 22-channel TDMA signal in the ship-to-shore direction which provides full-duplex teleprinter operation. Each active mobile terminal transmits a single burst per frame. Burst synchronization is derived from the continuously received TDM signal which has a frame duration which is exactly one-sixth that of the TDM frame (see Table 3); the complementary unique word provides the synchronization gate pulse. The



NOTES:

1. THE FIRST BIT TRANSMITTED IS WRITTEN TO THE LEFT (SEE TELEGRAPH CHARACTER 'A' ABOVE). IN THE ASSIGNMENT CHANNEL THIS CORRESPONDS TO THE LEAST SIGNIFICANT BIT.
2. IN THE TELEGRAPH CHANNEL THE FIRST BIT TRANSMITTED INDICATES THE TYPE OF CHARACTER FIELD. WHEN THE FIRST BIT IS A ZERO, THE SUBSEQUENT 5-BIT CHARACTER FIELD REPRESENTS A C.C.I.T.T. NO. 2 ALPHABET CHARACTER. WHEN THE FIRST BIT IS A ONE THE SUBSEQUENT 5-BITS REPRESENT LINE CONDITIONS FOR SIGNALING.

Figure 4. TDM Channel Format (shore-to-terminal)

TABLE 3. TDM SIGNAL CHARACTERISTICS

Modulation	2-phase coherent PSK (CPSK), unfiltered
Ambiguity	Differential coding; phase transition of 180° on a binary "one"
Data Rate	1,200 bit/s; ±1 part in 10 ⁴ [3]
Frame Length	0.29 s (348 bits)
Synchronization	20-bit unique word; complementary unique word transmitted every sixth frame
Unique Word	0111 1010 1100 1101 0000
Complementary Unique Word*	1000 0101 0011 0010 1111
Coding	57 bits plus 6 parity bits encoded with BCH (63, 57) code; the generating polynomial is (6, 1, 0)
Telegraph Channel	12 bits per frame (two 6-bit characters)
Character Code	C.C.I.T.T. Code #2 plus line state characters, mark = 111111 and space = 100000

* The first bit is the left bit in the sequence.

TDMA frame format is shown in Figure 5, and the signal characteristics are summarized in Table 4.

TABLE 4. TDMA SIGNAL CHARACTERISTICS

Modulation	2-phase CPSK, filtered
Spectrum Roll-off	At least 40 dB below the unmodulated carrier level in a 6.3-kHz band 50 kHz on either side of the carrier center frequency for all possible bit patterns.
Ambiguity	Differential coding; phase transition of 180° occurs on a binary "one." Input data in NRZ format.
Data Rate	4,800 bit/s; ±1 part in 10 ⁴ [3]
Frame Length	1.74 s
Burst Duration	37.7 ms (181 bits)
Preamble*	109 bits
Carrier Recovery Field	50 bits, all "0's"
Bit Timing Field	29 bits, all "1's"
Unique Word	30 bits
	0000 1000 0101 0011 0101 1001
	1111 11
Telegraph Channel	72 bits per burst, 12 six-bit characters
Character Code	C.C.I.T.T. Code #2, plus line state characters, mark = 111111 and space = 100000

* The first bit is the left bit in the sequence.

TABLE 5. VOICE CHANNEL CHARACTERISTICS

Modulation	Frequency modulation (FM), single channel per carrier (SCPC)
Peak Frequency Deviation	12 kHz
rms Speech Deviation	3.8 kHz for average call
rms Data or Facsimile Tone Deviation	Nominally 4.2 kHz
Baseband	300-3,000 Hz
Baseband Reference Level	0 dBr is defined as that level causing 12-kHz peak deviation (8.5-kHz rms deviation for sinusoidal test tone)
Companding	2:1 syllabic as specified in C.C.I.T.T. Rec. G. 162 [4], with unaffected level equal to 0 dBr
Emphasis	None
Peak Clipping Level	0 dBr

TABLE 6. SHORE-TO-SHIP LINK PARAMETERS (BEAM EDGE), MEDIUM-POWER MODE (26 dBW)*

Parameter	Voice	TDM
Shore-to-Satellite (6 GHz)		
Shore Station Transmitter Power (dBW)	11.9	9.5
Combining Network and Feeder Loss (dB)	2.1	2.1
Shore Station Antenna Gain (dBi)	56.2	56.2
Shore Station e.i.r.p. (dBW)	66.0	63.6
Free Space Attenuation (dB)	200.8	200.8
Absorption with Precipitation 10° Elevation (dB)	0.7	0.7
Polarization Coupling Loss (dB)	0.4	0.4
Satellite G/T (dB/K)	-19.5	-19.5
Saturation Flux Density (dBW/m ²)	-87.0	-87.0
Up-Link C/N ₀ (dB-Hz)	73.2	70.8
Satellite-to-Ship (1.5 GHz)		
Intermodulation, Noise and Pilot Losses (dB)	1.4	1.4
Small Carrier Suppression (dB)	0.0	0.6
Satellite e.i.r.p. (dBW)	15.6	12.6
Free Space Loss (dB)	188.5	188.5
Polarization Coupling Loss (dB)	0.4	0.4
Absorption, 5° Elevation (dB)	0.4	0.4
Short-Term Fading (dB)	4.0	4.0
Additional Wet Radome Degradation (dB)	0.5	0.5
Ship Terminal G/T (including losses due to dry radome and pointing error) (dB/K)	-4.0	-4.0

(Continued)

Parameter	Voice	TDM
Down-Link C/N ₀ with Short-Term Fade (dB-Hz)	46.4	43.4
Down-Link C/N ₀ without Short-Term Fade (dB-Hz)	50.4	47.4
C/N ₀ Intermodulation (dB-Hz)	>70.0	>67.0
Total Link C/N ₀ without Short-Term Fade (dB-Hz)	50.4	47.4
Total Link C/N ₀ without Fade (dB-Hz)	51.7	48.7

* In the high-power mode intermodulation noise will be greater. Also, the shore-to-ship satellite repeater gain will increase, so that the up-link power per carrier will decrease.

TABLE 7. SHIP-TO-SHORE LINK PARAMETERS (BEAM EDGE)

Parameter	Voice or TDMA/Request
Ship-to-Satellite (1.6 GHz)	
Ship Transmitter Power (dBW)	14.8
Diplexer Loss (dB)	0.6
Ship Antenna Gain (dBi)	23.0
Dry Radome Loss (dB)	0.2
Ship e.i.r.p. (dBW)	37.0
Free Space Attenuation (dB)	189.1
Additional Wet Radome Loss (dB)	0.3
Absorption for 5° Elevation (dB)	0.4
Polarization Coupling Loss (dB)	0.4
Short-Term Fading (dB)	4.0
Satellite G/T (dB/K)	-17.0
Saturation Flux Density (dBW/m ²)	-99.5
Up-Link C/N ₀ with Short-Term Fade (dB-Hz)	54.4
Up-Link C/N ₀ without Short-Term Fade (dB-Hz)	58.4
Satellite-to-Shore (4 GHz)	
Gain Compression and Intermodulation Loss, Worst-Case (dB)	1.8
Satellite e.i.r.p. without L-Band Short-Term Fade (dBW)	-4.6
Free Space Loss (dB)	197.1
Absorption, 10° Elevation (dB)	0.2
Polarization Coupling Loss (dB)	0.4
Excess Attenuation with Precipitation (dB)	0.5
Excess Sky Noise Degradation with Precipitation (dB)	1.2
Shore Terminal G/T, (clear sky, 10° elevation) (dB/K)	32.2
Down-Link C/N ₀ without Short-Term L-Band Fade (dB-Hz)	56.8
Down-Link C/N ₀ with Short-Term L-Band Fade (dB-Hz)	52.8
Worst-Case C/N ₀ Intermodulation (dB-Hz)	>62.0
Total Link C/N ₀ without Short-Term L-Band Fade (dB-Hz)	53.8
Total Link C/N ₀ with Short-Term L-Band Fade (dB-Hz)	49.8
Total Link C/N ₀ without Fade (dB-Hz)	56.0

Satellite parameters are worst-case specification values for worst-case geometries. For other ship locations, the corresponding values of carrier-to-noise density could be as much as 6 dB higher considering the satellite antenna gain pattern, clear weather conditions, and range differences.

The chosen link parameters provide a bit-error rate (BER) for the TDM signal of no more than 1×10^{-5} even with down-link fading up to 5.3 dB. A subjective evaluation of typical FM voice modem equipment [5] shows that the transmission objectives in Table 2 are achieved on each link during unfaded operation. Under practical operating conditions, voice and telex services can be expected to meet or exceed these transmission objectives.

Communications capacity

The commercial communications capacity in the shore-to-ship direction of each MARISAT satellite depends upon the operating level mode of the satellite L-band transmitter. In the low-power mode, each satellite supports the design capacity of one voice and 44 telex channels. Since the L-band e.i.r.p. measured after launch for each satellite exceeds the specifications by 25 percent, the medium-power mode capacity exceeds expectations and amounts to eight voice channels and 44 telex channels (2 TDM carriers). In the high-power L-band mode, the space segment capacity is approximately twice that in the medium-power mode. Voice and teleprinter capacity can be traded within the constraints of the MARISAT frequency plan. In the shore-to-ship direction, the satellite L-band power required for a voice channel is equivalent to that used for 44 telex channels. In the ship-to-shore direction, the space segment capacity is independent of the use of the UHF system, and each satellite is capable of matching the shore-to-ship capacity in the high-power mode.

Frequency plan

The radio frequency plan for the MARISAT system minimizes performance degradation due to intermodulation distortion, especially on the power-limited satellite-to-ship link. Using 25 kHz as the basic spacing increment, the frequency plan provides 13 channels without third-order intermodulation products. The number of possible channel assignments is sufficient to accommodate the medium-power mode capacity. Additional channel assignments for operation in the L-band high-power mode will be chosen to minimize additional interfering intermodulation products.

With the exception of the beacon and the request channel, the return link (ship-to-shore) frequency configuration is paired with that of the forward link. Corresponding C-band (shore/satellite) frequencies are

separated by 2,225.00 MHz, while L-band (satellite/ship) frequency pairs are 101.5 MHz apart. A typical MARISAT frequency plan [2] is shown in Figure 7.

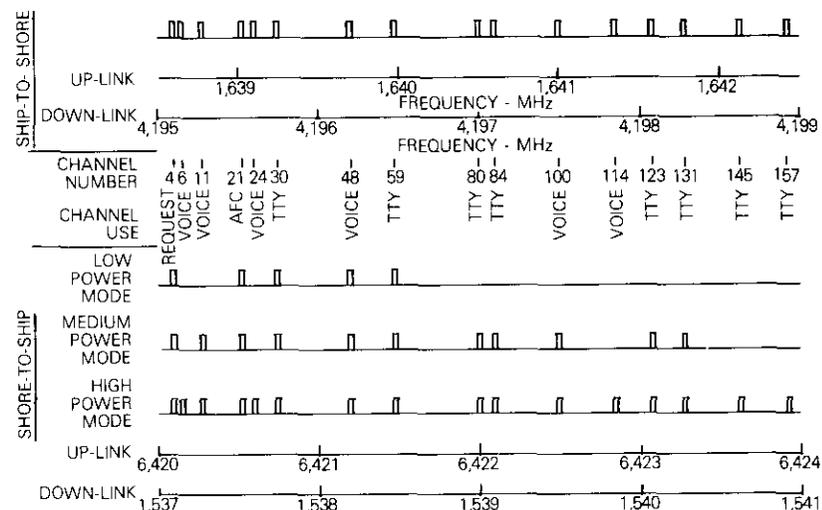


Figure 7. Typical MARISAT Frequency Plan

Channel assignment and signaling

Channel assignment and network control are provided by the shore stations. A shipboard calling party can request access to a working channel via the random-access request channel, and shoreside calling parties can request access to a particular ship terminal via a shore station. An assignment message in the TDM carrier announces the call to the ship for both terminal- and shore-originated calls. When the ship terminal is not busy, it automatically responds to an assignment message addressed to it. When it is busy, it responds to clearing messages but not to routine call announcement messages. Once a satellite circuit has been established between terminal and shore, selection and supervisory signaling are performed within the working channel. Except for the message request, the terminal transmitter can be activated only by an assignment message command from the shore station.

The combination of assignment channel signaling and working channel signaling in the MARISAT system has enabled the extension of terrestrial

telegraph and telephone networks to maritime mobile terminals. Mobile terminal complexity has been minimized by relegating many of the signaling and switching functions to the shore stations and gateways, thus providing redundancy in the critical clearing functions. Minimizing the normal uses of the assignment channel for signaling permits a large amount of working channel traffic to be supported without significant delays. Signaling for additional services can be gradually introduced utilizing the basic assignment channel signaling capabilities and additional working channel protocols.

Request channel

Each ship terminal formats the request message burst as shown in Figure 6. The assigned terminal ID code is fixed and automatically inserted into each request burst. Other request message data, selectable or inserted automatically, are as follows:

- a. shore station (e.g., Southbury),
- b. priority (routine, safety, urgent, or SOS),
- c. ocean area (up to 15 designators possible),
- d. type of channels (e.g., duplex voice),
- e. terrestrial network (e.g., COMSAT GENERAL),
- f. nature of request (assignment or release).

The 39-bit request message field is automatically encoded with a BCH code by the ship terminal to aid in error detection at the shore station. Although all mobile terminals share a common "request" frequency, the probability of burst overlap from two or more terminals is small because the bursts are short (38 ms). If response is not received from the shore station within six seconds, a second request burst can be transmitted.

Assignment messages

The shore station responds to a channel request by transmitting either a delay message or a channel assignment, causing automatic tuning of the ship terminal. As shown in Figure 4, each 57-bit assignment message contains the following elements:

- a. ID of addressed terminal(s),
- b. message type (e.g., new channel assignment, "queue" message, "invalid request" message, or "clearing" message),
- c. channel type (e.g., telex or telephone),
- d. channel number (frequency and time slot),

When a new channel is assigned, a single assignment message contains all the information to establish the satellite portion of the link automatically.

Call assignments are addressed either to individual ships or to groups of ships. The distinction between individual addresses and broadcast addresses is made in the 7-character (21-bit) "SHIP ID" address field by reserving certain codes for broadcast addresses.

Tuning

All frequency pairs and time slots are selected automatically when the terminal decodes its addresses in the assignment channel. TDM/TDMA tuning is independent from voice channel tuning so that a TDM carrier may be received continuously for assignment and control. In the MARISAT network, all the ship terminal TDM receivers are tuned to 1,537.750 MHz when the terminal is idle unless commanded to change. After a telegraph call requiring TDM retuning is cleared, the terminal receiver automatically retunes to 1,537.750 MHz.

The paired TDMA transmit frequency is 1,639.250 MHz. The request channel frequency is 1,638.600 MHz and is the same for all terminals. Use of the request channel does not interrupt reception of the TDM carrier. For a voice grade call, each MARISAT terminal can tune automatically to any one of 159 frequency pairs in 25-kHz increments, starting at 1,537.025-MHz receive/1,638.525-MHz transmit and ending at 1,540.975-MHz receive/1,642.475-MHz transmit.

Working channel signaling

After a satellite channel has been assigned, normal call processing is performed in the assigned working channel. Protocols for telegraph and telephone calls are consistent with C.C.I.T.T. recommendations [6]-[8] and allow efficient interworking with terrestrial networks. The protocols are restricted to essential functions such as the following:

- a. verification of MARISAT channel availability,
- b. transfer of selection information,
- c. answer supervision,
- d. clearing supervision.

After a connection has been extended to a transit switching center, the working channel is transparent to the shore station except for supervisory clearing signals. Hence MARISAT users can develop additional signaling protocols for special services such as store and forward message switching or PBX extension signaling.

At the U.S. shore stations, the terrestrial trunks, which are equipped for C.C.I.T.T. type B signals with keyboard selection, connect the shore station to international transit centers (gateways). The signal timing and mark/space (M/s) status for a shore-originated telex call transiting a U.S. shore station is shown in Figure 8. This signaling is characterized by 2-way continuity verification, isolation of the satellite link from the gateway link until both links have been verified, and prompt response to gate seizure and answerback inquiry.

For a ship-originated call, the mobile terminal transmits keyboard selection signals after establishing a satellite circuit and exchanging answerbacks. During a call, the shore station monitors the line for clearing either by the mobile terminal or the shoreside party. The satellite link is designed to preclude premature clearing caused by short-term interruptions in the received radio signal. However, loss of the mobile terminal's radio signal for more than a minute will cause the shore station to clear the channel.

For telephone channel signaling, the MARISAT shore station operates as an additional transit switch between a terrestrial transit center and the mobile terminals. In addition to performing a splitting function (between calling and called parties) during the call setup, the shore station provides concentration and routing for the terrestrial trunks.

The satellite channels are equipped for single supervisory frequency (SF) signaling and multifrequency selection signaling. At the U.S. shore stations, the terrestrial trunks are equipped for C.C.I.T.T. #5 signaling. Signaling for a telephone channel is shown in Figures 9 and 10, which indicate the sequence for a ship-originated call. Characteristics of the satellite link signals include the following:

- a. compelled SF sequence for link verification;
- b. supervisory frequency (SF) = 2,600 Hz;
- c. SF recognition time = 125 ± 25 ms;
- d. multifrequency ship-to-shore per C.C.I.T.T. Recommendation Q23 (pushbutton telephone) terminated by " ";
- e. audible signals at 425 Hz from shore station to ship consistent with C.C.I.T.T. Recommendation Q35.

Calls are operator-assisted to obtain reliable billing data for shore-originated calls and to permit person-to-person calls from mobile terminals. After a through connection has been established, the shore station is transparent except for clearing signals initiated either from the shore end or the mobile terminal. As in the case of the telegraph calls, the

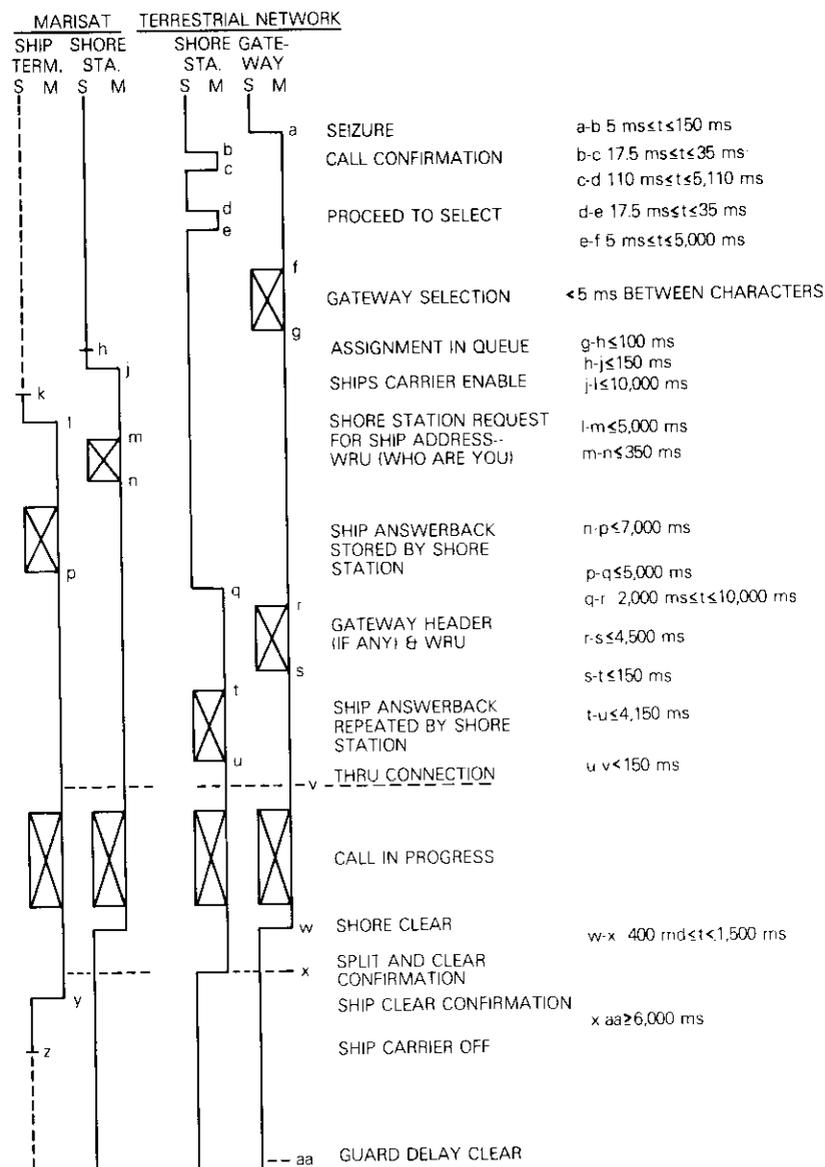
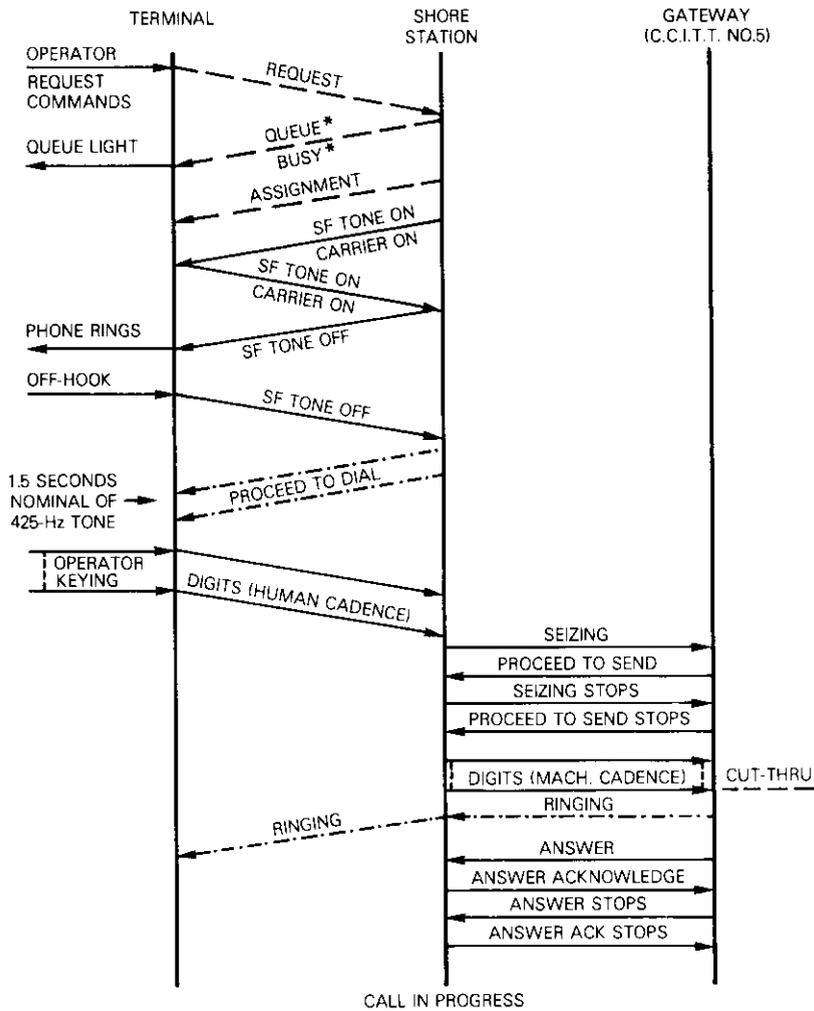


Figure 8. Shore-Originated Telex Call



* IF REQUIRED

- IN-BAND SIGNALING
- OUT-OF-BAND SIGNALING
- - - AUDIBLE TONE

NOTE: RINGING MAY NOT BE NEEDED IF THE HANDSET IS ALREADY OFF-HOOK. HOWEVER, THE SF SEQUENCE SHALL BE EXACTLY AS SHOWN.

Figure 9. Signaling Sequence, Voice Call Setup Automatic Mode (terminal-to-shore)

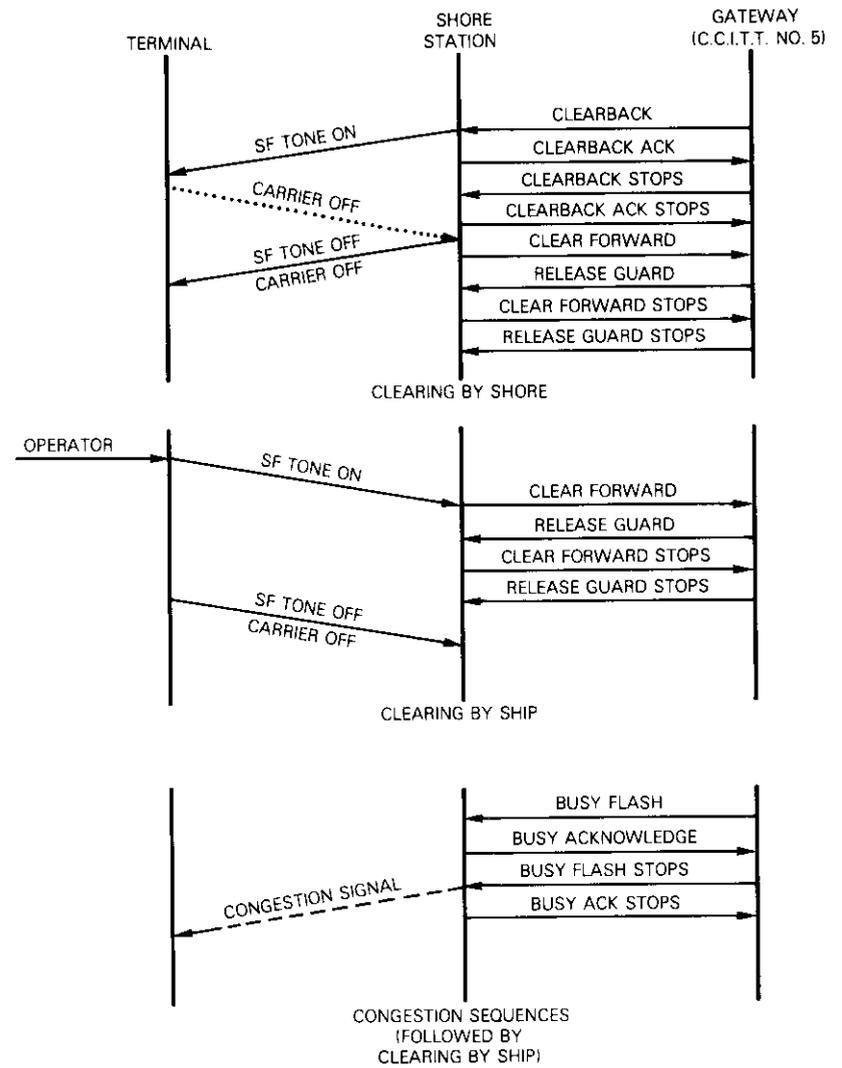


Figure 10. Signaling Sequence, Voice Call Clearing and Congestion, Automatic Mode

satellite link design precludes premature clearing caused by interruptions shorter than one minute in the received radio signals.

Satellite characteristics

The spin-stabilized MARISAT satellites [1], shown in Figure 11, operate in geostationary orbit. Each satellite had a lift-off weight of 654.6 kg (1,445 lb) and was launched from the Kennedy Space Center at Cape Canaveral, Florida, by a Thor-Delta 2914 vehicle. Subsequent to apogee motor burn-out, the on-orbit weight is about 326.2 kg (720 lb). The cylindrical body of the spacecraft with the attached solar cell array has an outside diameter of 2.16 m (85 in.), a height of 1.63 m (64 in.), and a dry weight of 283.1 kg (625 lb). The height of the spacecraft including the 20.4-kg (45-lb) despun antenna farm is 3.81 m (150 in.).

The MARISAT spacecraft carries four despun antenna systems, each providing earth coverage. The UHF array consists of three 1.22-m (48-in.) long helical elements at the vertices of a 1.22-m equilateral triangle. The L-band antenna, which is mounted in the center of the UHF array, consists of a 4-element helical array; each element is about 0.38 m (15 in.) long. This right-hand circularly polarized array provides a minimum gain of about 18 dBi at beam center and 14 dBi at edge of earth. Two separate C-band antennas, for transmissions between shore and satellite, are conventional circular horn antennas optimized for earth coverage. These horns, sized for 6 and 4 GHz (for receive and transmit, respectively), have nominal gains of 19 dBi at beam center and 15 dBi at edge of earth in their communications bands. The transmit antenna is left-hand circularly polarized; the receive antenna is right-hand circularly polarized.

The antenna farm is despun and earth oriented; the remainder of the spacecraft spins at 100 rpm. A noncontacting tricoaxial rotary joint couples the RF signals in the UHF- L-, and C-bands between the spinning and despun sections. All electronics and housekeeping subsystems are mounted within the spinning cylindrical body.

As shown in Figure 12, each satellite contains three separate, fully redundant repeaters. One all-solid-state repeater, operating at UHF (300/250 MHz), provides service to the U.S. Navy. The C- to L-band (C/L) and L- to C-band (L/C) repeaters provide full-duplex service to commercial shipping.

Ship transmissions are received at 1,638.5–1,642.5 MHz and undergo a single translation of 2,556.5 MHz to 4,195–4,199 MHz through the L/C repeater, which has a nominal 1-dB bandwidth of 4 MHz. With the exception of the 5-W C-band TWTA, solid-state devices are employed throughout the repeater. During normal on-station operation, telemetry beacon signals at 3,945.5 MHz and 3,954.5 MHz are injected into the TWT at a low input

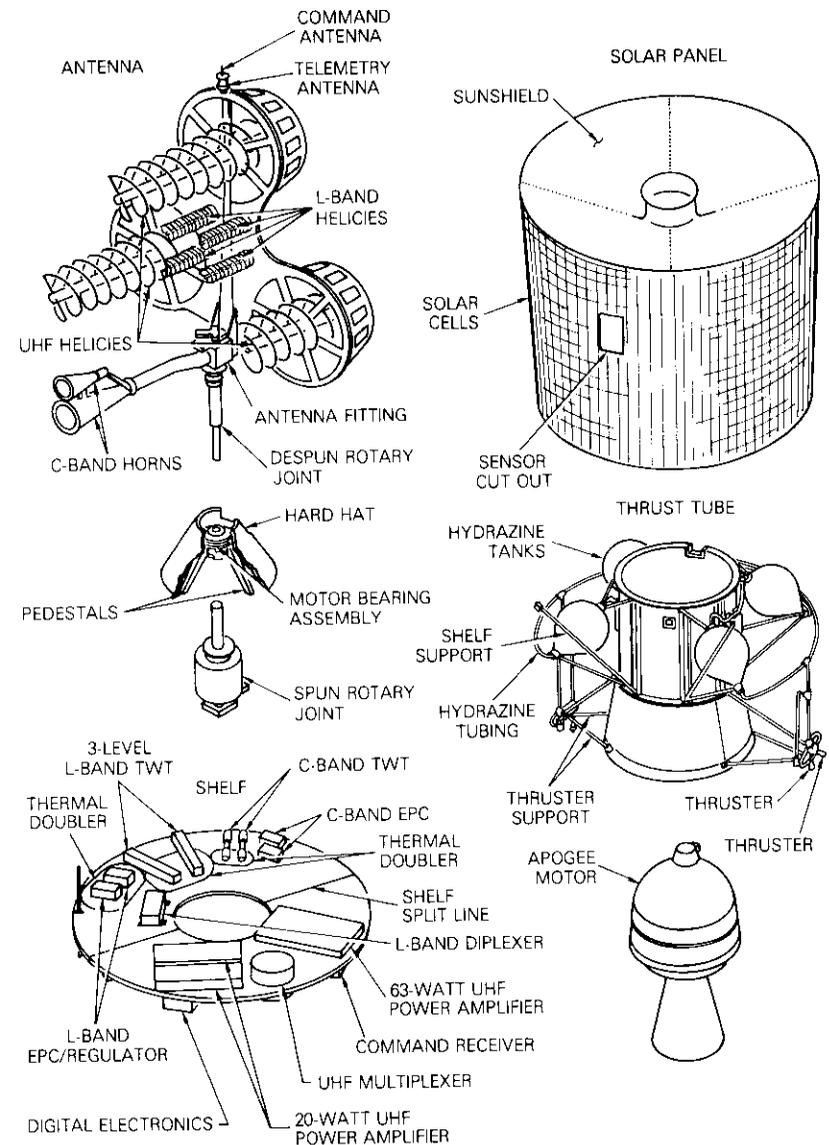


Figure 11. MARISAT Spacecraft, Exploded View

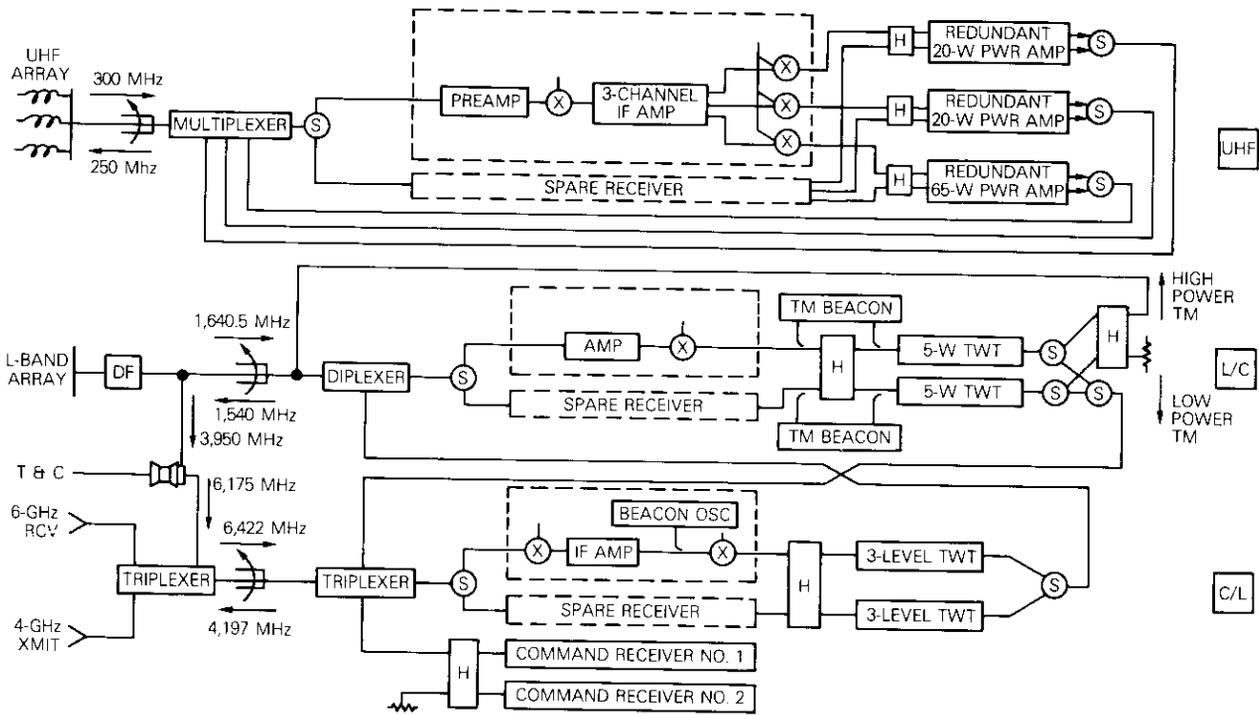


Figure 12. Satellite Communications Subsystem Block Diagram

TABLE 8. MARISAT SHORE-TO-SHIP REPFATER CHARACTERISTICS

Satellite Parameter	Specification Value	Test Data		
		F-1	F-2	F-3
Receive Band, -1 dB (MHz)	6,420-6,424			
Transmit Band, -1 dB (MHz)	1,537-1,541			
Repeater Bandwidth at 10-dB Output Backoff (MHz)	+2.0, -2.0	+3.45, -3.20	+3.15, -3.10	+4.00, -3.32
Antenna Beamwidth (deg)	19			
Axial Ratio (dB)				
Receive	<4	3.1	3.0	2.6
Transmit	<4	2.9	2.8	2.7
Saturation Flux Density (dBW/m ²)	-87 ± 2	-88.1	-88.0	-87.5
Saturation e.i.r.p. (dBW)				
Low	20	21.5	21.6	21.2
Medium	26	27.3	27.4	27.0
High	29.5	30.7	20.9	20.6
G/T (dB/K)	-25.4	-19.7 ^a	-19.8 ^a	-19.5 ^a
Translation Frequency Drift				
Short Term	±10 ⁻⁶			
Long Term	±10 ⁻⁶	3 × 10 ^{-6b}	2.5 × 10 ^{-6b}	5.5 × 10 ^{-6b}
Carrier-to-Intermodulation, 2-Carrier Saturation (dB)	>3	10.8	11.6	10.6
AM/PM Saturation (deg/dB)	<7.5	5.6	4.0	6.4 [3]
Group Delay (ns)	100	54	58	40
Repeater Gain Slope (dB/MHz)	1.5	0.4	0.3	0.48
Receive Out-of-Band Response ≥ 20 dB Down (MHz)	±10 MHz	well within specifications	-6.9, +7.6	-8.7, +8.45
In-Band Spurious Outputs (dB)	-20	<-65	<-65	<-65
Antenna Polarization				
Transmit		Right-hand circular polarization, L-band		
Receive		Right-hand circular polarization, C-band		
Normal Operation		At saturation; single or multiple channels per carrier		

^a Excluding increased noise due to output stage nonlinearities.

^b Change over total temperature range, excluding aging.

TABLE 9. MARISAT SHIP-TO-SHORE REPEATER CHARACTERISTICS

Satellite Parameter	Specification Value	Test Data		
		F-1	F-2	F-3
Receive Band, -1 dB (MHz)	1,638.5-1,642.5			
Transmit Band, -1 dB (MHz)	4,195-4,199			
Repeater Bandwidth at 10-dB Output Backoff (MHz)	+2, -2 19	+3.3, -3.6	+3.1, -3.15	+2.93, -3.35
Antenna Beamwidth (deg)				
Axial Ratio (dB)				
Transmit	<4	3.7	2.4	2.0
Receive	<4	3.7	3.6	3.8
Saturation Flux Density (dBW/m ²)				
High Gain Step	-99.5, ±2.3	-99.33	-99.6	-100.1
Low Gain Step	-96 ±2	-96.1	-95.9	-96.7
Saturation e.i.r.p. (dBW)	18.8	19.9	20.0	20.4
G/T (dB/K)	-17	-17.1	-16.9	-16.5
Translation Frequency Drift				
Short Term	10 ⁻⁶			
Long Term	10 ⁻⁵			
Carrier-to-Intermodulation, 2-Carrier Operation (dB)		4 × 10 ^{-6a}	8 × 10 ^{-6a}	5 × 10 ^{-6a}
AM/PM Saturation (deg/dB)	>9	11.0 ^b	10.2 ^b	11.1 ^b
Group Delay (ns)	<8	4.8	3.1	3.04
Repeater Gain Slope (dB/MHz)	100	55	58	47
Repeater Out-of-Band Response > 20 dB Down (MHz)	<1.5	0.3	0.25	0.4
Spurious Output (dB)				
In-Band	±10 MHz	well within specifications	-7.3, +7.8	-7.70, +7.15
Out-of-Band	<-60	<-65	<-60	<-65
Antenna Polarization	<-60	<-70	<-60	<-70
Transmit				
Receive				
Normal Operation				

Left-hand circular polarization, C-band
 Right-hand circular polarization, L-band
 Linear, Single or multiple channels per carrier

^a Change over total temperature range, excluding aging. ^b -3-dB saturation.

level. In this mode the TWT couples to the C-band transmit horn antenna. During transfer orbit when telemetry signals saturate the C-band TWT, a high-level injection mode couples the TWT output to a lower gain, broad-beam bicone telemetry and command antenna.

There is a wide dynamic range in the signals received by the L/C repeater from ship transmitters operating over the full coverage zone of the spacecraft. Therefore, the repeater has been configured so that under fully loaded conditions at least 10-dB input backoff remains at the TWT. Such linear operation of the TWT minimizes small signal suppression effects and intermodulation products.

In the shore-to-ship direction, shore station signals at 6,422 MHz are received by an earth coverage horn antenna. Without pre-amplification, these signals are translated through a dual-conversion process to 1,539 MHz. Filtering and amplification within the all-solid-state receiver are performed in the IF band of 88.67 ± 2 MHz. The final stage power amplifier consists of redundant L-band TWTs which have three switch-selectable levels of operation (to share available prime power with corresponding selectable modes of UHF power amplifier operation). The TWT, which is capable of providing saturated output levels of 7, 30, or 60 W, is connected to an input/output diplexer and drives the L-band quad-helix antenna. A 1,541.5-MHz cw beacon, inserted at IF, provides an L-band signal which can be used by ship terminals for acquisition and tracking of the satellite.

In both the L/C and C/L repeaters, translation frequencies are maintained to within $\pm 1 \times 10^{-5}$ under all conditions during the lifetime of the spacecraft. MARISAT repeater characteristics are given in Tables 8 and 9.

Ship terminals

Before a ship terminal is allowed access to the MARISAT system, it must comply with technical requirements established by the MARISAT Joint Venture [9]. Ship terminals must be able to transmit and receive duplex teleprinter signals and receive shore-originated one-way teleprinter messages, including broadcasts. Duplex voice grade operation with and without companders is optional. In addition, the terminals must automatically recognize any assignment message addressed to the terminal, respond automatically to applicable commands, and clear channels as required.

Important RF characteristics for the terminal are summarized in Table 10. All transmissions are at the same power level regardless of ship location within the satellite coverage area. Further, the TDMA signals and

request messages use similar transmission characteristics so that common equipment and terminal costs are reduced.

TABLE 10. MOBILE TERMINAL CHARACTERISTICS

G/T	≥ -4.0 dB/K at elevation angles $\geq 5^\circ$
e.i.r.p. of a Single Carrier in the Direction to the Satellite	37 ± 1 dBW
Spurious and Noise Output e.i.r.p. in any 4-kHz Band	Below the spectrum envelope defined by the following data points:
	e.i.r.p. per 4 kHz (dBW)
	Frequency (MHz)
	1,535.0
	1,622.5
	1,636.5-1,645.0 ^a
	1,660.0
	1,750.0
	Below 1,535.0, above 1,750.0 ^b
Harmonic Output e.i.r.p.	< -22 dBW for any frequency up to 40 GHz
Frequency Accuracy	± 250 Hz maximum
Power Density Spectrum of Phase Noise Induced on a TDMA Carrier	Not exceeding the envelope shown in Figure 13

^a All spurious signals shall be at least 60 dB below the rated power output level in this band, excluding a ± 25 -kHz band about the transmitted carrier where the spurious signal shall be limited to 50 dB below the carrier.

^b Excluding harmonics.

Antenna system characteristics are given in Table 11. The antenna must have the capability of being steered toward a quasi-geostationary (orbit inclination $\leq 5^\circ$) satellite whose elevation angle is 5° or higher. The antenna beam must be continuously steered to satisfy the G/T and e.i.r.p. requirements.

TABLE 11. ANTENNA SYSTEM CHARACTERISTICS

Peak Sidelobes	≤ 8 dBi
Sidelobes Beyond 45°	≤ 0 dBi
Axial Ratio	≤ 2 dB on-axis
Polarization	Right-hand circular for both receive and transmit, in accordance with the definition in C.C.I.R. Report 321.
Gain	Antenna gain at both the receive and transmit frequencies shall be such that the specified G/T and e.i.r.p. requirements are satisfied.

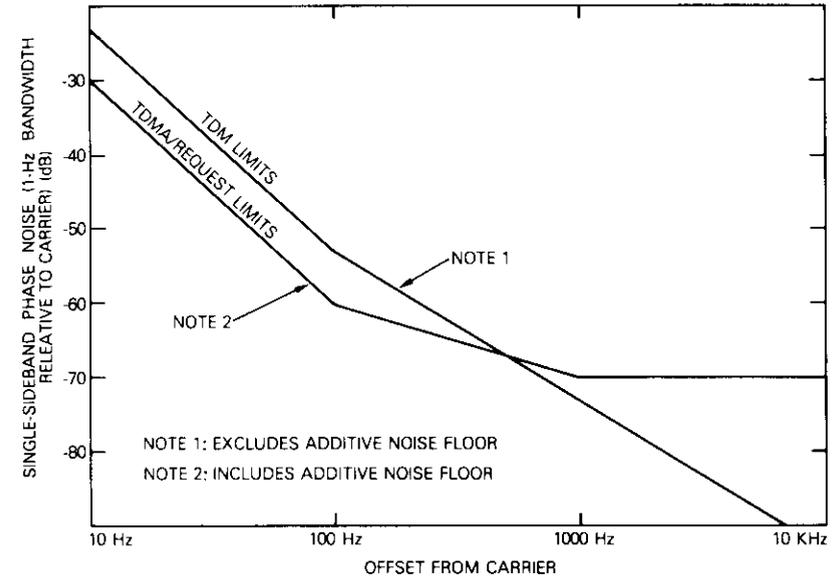


Figure 13. Phase Noise Limits

Manufacturers may design terminals to their own environmental specifications. Typical environmental conditions for terminals operating on a wide variety of commercial vessels are given in Table 12.

TABLE 12. TYPICAL ENVIRONMENTAL CONDITIONS

Ambient Temperature	-40°C to 65°C for equipment installed above deck 0°C to 55°C for equipment installed below deck
Relative Humidity	$\leq 95\%$
Spray	Solid droplets from any direction
Icing	≤ 1 inch of ice
Precipitation	≤ 10 cm/hr
Wind	Normal operation with relative average wind ≤ 140 km/hr (75 knots)
Shock and Vibration	As experienced in commercial vessels
Prime Power Variations	
Frequency	$\pm 6\%$
Voltage	$\pm 10\%$
Radio Frequency Interference	As experienced in commercial vessels

(Continued)

TABLE 12. (Continued) TYPICAL ENVIRONMENTAL CONDITIONS

Vessel Motions	Amplitude	Period	Above Deck
			Equipment Installation Distance from Motion Axis
Roll	$\pm 25^\circ$	4π s	30 m
Pitch	$\pm 7.5^\circ$	3π s	76 m
Yaw	$\pm 4^\circ$	100 s	70 m
Surge	± 0.1 g		
Sway	± 0.1 g		
Heave	± 0.25 g		
Turning Rate	$6^\circ/\text{s}$		
Headway	55.5 km/hr (30 knots)		

Shore stations

The MARISAT shore stations contain two major equipment groups: the RF system and the communications and control equipment. An RF system consists of a 12.8-m (42-ft) diameter parabolic antenna, redundant uncooled parametric amplifiers, and klystron power amplifier subsystems. The system transmits and receives C-band communications, signaling carriers, and L-band test signals, as well as AFC (automatic frequency control) signals, which are transmitted and received in both bands. Communications electronics and control equipment is used for compensation for frequency errors, modulation/multiplexing, channel access and signaling, network control, monitoring, and inter connection with terrestrial communications facilities.

Frequency compensation

One critical factor in the MARISAT system is the compensation for frequency translation errors. Significant frequency translation errors can occur on the satellite links, as shown in Table 13. Removal or compensation for unwanted frequency offsets is important for two major reasons. First, the equipment problems associated with AFC are alleviated. Second, the TDMA signal is detected on a burst-by-burst basis and the acquisition time for carrier and bit timing recovery depends upon residual frequency offset.

The shore station transmit and receive frequencies are adjusted automatically to compensate for satellite frequency translation errors. Each shore station has an L-band transmit/receive capability; AFC pilots looped

TABLE 13. FREQUENCY TRANSLATION ERRORS (MAXIMUM)

Source	Frequency Translation Error		
	Shore to Ship (Hz)	Ship to Shore (Hz)	Compensation by AFC
Satellite Translation	48,830	25,565	Yes
Shore Station Up- and Down-Converters	960	630	Yes
Shore-Satellite Doppler from Orbit Inclination (5°)	640	420	Yes
Ship Motion Doppler (15 m/s)	77	82	No
Ship-Satellite Doppler from Orbit Inclination (5°) ^a	205	219	No ^a
Ship Down- and Up-Converters ^b	240	250	No

^a The effect of ship-satellite errors from orbit inclination is compounded for ships in the southern hemisphere by the use of simple shore-satellite AFC loops, although ships in the northern hemisphere will see a reduced inclination error.

^b Ship frequencies are assumed to be aligned over a year and to have an aging rate of $\leq 1.5 \times 10^{-7}$ parts/year.

through each of the two MARISAT transponders provide control signals for the automatic adjustment.

To reduce the L-band frequency errors seen by the mobile terminals, each MARISAT shore station compensates for frequency translation errors on its shore-to-ship (C/L) communications signals. TDM and telephony signals are combined at the transmit intermediate frequency (between 88.5 and 92.5 MHz in the U.S. shore stations) and translated to the 6,420- to 6,424-MHz frequency band in two up-conversion steps. The stages of up conversion are shown in the top half of Figure 14.

Compensation for the frequency translation error is achieved in the first up-converter mixer, which obtains its mixing signal from the C/L pilot receiver. The AFC pilot is combined with the communications signals after the first up-converter mixer so that its transmitted frequency remains independent of the frequency compensation.

After translation by the C/L satellite repeater, the AFC pilot is received at an L-band frequency, which contains the cumulative Doppler and satellite translation errors, and compared against a locally generated reference. A voltage-controlled crystal oscillator (VCXO) is forced to run with a compensating frequency error to achieve phase lock with the received C/L pilot signal. The multiplied output of this VCXO is the mixing signal for the first up-converter stage of the C-band up-converter.

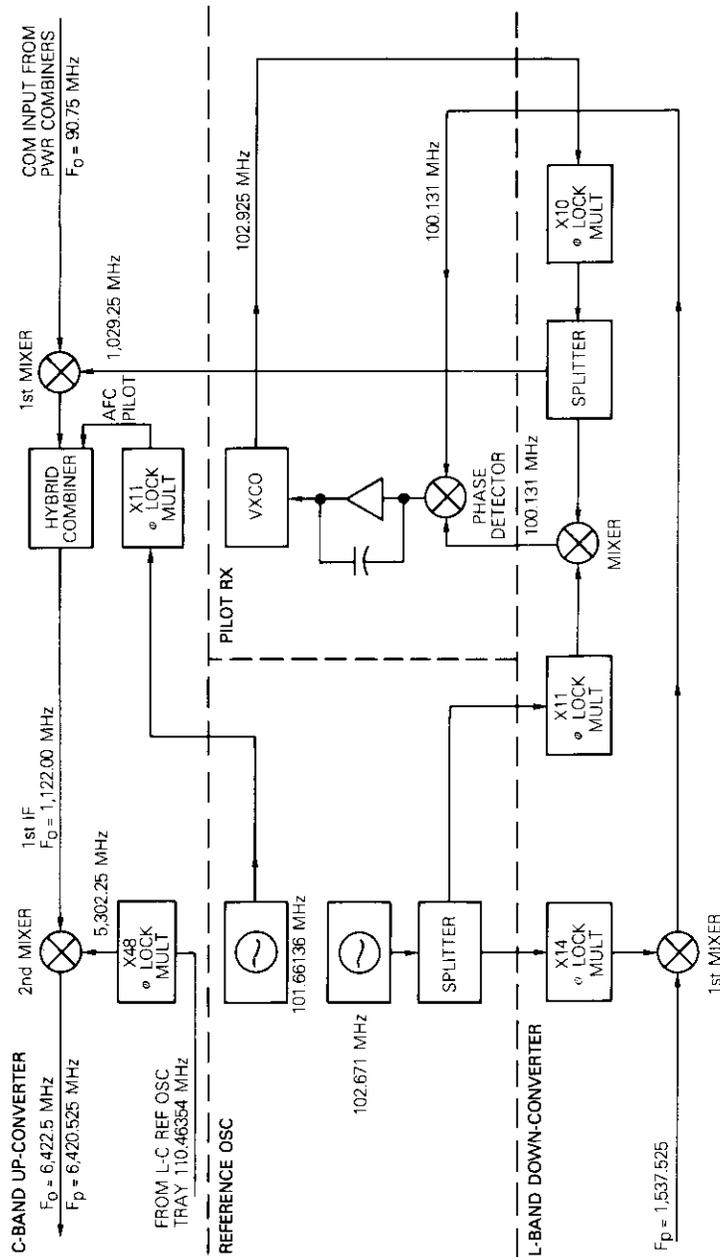


Figure 14. Up-Conversion and C/L AFC

The C/L AFC operates in an open loop configuration with the transmitted communications carriers offset in frequency to compensate for subsequent frequency translation errors of the satellite. Since the system is open loop, several shore stations at approximately the same latitude can utilize a single AFC pilot transmitted by only one of the stations (system reference station). The frequency error introduced by sharing the pilot is caused by the difference in shore station to satellite Doppler shifts among the stations. For 3° satellite inclination and mid-latitude stations, a latitude difference of 10° would introduce a frequency error of only 60 to 70 Hz.

The ship-to-shore frequency translation error is removed in each shore station's C-band receiver. An L/C AFC pilot is transmitted through the satellite L/C repeater and received by the shore station's AFC receiver. In the U.S. shore stations, a double conversion process translates the 4,195- to 4,199-MHz communications band down to the station IF band of 87-91 MHz as shown in Figure 15. The first translation uses a fixed frequency mixing signal, while the second translation uses a mixing frequency phase locked to the received pilot frequency.

Network control processor

The network control processor (NCP) controls telex and telephone calls which transit a MARISAT shore station. It processes all control messages between ships and the shore terminal, monitors telex and telephone channels for new calls or clearing old calls, and controls the switches (and line signaling) for routing all calls. Its main functions fall into the following categories:

a. Automatic processing of telex calls. The NCP detects call requests and automatically sets up and clears both ship-to-shore and shore-to-ship calls. Alternatively, it can set up and clear shore/ship calls manually under control of an operator in the telex supervisory position. The telex operator can also initiate and receive calls from the telex gateways on dedicated operator trunks.

b. Automatic setup and clearing of voice calls. In the automatic mode, the NCP detects and sends the signals and control information necessary to set up ship-to-shore voice calls automatically. Shore-initiated calls are always routed to the operator for connection to a ship. All signaling and connections in the voice switch are controlled by the NCP program.

c. Manual voice call processing. A single voice operator position controls manual setup of calls. The voice operator can originate or receive calls with either ships or the terrestrial voice network. The

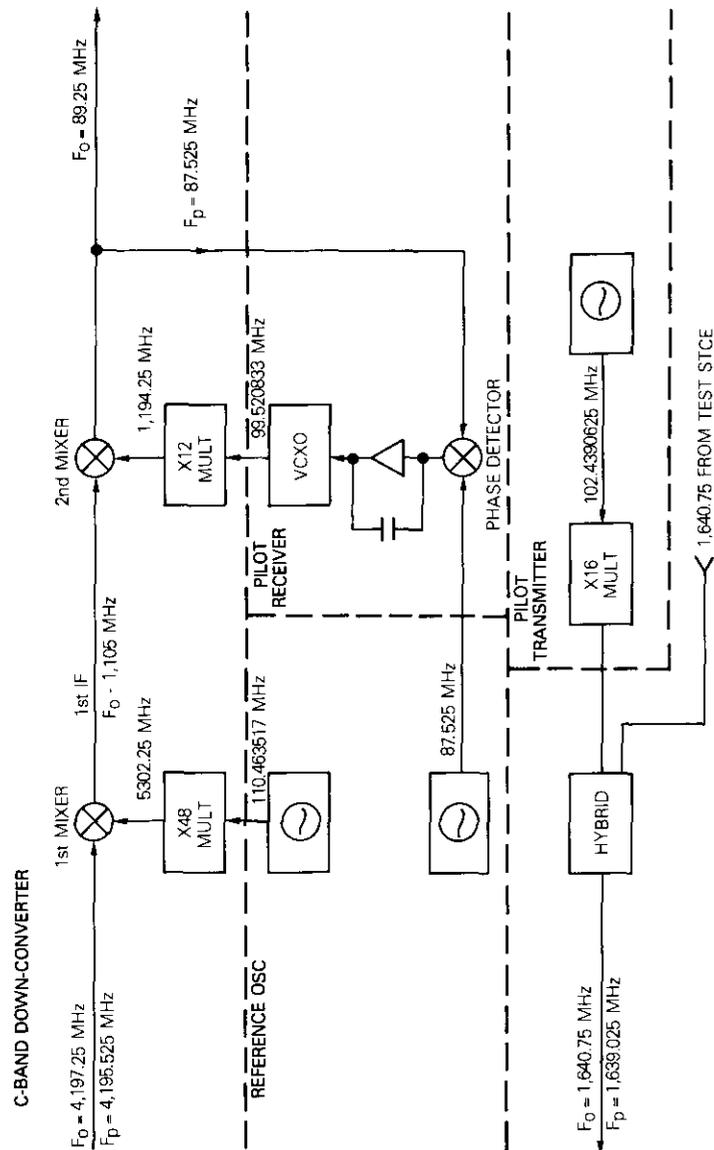


Figure 15. Down-Conversion and L/C AFC

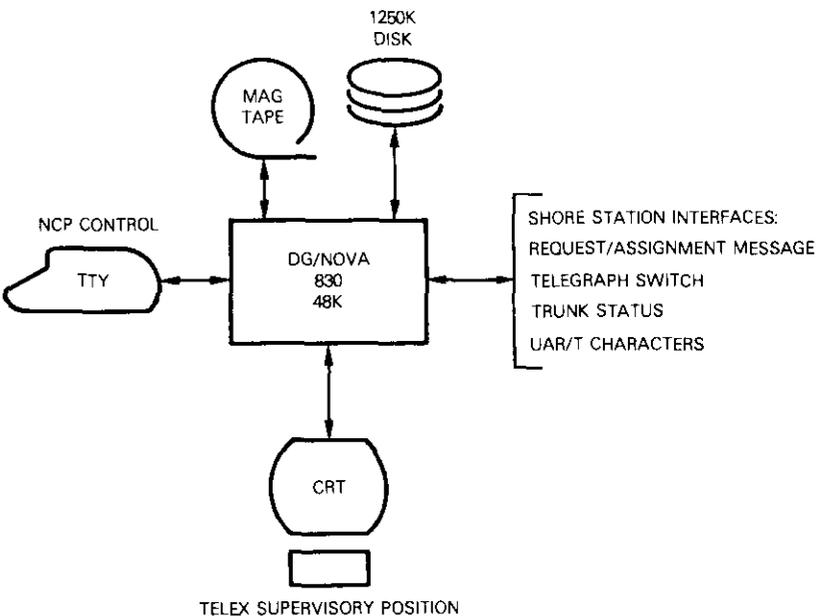


Figure 16. NCP Hardware and Interfaces

voice operator can answer, hold, connect, or re-enter voice calls and maintains a queue of requests for voice service. The NCP displays trunk, call status, and queue information on the operator's CRT. The NCP program interacts with the operator to perform all call control functions.

d. *Data logging.* The NCP logs statistical telex and telephone traffic data on magnetic tape or disk. Log tapes are processed off-line.

e. *NCP operator control.* The NCP program allows the NCP operator to control, monitor, and test the MARISAT shore station using teletype operator commands and error messages.

f. *Automatic failure detection.* As the NCP program processes calls, it maintains failure data which may lead to automatically removing a hardware unit from service. As part of this process, the NCP makes test telex calls to a test ship terminal located at the shore station.

g. *Support functions.* To support its major functions, the NCP performs a number of background tasks including maintaining internal operating statistics and running timers.

Figure 16 shows the NCP and its associated peripheral equipment. The

program runs on a standard Data General Nova 830 computer using a maximum of 48K of 16-bit core memory. The teletype controls and monitors the operation of the NCP and all call processing functions. The magnetic tape unit is used primarily for data logging, and in some cases to reload the program or NCP trunk definition data. The disk unit is used for reloading the program and as a backup data logging device for the magnetic tape unit. Certain parts of the program are disk resident and loaded by the operating program.

Conclusion

The MARISAT system has introduced a new communications era for commercial maritime interests. For the first time, high-quality voice, data, facsimile, and teletypewriter services with ships can be directly interconnected with domestic and international networks on a real-time basis.

The MARISAT system design has sufficient flexibility to accommodate new and improved capabilities. New services are being considered to meet the requirements of the shipping and offshore industries. Although MARISAT initially employs only one shore station in each ocean area, the design concepts will permit multiple-shore-station operation. Therefore, any ship will be able to communicate directly with any shore station in the satellite coverage area, leading to the development and expansion of maritime satellite communications into a highly used international service.

Acknowledgment

The individuals who made valuable contributions to the success of the MARISAT program are far too numerous to mention here. However, the authors wish to acknowledge the efforts of E. J. Martin, L. M. Keane, F. Giorgio, I. N. Knight of COMSAT GENERAL; C. Wolejsza of COMSAT Laboratories and R. McClure, formerly of COMSAT Laboratories, in the design of the MARISAT System.

References

- [1] L. M. Keane and E. J. Martin, "MARISAT—Gapfiller for the Navy Satellite Communications," *Signal Magazine*, Vol. 29, No. 3, November/December 1974.
- [2] C. Dorian, T. O. Calvit, and D. W. Lipke, "MARISAT Design and Operational Aspects," *Journal of the British Interplanetary Society*, Vol. 29, No. 5, May 1976.

- [3] International Telegraph and Telephone Consultative Committee (C.C.I.T.T.), "Standardization of Data-Signalling Rates for Synchronous Data Transmission in the General Switched Telephone Network," *Fifth Plenary Assembly, Green Book*, Vol. VIII, Geneva, 4-15 December, 1972.
- [4] International Telegraph and Telephone Consultative Committee (C.C.I.T.T.), "Characteristics of Compandors for Telephony," Recommendation G.162, *Fifth Plenary Assembly, Green Book*, Vol. III-1, Geneva, 4-15 December, 1972.
- [5] International Radio Consultative Committee (C.C.I.R.), "Subjective Evaluation Results of a Frequency Modulation Equipment for Use in the Maritime Mobile Satellite Service," Study Group 8, U.S. Contribution No. 8/1/19, March 2, 1977.
- [6] International Telegraph and Telephone Consultative Committee (C.C.I.T.T.), "Telegraph Operation and Tariffs," *Fifth Plenary Assembly, Green Book*, Vol. II-B, Geneva, 4-15 December, 1972.
- [7] International Telegraph and Telephone Consultative Committee (C.C.I.T.T.), "Telegraph Technique," *Fifth Plenary Assembly, Green Book*, Vol. VII, Geneva, 4-15 December, 1972.
- [8] International Telegraph and Telephone Consultative Committee (C.C.I.T.T.), "Telephone Signalling and Switching," *Fifth Plenary Assembly, Green Book*, Vol. VI-1-VI-4, Geneva, 4-15 December, 1972.
- [9] COMSAT General Corporation "Technical Requirements for MARISAT Mobile Terminals," June 17, 1977.



David W. Lipke received B.S. and M.S. degrees in electrical engineering from Massachusetts Institute of Technology in 1956. Prior to joining COMSAT, he was the Communications Analyst for NASA's Applications Technology Satellite Program. He joined COMSAT in 1966 and was involved in systems engineering for the Aeronautical and U.S. Domestic Satellite Programs. He later served as Manager of the Special Projects Department. Since 1973, Mr. Lipke has been Director, Maritime Systems Engineering, with responsibility for planning and engineering of the commercial portion of the MARISAT system. He is currently Chairman of the EASCON Board of Directors.



D. W. Swearingen received a B.S.E.E. from the Georgia Institute of Technology in 1963 and an M.S.E.E. from Stanford in 1965. He joined the Spectrum Utilization Department at COMSAT in 1970 after two years in the U.S. Army and three years as an engineer with the ITT Telecommunications Transmission Department. Since 1974, he has been Manager of the System Design Department with responsibility for the MARISAT commercial communications system design, specification of ship terminal and shore station equipment

characteristics, and interference coordination.

Jeremy F. Parker received a B.S.C. degree in Physics from the Imperial College of London University in 1963. Since joining COMSAT in 1968, he has participated in the planning and implementation of both the SPADE and MARISAT systems. Since 1974, he has served as Senior Design Engineer with the MARISAT program and has been largely responsible for the MARISAT access and control design.



Ernst E. Steinbrecher received the degree of Diplom Ingenieur from the Institute of Technology in Vienna, Austria. He was engaged in microwave research and development for the German Air Force and later headed his own laboratory in Vienna, working in the field of communications techniques. From 1953 to 1966, he held positions at the U.S. Army Signal Corps Engineering Laboratories and several other organizations, where he developed microwave components and explored new means of signal processing.



He joined COMSAT in 1966. For eight years he was a member of the technical staff of the RF Transmission Laboratory. Since 1974, he has been a Senior Design Engineer in the System Design Department where he assisted in the design of the MARISAT System.

Thomas O. Calvit received his B.S.M.E. from Louisiana State University in 1955 and M.S.E. from the University of Michigan in 1959. From 1955 to 1962 he served as an officer in the U.S.A.F. He then worked for General Electric's Radio Guidance Operation, where he was program manager for studies of communications and range instrumentation satellites. During the past 9 years at COMSAT and COMSAT GENERAL, he has been extensively involved in the development of mobile satellite systems and participated in the conception and development of the MARISAT system. As Manager, Systems Analysis, he is responsible for MARISAT communications performance analysis.



Hans Dodel attended the Technische Universitaet Stuttgart, and in 1969 received his M.S. in E.E. from the University of Kansas. He then joined COMSAT, where in the Systems Engineering Division he worked on frequency planning and transmission impairment analysis for the INTELSAT III and IV space segments. Later, in the Systems Studies Division of COMSAT Laboratories, he was a member of the INTELSAT V Study Team, and was responsible for tradeoff studies, link budgets, and transmission analyses. He is presently a senior engineer in the Systems Analysis Department, where he is responsible for MARISAT transmission analysis, evaluation of traffic statistics, and analysis of new applications.

Echo canceller with adaptive transversal filter utilizing pseudo-logarithmic coding

O. A. HORNA

(Manuscript received April 11, 1977)

Abstract

This paper describes a digital echo canceller (DEC) with a novel adaptive transversal filter capable of enhancing the echo return loss by more than 22 dB. The residual echo is removed by an adaptive center clipper.

Pseudo-logarithmic analog-to-digital conversion of the signal is employed for convolutional processing, hence reducing the memory size and simplifying the convolution processor. A new convergence algorithm using nonlinear stabilization of the feedback loop substantially reduces the cross-correlator complexity and shortens the convergence time of the transversal filter. Hence, the DEC performs well even during "double talk" and "phase roll," which are among the most adverse circuit conditions an echo canceller may encounter. The center clipper, which has an adaptive threshold, does not introduce perceptible distortion while eliminating residual echo.

Introduction

In long-distance telephone connections involving both 4-wire and 2-wire links, an echo signal may be generated primarily at the hybrid coil that converts 4-wire to 2-wire transmission. The echo return loss (ERL) is defined as the normalized ratio of speech signal power to echo signal power.

When it is measured directly at the hybrid coil, it has a distribution which is approximately normal with a mean μ_{ERL} of 11–15 dB and a standard deviation σ_{ERL} between 3 and 5 dB [1], [2]; thus, the ERL can be 5 dB or less in as many as 10–12 percent of the cases.

The subjective disturbing influence of echo increases with the round trip delay, t_{RD} , of the connection. Even for a relatively short delay of 20 ms, the required mean ERL must be well in excess of 11 dB to provide satisfactory service [3], [4]. With round trip delays encountered in long-distance international terrestrial connections ($t_{RD} > 50$ ms) or satellite communications ($t_{RD} \approx 500$ ms), μ_{ERL} must be greater than 30 dB, necessitating an echo control device. Considerable effort has been directed [5], [6] toward improving echo suppressors, which are basically voice activated switches interrupting the return path in the 4-wire section. However, these devices may not provide adequate quality for 2-way communications, especially in the "double talk" condition when $ERL < 15$ dB and $t_{RD} > 100$ ms. Therefore, different approaches have been investigated.

As reported by Sondhi [7], Flanagan et al. synthesized a transversal filter approximating the measured impulse response of the echo path, *i.e.*, the 4-wire carrier system, hybrid coil and 2-wire section. Subtraction of the transversal filter output (*i.e.*, the simulated echo signal) from the actual echo resulted in effective echo cancellation (as opposed to suppression). This concept was improved by Kelly, who suggested the use of self-adaptive multipliers for the tap gain settings of the transversal filter.

Sondhi and later Campanella et al. [8] derived the linear theory of convergence and stability of the echo canceller adaptive filter loop. Campanella et al., in cooperation with the Nippon Electric Company, developed a digital echo canceller which underwent an extensive field trial on satellite links, demonstrating that echo cancellers can significantly improve the performance of circuits with long delay. However, the earlier design exceeds the limits of economic viability [9] because the equipment is nearly two orders of magnitude more complex than that of the most sophisticated echo suppressors [6].

This paper presents the design philosophy of an echo canceller whose complexity, size, and cost are comparable to those of echo suppressors. The end result can be judged from the composite photographs in Figure 1, which compare the latest model of the new echo canceller with the version used in worldwide tests.

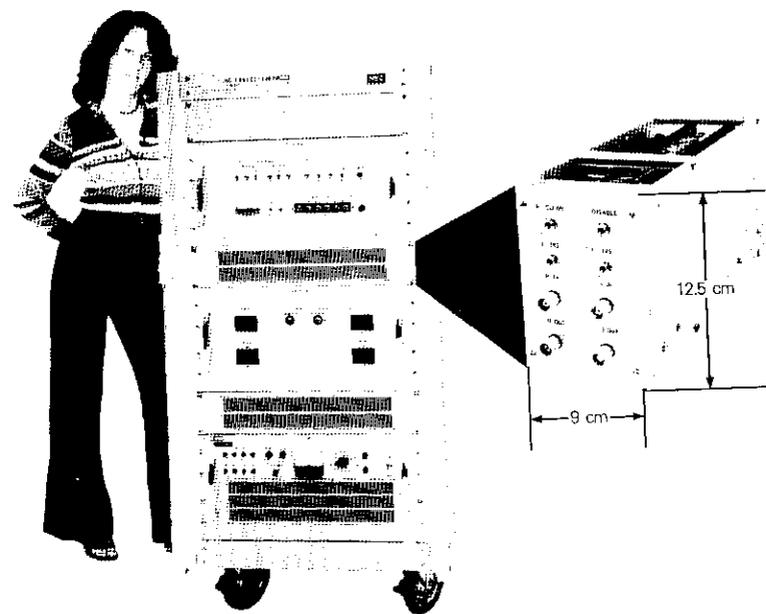


Figure 1. Size Comparison of the Echo Canceller Used in the Worldwide Tests with the New Echo Canceller Model

Digital echo canceller design

Figure 2 shows the echo canceller inserted in the 4-wire circuit, usually at some distance from the hybrid coil (*i.e.*, with some delay t_D) where the echo is generated. In the H -register, n samples, h_i , of the unit impulse response of the echo path are stored in digital form. The X -register stores n most recent speech samples, x_{i+j} , from the analog-to-digital converter (ADC).

During the j th sampling period, the j th estimate of the echo, r_j , is computed in the convolution processor:

$$r_j = \sum_{i=1}^m h_i \times x_{i+j} = H(i) * X(j) \quad (1)$$

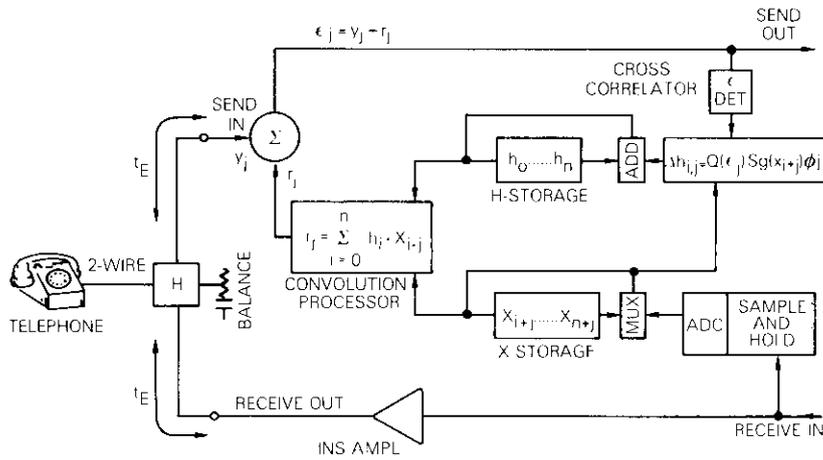


Figure 2. Block Diagram of the DEC [10]

This estimate is subtracted from the true echo, y_j , resulting in the residual echo signal sample,

$$\epsilon_j = y_j - r_j \quad (2)$$

at the output of the subtractor. This residual signal, which serves as an error signal, is nonlinearly quantized according to a function $Q(\epsilon_j)$. A correction, $\Delta h_{i,j}$, for every impulse response sample h_i is computed by using a modified steepest descent method [7], [10], as follows:

$$\Delta h_{i,j} = \mu \epsilon_j \frac{x_{i+j}}{\sum_i x_{i+j}^2} = \mu Q(\epsilon_j) \text{Sgn}(x_{i+j}) \phi_j \quad (3)$$

where ϕ_j is a function of $x_{i+j}/\sum_i x_{i+j}^2$. The correction term $\Delta h_{i,j}$ is added with the proper sign to every corresponding sample h_i so that a new value, h'_i , is obtained

$$h'_i = h_i + \Delta h_{i,j} \quad (4)$$

This process is repeated to compute the next [($j + 1$) th] estimate, r_{j+1} , for a new set of x_{i+j+1} speech samples. The new error value, ϵ_{j+1} , is used to compute another set of corrections, leading ultimately to a close replica of the impulse response which is built up in the H -register and continuously updated whenever changes occur. This algorithm ensures convergence

[8], [10] if $|\epsilon_{j+1}| < |\epsilon_j|$, i.e., if μ in equation (2) is less than 2.

Values of ERL > 30 dB are required in circuits with delay in excess of 100 ms; therefore, the echo canceller must attenuate the echo signal power $P(y_j)$ by more than 26 dB under the worst-case condition; i.e., the echo return loss enhancement (ERLE) must be

$$\text{ERLE} = 10 \log_{10} \frac{P(\epsilon_j)}{P(y_j)} = -10 \log_{10} \frac{\bar{\epsilon}_j^2}{\bar{y}_j^2} \geq 26 \text{ dB} \quad (5)$$

The dynamic range of speech is about 40 dB, and the possible amplitude range of the different impulse responses exceeds 24 dB [2], [11]. Therefore, the quantization of speech samples x_{i+j} requires 12 bits and that of the impulse response samples h_i at least 9 bits.

The end delay, t_E , for satellite circuits connected to the terrestrial telephone networks is usually less than 15 ms; consequently, the impulse response can be delayed by twice as much, or 30 ms. If a sampling period, $T_S \leq 125 \mu s$, and approximately < 30 samples h_i are used to emulate the echo response (see Appendix A), at least $n \geq 2t_E/T_S = 250$ samples must be stored in the H - and X -registers for processing.

The echo estimate r_j is computed according to equation (1) by performing 250 multiplications of 12 by 9 bits and 250 additions of these products during every sampling period.

As the operational speed of the convolution processor is limited mainly by the access time of available memories, parallel processing in the multiplier seems to be the only practical means of avoiding custom-built circuits. The parallel multiplier with memories represents most of the hardware and cost of present echo canceller designs [8], [9].

DEC with logarithmic coding

With reference to Figure 3, the following steps have been taken to reduce the hardware complexity:

a. Speech samples x_{i+j} are encoded in a sign plus magnitude 13-segment pseudo-logarithmic 7-bit format with a linear segment around zero (similar to A -law format, see Appendix B) yielding the resolution and the dynamic range of an 11-bit linear code. Impulse response samples h_i are encoded in a similar 11-segment 8-bit format whose dynamic range is equivalent to that of a 10-bit linear code. This type of coding saves more than 28 percent of the memory space.

b. Multiplication in the convolution processor [see equation (1)], performed as an addition of the pseudo-logarithms of x_{i+j} and h_i ,

The analog value of r_j is then subtracted from echo signal y_j . For simplicity, assume unity digital-to-analog scaling factors. Then the zero exponents $e_x = e_h = 0$ indicate that the amplitude of the samples is in the linear segment of the chosen code and that their equivalent absolute values are

$$\begin{aligned} |x| &= 2 \times m_x \\ |h| &= 2 \times m_h \end{aligned} \quad (6)$$

For $e_x > 0$ and $e_h > 0$, samples are encoded quasi-logarithmically. The equivalent values of the samples are

$$|x| = 2^{e_x}(1 + m_x) \quad (7)$$

$$|h| = 2^{e_h}(1 + m_h) \quad (8)$$

where by definition the mantissas are $m_x, m_h < 1$ (see Appendix B). As there are four possible combinations of $|x|$ and $|h|$ formats entering the convolution processor's multiplier, four different algorithms must be used to form the product P_i .

If $e_x = e_h = 0$ and $|x|$ and $|h|$ are only three bits wide [see equation (6)], the product is

$$|P_i| = 2^2 \times m_x \times m_h \quad (9)$$

Product $|P_i|$ rounded off to four bits is found in a 256-bit ROM, a readily available hardware component. If either $e_x = 0, e_h \geq 1$, or $e_x \geq 1, e_h = 0$, then $|P_i|$ is

$$|P_i| = 2^{e_x} \times (1 + m_x) \times 2 \times m_h = 2^{e_x+1}(m_h + m_x \times m_h) \quad (10)$$

In this case, the 4-bit product $m_x \times m_h$ is read from the ROM and added to the mantissa with the zero exponent. The sum $m_h + m_x \times m_h < 2$ is ≤ 5 bits wide; therefore, a simple 4-bit MSI adder can be used. If both $e_x \geq 1$ and $e_h \geq 1$, an approximate expression for the binary logarithms is used in computing $|P_i|$. For \log_2 , the following formula is used:

$$\log_2(1 + m) \doteq m \quad (11)$$

and for antilog₂:

$$\text{antilog}_2 m = 2^m \doteq 1 + m \quad (12)$$

where $0 < m < 1$. The \log_2 of the product $|P_i|$ is therefore

$$\begin{aligned} \log_2 |P_i| &= e_x + \log_2(1 + m_x) + e_h + \log_2(1 + m_h) \\ &= e_x + e_h + m_x + m_h = E_i + M_i \end{aligned} \quad (13)$$

where exponent E_i is the whole part of the sum of equation (13) and mantissa M_i is the remainder < 1 . Product $|P_i|$ is thus

$$|P_i| = \text{antilog}_2(E_i + M_i) = 2^{E_i}(1 + M_i + \epsilon_M) \quad (14)$$

where ϵ_M is a correction which is a function of m_x and m_h . Equation (13) can obviously be implemented by two 4-bit adders, one for the exponent and the other for the mantissa. Multiplication by 2^{E_i} according to equation (14), where E_i is an integer $E_i \in \{2, \dots, 14\}$, is performed by shifting the partial products by E_i binary places before entering the accumulator (Figure 5).

In addition to circuit simplification, there are two other reasons for using this seemingly complicated multiplication scheme:

a. SPEED: Multiplication is performed by one read command from the ROM and a single addition. The entire operation of multiplying $x_{i+j} \times h_i$ and adding the product into the accumulator can be done in one clock period, T , which may be as low as 200 ns when 500 x_{i+j} and h_i samples are required for processing.

b. ZERO-CROSSING: This multiplication algorithm provides a simple solution to the problem of zero values associated with logarithmic encoding.

Error analysis

There are four main contributions to the total energy of the error signal $P(\epsilon_j) = k E[\epsilon_j^2]$ which limit the degree of cancellation:

- uncorrelated noise appearing at the send side,
- truncation of the impulse response $H(t)$,
- quantization noise,
- error due to the \log_2 approximation.

During normal operation the level of the send side noise is usually < -50

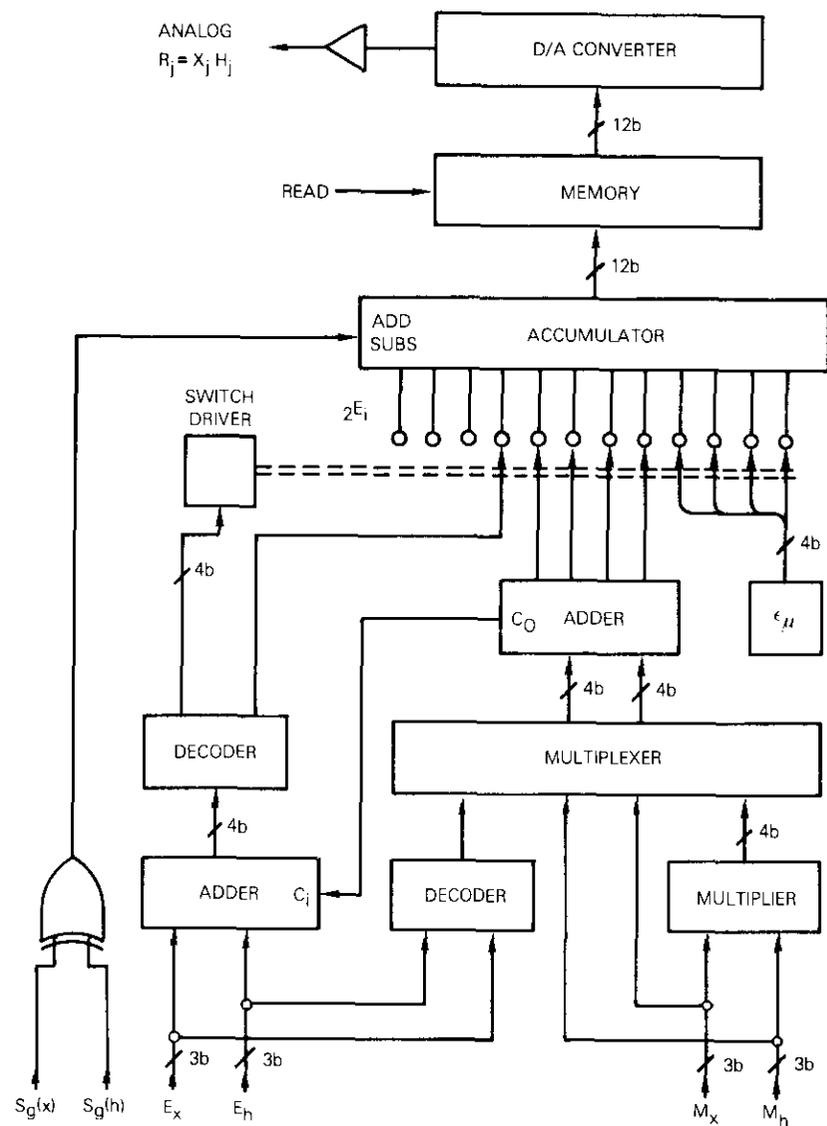


Figure 5. Block Diagram of the Convolution Processor

dBm0 and the number m of impulse response bytes h_i usually occupies only a small fraction of the capacity ($n \gg m$) of the memory (Appendix A).

Under these conditions, the degree of cancellation is determined mainly by the quantization errors Δh and Δx of samples h_i and x_{i+j} , respectively, entering the convolution processor [8], [11]. As shown in equation (1), the j th estimate of echo r_j is computed as a sum of products $P_i = h_i \cdot x_{i+j}$:

$$r_j = \sum_{i=1}^m P_i = \sum_{i=1}^m h_i \cdot x_{i+j} \quad (15)$$

If the process is stationary and samples h_i of stored response H_i converge to their optimal values, every product P_i will contain only a quantization error due to finite quantization steps Δ_h and Δ_x of samples h_i and x_{i+j} :

$$\begin{aligned} P_i + \Delta P_i &= (h_i + \Delta_h)(x_{i+j} + \Delta_x) \\ &= h_i x_{i+j} + \Delta_h x_{i+j} + \Delta_x h_i + \Delta_x \Delta_h \end{aligned} \quad (16)$$

Neglecting $\Delta_x \Delta_h$ in a first approximation leads to the quantization error

$$\Delta P_i = \Delta_h x_{i+j} + \Delta_x h_i \quad (17)$$

Because Δ_h and Δ_x are uncorrelated, the variance σ_i^2 is

$$\sigma_i^2 = E[(\Delta P_i)^2] = \frac{\Delta_h^2}{12} x_{i+j}^2 + \frac{\Delta_x^2}{12} h_i^2 \quad (18)$$

Since both samples x_{i+j} and h_i are encoded in a pseudo-logarithmic format [see equation (8)] except for a narrow linear range around zero [see equations (6) and (7)], quantization steps Δ_h and Δ_x are approximately a constant fraction $\delta \ll 1$ of sample h_i or x_{i+j} . Therefore,

$$\Delta_h \doteq \delta \cdot |h_i| \quad (19)$$

$$\Delta_x \doteq \delta \cdot |x_{i+j}| \quad (20)$$

Substituting into equation (18) yields

$$\sigma_i^2 = \frac{\delta^2}{6} h_i^2 x_{i+j}^2 \quad (21)$$

and the expected value of the error ϵ_j^2 of all products P_i of convolution (15) is

$$E[\epsilon_j^2] = \sum_{i=1}^m \sigma_i^2 = \frac{\delta^2}{6} \sum_{i=1}^m h_i^2 x_{i+i}^2 \quad (22)$$

which is proportional to the total power of the estimated echo signal $P(r_j) = r_j^2/R$. At the end of the convergence process, the estimated and true echo signals are nearly equal, whereby

$$y_j^2 \doteq r_j^2 = \sum_{i=1}^m h_i^2 x_{i+i}^2 \quad (23)$$

If equations (22) and (23) are substituted into equation (5) and $\delta^2 = 2^{-B}$, ERLE is a function of the equivalent number of quantization steps, 2^B , of the logarithmically encoded signals:

$$\text{ERLE} = -10 \log_{10} \frac{\delta^2}{6} = 3B + 7.8 \text{ dB} \quad (24)$$

The 7-bit pseudo-logarithmic code has a resolution, B , of 5 bits; *i.e.*, the expected enhancement is 22.8 dB, which is in relatively good agreement with measured values (as will be shown in Figure 11) for higher levels of signal $X(t)$.

For small amplitudes $|h_i|$ and $|x_{i+j}|$, equation (19) does not hold because the encoding is linear. That is, the quantization error increases with decreasing $|P_i|$. In the present DEC design this is the main reason for decreasing ERLE for low-level receive-in signals.

The relative errors of the \log_2 and antilog₂ approximations [see equations (11) and (12)] are

$$\epsilon_l = \log_2(1+m) - m = \Phi(m) \quad (25)$$

$$\epsilon_a = 2^m - ((1+m)) = \Phi_a(m) \quad (26)$$

where for given m the errors are approximately related, *i.e.*,

$$\epsilon_a = -\epsilon_l \rightarrow \Phi_a(m) = -\Phi(m) \quad (27)$$

Figure 6 shows $\Phi(m)$ as a function of m . For $m \doteq 7/16$ the error has a

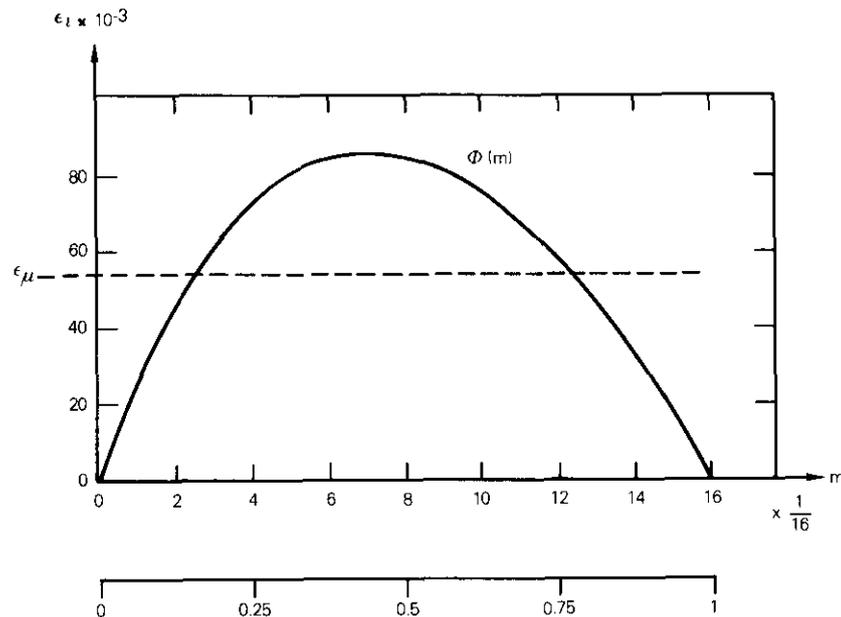


Figure 6. Error ϵ_l of the \log_2 Approximation as a Function $\Phi(m)$ of mantissa $0 < m < 1$

peak value, $\epsilon_p = 0.086$. As suggested by Dill, this error can be halved by adding $\epsilon_p/2$ to mantissa m when computing the \log_2 approximation [13].

It is advantageous in the convolution process [equation (1)] to zero the average relative error, μ , *i.e.*, to add a constant term

$$\epsilon_\mu = \int_0^1 \Phi(m) dm = 0.0573 \quad (28)$$

to the mantissa m of \log_2 or to subtract ϵ_μ when computing 2^m . For binary product $|P_i|$ according to equations (11)-(14), a correction ϵ_M , which is the binary representation of ϵ_μ , is applied, *i.e.*,

$$\epsilon_M = 0.0000111 \quad (29)$$

This correction can be directly added to the output of the 2^{B_i} multiplier, as shown in Figure 5, because the mantissa M_i is represented by only

3 bits. The rms value of the relative error of the \log_2 or antilog₂ approximation is

$$\epsilon_{rms} = \left(\int_0^1 [\Phi(m) - \epsilon_\mu]^2 dm \right)^{1/2} = 0.027 < 2^{-5} \quad (30)$$

which is also equal to the standard deviation, σ , because the average has been adjusted for zero.

If it is assumed that all three errors in computing the product according to equations (11) to (14) are uncorrelated and the central limit theorem is applied, the rms error of $|P_i|$ is $\sigma_p = \sigma \sqrt{3} \doteq 0.0468 \doteq 3 \times 2^{-6}$. Since a sufficiently low level of quantization noise is achieved with a 3-bit mantissa, the rms and probable errors $0.6745 \sigma_p \doteq 2^{-5}$ of the logarithmic approximation with a constant correction $\epsilon_M = \epsilon_\mu$ are sufficiently below this noise.

Correction algorithm

The h_i samples stored in the H-register are encoded into another 8-bit format with a 3-bit exponent $e_h \{0, 1, \dots, 6\}$ and a 4-bit mantissa, $0 < m_h < 1$. If $e_h = 0$, i.e., h_i is in the linear segment of the code around zero, the correction formula is the same as that given by equation (4):

$$h'_i = h_i + \Delta h_{i,j} = \text{Sgn}(h_i) \times |2m_h| + \Delta h_{i,j} \quad (31)$$

For exponents $e_h \geq 1$, the sample h_i is in pseudo-logarithmic format [equation (8)]:

$$\log_2 |h_i| = e_h + m_h = \log_2 2^{e_h} (1 + m_h) \quad (32)$$

The increment $\Delta_{i,j} \ll 1$ is added to the mantissa m_h of $\log_2 |h_i|$ to establish the new corrected value h'_i . In linear format this is equivalent to the product

$$|h'_i| = |h_i| (1 + \Delta_{i,j}) = |h_i| + |h_i| \Delta_{i,j} \quad (33)$$

Thus, the sample h_i is multiplied by $(1 + \Delta_{i,j}) > 1$ or $(1 - \Delta_{i,j}) < 1$ according to the sign of $\Delta_{i,j}$. In effect, the correction step $\Delta h_{i,j}$ is directly proportional to the amplitude of the sample h_i :

$$|\Delta h_{i,j}| = |h_i| \Delta_{i,j} \quad (34)$$

With this correction algorithm, the time of convergence, τ_{CN} (see Appendix C and References 7 and 8), is less dependent on the amplitude of the input response $H(i)$. For large values of h_i , the convergence is nearly eight times faster than with the correction method of equation (4), which adds the same correction to each value h_i .

The speed of convergence resulting from this adaptive algorithm permits the use of only one fixed value of correction increment $\Delta_{i,j} = \pm 2^{-4}$. However, another two-valued function $F(\epsilon_j)$ must be used for computing the correction $\Delta h_{i,j}$ [equation (3)] to satisfy the condition of convergence in the whole dynamic range of $X(i+j)$ and $H(i)$:

$$F(\epsilon_j) = 0, \quad \frac{|\epsilon_j|}{\beta X_j} < 1$$

$$F(\epsilon_j) = 1, \quad \frac{|\epsilon_j|}{\beta X_j} \geq 1 \quad (35)$$

$F(\epsilon_j)$ is used to initiate the correction process only if the error $|\epsilon_j|$ is greater than a fraction β of the rms value X_j of the input signal. $F(\epsilon_j)$ is implemented by making the error sensor reference voltage proportional to X_j so that the detector is less sensitive to higher level receive-in signals and vice-versa.

The correction $\Delta_{i,j}$ is added to the fourth bit of the mantissa, which does not enter the convolution processor, creating a "digital inert zone" in the correlator's loop. This inert zone contributes substantially to the stability of the H -response during extreme and transient conditions and in the presence of uncorrelated noise at the send-in path.

Generation of $\Phi(t)$ function

As described above, the functions of the correction loop and the adaptive center clipper are dependent on the rms value X_j of the m most recent speech samples. Since the computation of $F(\epsilon_j)$ does not require high precision, the true rms value X_j can be approximated by averaging the absolute values of $|x_{i+j}|$ with amplitudes greater than a threshold V_x :

$$X_j \doteq \bar{X}_j = 1/T \int_0^T \int_{V_x}^{\infty} |x_{i+j}| dx dt \quad (36)$$

For a sinusoid the average-to-rms ratio is $\mu/\sigma_s = 0.9$, and for white

noise it is 0.8. Therefore, the error involved in using an average \bar{X}_j as a reference for $F(\epsilon_j)$ and Φ_{i+j} [see equations (35) and (37)] cannot adversely influence the convergence of the canceller. This was confirmed by other experiments [8].

Figure 7 is a block diagram of a circuit which generates X_j and another

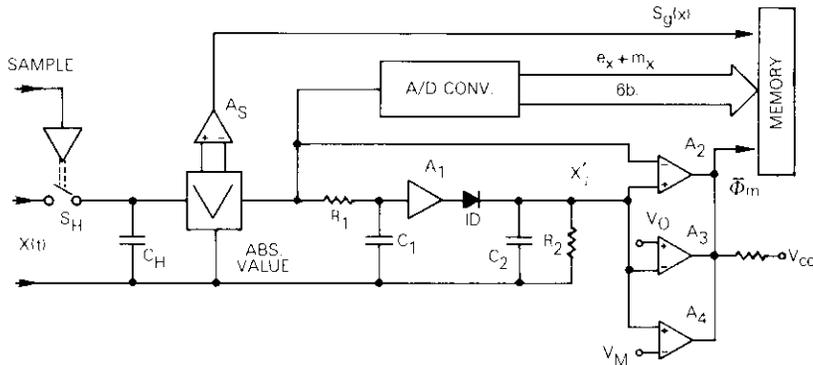


Figure 7. Block Diagram of the Pseudo-rms Detector and Φ_{i+j} Function Generator

two-valued function Φ_{i+j} used in the correlation loop. The sampled input signal is converted into a sign signal $\text{Sgn}(x_{i+j})$ and its absolute value $|x_{i+j}|$, which is averaged by low-pass filter $R_1 C_1$. A buffer amplifier feeds this voltage \bar{X}_j into a peak detector (diode ID) which stores the $[\bar{X}_j]_{max}$ in capacitor C_2 . If the ratio $R_2 C_2 / R_1 C_1$ is properly chosen, DC output voltage \bar{X}_j is approximately equal for sine wave and white noise signals with the same (true) rms values.

Every sample $|x_{i+j}|$ is compared (A_2, A_3, A_4) with this pseudo-rms value \bar{X}_j , the noise level threshold V_0 , and the overload level V_M . The comparator decisions are combined into a 2-valued function Φ_{i+j} [compare ϕ_{i+j} in equation (3)]:

$$\Phi_{i+j} = 1, \text{ for } \frac{|x_{i+j}|}{\alpha \bar{X}_j} > 1 \text{ and } V_0 < |x_{i+j}| < V_M$$

$$\Phi_{i+j} = 0, \text{ otherwise} \tag{37}$$

where α is a constant $1 \leq \alpha \leq 2$. The value of Φ_{i+j} is stored in the memory

as a part of the digital signal x_{i+j} and fed into the correlation processor to compute the H -response corrections:

$$\Delta h_{i,j} = \text{Sgn}(\epsilon_j) \text{Sgn}(x_{i+j}) \Phi_{i+j} F(\epsilon_j) \tag{38}$$

These corrections are used to correct the old values of h_i ; i.e., the correction is made only if $\Phi_{i+j} F(\epsilon_j) = 1$.

Because of the different buildup and release time constants $R_1 C_1$ and $R_2 C_2$, \bar{X}_j is a nonlinear function of time. Therefore, the function Φ_{i+j} is also time dependent, i.e., $\Phi_{i+j}(t)$. During the initial $H(t)$ response buildup, function $\Phi_{i+j}(t)$ only gradually increases its threshold $\alpha \bar{X}_j$. This contributes to the initial rapid convergence under the transients and to stability under the steady-state condition [14].

Stability of the correction loop

Under idealized conditions, the echo canceller in the circuit shown in Figure 2 can be represented by an adaptive filter which identifies an unknown system (see Figure 8). The discrete coefficients h_i must be adjusted

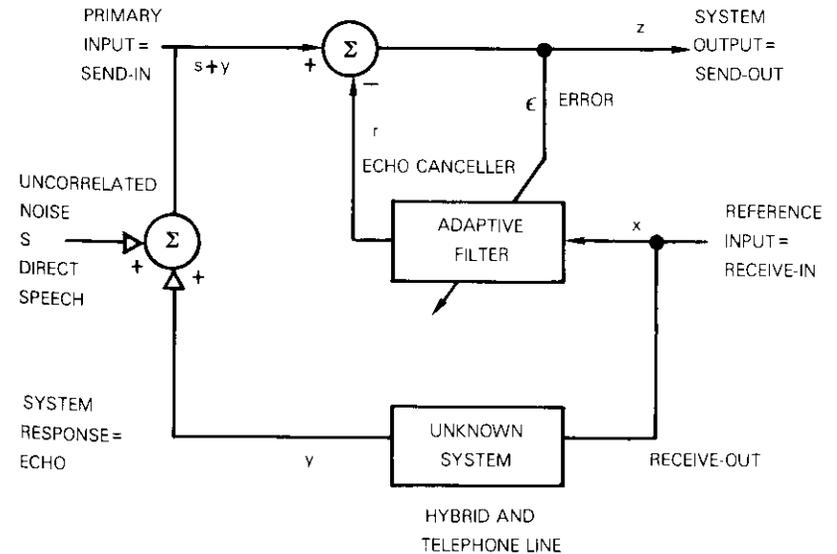


Figure 8. Equivalent Diagram of the DEC

so that the adaptive filter output emulates the unit impulse response $H(t)$ of the unknown system with minimum mean square error $E[\epsilon^2]$ (see Appendix A).

This process, also known as an identification process, is based on the well-known relationship [10] between input/output signal cross-correlation $R_{xy}(\tau)$ and system response $H(t)$:

$$R_{xy}(\tau) = \int_{-\infty}^{+\infty} H(t) R_{xx}(\tau - t) dt \quad (39)$$

where $R_{xx}(\tau - t)$ is the autocorrelation function of the input signal x . If it is assumed that no additive noise signal is present (its influence will be analyzed later), $H(t)$ can be unambiguously determined from equation (37), by using an input signal x , where $R_{xx} = \text{constant}$ for $\tau - t = 0$ and $R_{xx} = 0$ elsewhere (e.g., a Poisson wave):

$$H(t - \tau) \approx R_{xy}(\tau) = E[x(t) y(\tau - t)] \quad (40)$$

Speech, however, is a highly correlated signal. For example, successive speech samples x_{i+j} and x_{i+1+j} taken at intervals of $125 \mu\text{s}$ have an autocorrelation coefficient $\rho_{xx} > 0.8$ [15]; therefore, a dither signal [16] must be added to x . Function Φ_{i+j} defined in equation (37) serves this purpose: as the coefficient α increases, the sequence of pulses $|x_{i+j}| > \alpha \bar{X}_j$ more closely approximates the statistical properties of the Poisson wave, but the number of corrections [equation (38)] which can be made every convolution cycle decreases. It was determined experimentally that, for continuous speech and for the used quantization code, the optimum compromise is α between $1.2 < \alpha < 1.6$. For $\alpha = 1.4$ the probability is that the speech sample $|x_{i+j}| > 1.4 \bar{X}_j$ is $p_x \leq 0.2$ [17].

The correction algorithm according to equation (38) approximates the true cross correlation by the discrete sign correlation

$$H(t - \tau) \approx R_{xy}(\tau) \doteq B_{xy}(i) = E[\text{Sgn}(x_{i+j} \Phi_{i+j}) \text{Sgn}(y_j \theta_j)] \quad (41)$$

where θ_j must be another random or pseudo-random process statistically independent of the dither signal Φ_{i+j} . Function $F(\epsilon_j)$ defined by equation (36) fulfills this condition. Function $F(\epsilon_j)$ can change its value once while Φ_{i+j} can change n times every convolution cycle; hence independence is ensured. During the H response buildup process $|y_j| > |r_j|$; therefore, according to equation (2)

$$\text{Sgn}(\epsilon_j) = \text{Sgn}(y_j) \quad (42)$$

Substituting into equation (41) and separating both dither functions makes it possible to compute the discrete unit impulse response $H(i)$:

$$H(i) \approx B_{xy}(i) = E[\Phi_{i+j} F(\epsilon_j) \text{Sgn}(x_{i+j}) \text{Sgn}(\epsilon_j)] \quad (43)$$

which proves that the simple correction algorithm [equation (38)] will result in identifying the unknown response $H(t)$.

The stability of the correction loop is ensured when the errors ϵ_j and ϵ'_j , before and after the corrections of the response $H(i)$, respectively, fulfill the inequality

$$||\epsilon'_j|| < ||\epsilon_j|| \quad (44)$$

The stability criterion has been derived under different limiting assumptions in References 7, 8, 10, 11, 12, and 14. Svoboda [18] and Pokorna [19] have presented an MSE convergence algorithm without any restriction on properties of the input signal vectors $X(j) = [x_{1+j}, x_{2+j}, \dots, x_{i+j}]$ and under a very weak condition. That is, from the set of products $P_i = h_i \cdot x_{i+j}$ used to compute the j th estimate, r_j [equation (15)], one of the highest absolute values, $|P_k| = |h_k \cdot x_{k+j}|$, is selected. The correction $\Delta h_{k,j}$ is then computed from

$$\Delta h_{k,j} = \epsilon_j \frac{x_{k+j}}{\sum_i (x_{i+j})^2} \quad (45)$$

and this value is used to update h_k :

$$h'_k = h_k + \Delta h_{k,j} \quad (46)$$

Note the similarities with equations (3) and (4), which were also derived in Reference 8 using the pseudo-inverse theory of non-square matrices. The correction reduces the error to

$$\epsilon'_j = \epsilon_j \left[1 - \frac{(x_{k+j})^2}{\sum_i (x_{i+j})^2} \right] \quad (47)$$

Then a second product with the next highest absolute value of P_k is chosen and the correction process is repeated with the new resulting values of ϵ'_j .

Since under all conditions

$$0 < \frac{(x_{i+j})^2}{\sum (x_{i+j})^2} \leq 1 \quad (48)$$

convergence is ensured. This algorithm also ensures the fastest convergence and minimizes the probability of "wrong" corrections [18]-[20].

The correction strategy defined by equations (34) and (37) is similar. Function Φ_{i+j} selects only the samples x_{i+j} with highest amplitudes. The correction step, which is proportional to the previous amplitude h_i , will be largest for the largest values of h_i and therefore the largest product $|P_i|$.

In echo cancellers, several hundred samples x_{i+j} are stored and only 10-30 participate in the computation of r_j . The square of the absolute value of the vector $\|X(j)\|^2 = \sum_i (x_{i+j})^2$, which is proportional to the total signal power entering the adaptive filter, changes relatively slowly and therefore can be approximated by the square of the average value \bar{X}_j^2 according to equation (36). Substituting from equations (34), (35), and (37) into equation (45) makes it possible to express the stability criterion as

$$\Delta h_{i,j} = \Delta_{i,j} |h_i| \leq \frac{\epsilon_j}{\bar{X}_j} \frac{x_{i+j}}{X_j} = \alpha \beta F(\epsilon_j) \Phi_{i+j} \quad (49)$$

The dimensionless constant multiplier $\Delta_{i,j}$ has a fixed value of $\pm 2^{-4}$; the value of α has been chosen as 1.4. The condition for correction is $F(\epsilon_j) \Phi_{i+j} = 1$; therefore, the coefficient β , which determines the sensitivity of the error detector [equation (35)], must be equal to

$$\beta \geq \frac{|h_i|}{1.4} 2^{-4} \quad (50)$$

to satisfy equation (49). This means that for optimum stable operation β should be adjusted according to the expected value of $|h_i|$. Although this adjustment is easy to implement, a compromise value for β which will satisfy equation (50) for an echo return loss of 14 dB can be used instead. This will cause a burst oscillation of the stored response $H(i)$ in circuits with ERL < 10 dB. However, at a frequency of one half the sampling rate (4 kHz), this oscillation causes no problem since the "reconstruction" filter high-frequency limit is 3.2 kHz. The influence of this instability on

response r_j is further attenuated because the fourth bit of mantissa m_h does not enter the convolution processor, and the adaptive loop tends to adjust the values of $H(i)$ to minimize the influence of "overcorrection." Therefore, the oscillation of $H(i)$ cannot be detected on the send output of the DEC and does not influence its performance.

For values of ERL > 20 dB, the probability, P_e , that the peaks of residual echo signal ϵ_j will be $> \beta X_j$, is $< 10^{-3}$. In this case no more than 16 correction steps are required to create the impulse response, and in this range of ERL the adaptive center clipper removes most of the residual echo signal. This heuristically explains the good performance of the DEC over the whole range of input signal levels (≈ 40 dB).

Double talk detector

In the equivalent diagram of the DEC (Figure 8), y is the echo signal as a result of a signal x ; s is an uncorrelated signal, e.g., the second party's speech signal plus noise of the tail circuit. If it is assumed that s , y , x , and r have zero means and are statistically stationary, which is true at least during a period of several syllables, the send-out signal is

$$z = s + y - r \quad (51)$$

Since s is uncorrelated with y and r , variance $E[z^2]$ is equal to the sum of average uncorrelated signal power levels

$$E[z^2] = E[s^2] + E[(y - r)^2] \quad (52)$$

The adaptive algorithm minimizes the square of errors ϵ_j , i.e., the variance $E[z^2]$, since the uncorrelated signal s remains constant. Thus, minimizing $E[z^2]$ minimizes the echo output power $E[(y - r)^2] \rightarrow 0$, theoretically without regard to the presence of uncorrelated speech signal s . However, this is true only when echo signal y is approximately equal to or larger than signal s , e.g., in circuits with very low ERL and low signal level s at the hybrid end of 2-wire sections.

However, under the double talk condition, when the send-out signal s is usually substantially larger than the echo y , the correction algorithm does not work properly because the minimum value of $E[z^2]$ remains relatively constant as a function of h_i . Hence, gradient $\nabla_j(\epsilon_j)$ of the error function is below the resolution of the error sensor (Figure 3), whose sensitivity is limited by the stability criterion [equation (50)] of the cross-correlation feedback loop. Under this condition, the correction estimate

according to equation (3) or (45) can lead to a solution which is different from the optimum (least mean square); *i.e.*, it can introduce distortion of the signal s [10].

The correction process must be inhibited immediately when a send-out signal s larger than echo y is detected because of the short convergence time, τ_{CN} . However, this fast convergence ensures correction of any distortion of the impulse response, usually during one syllable. A relatively simple double talk detector as shown in Figure 9 can give satisfactory results even when the echo path is unstable or has some phase roll.

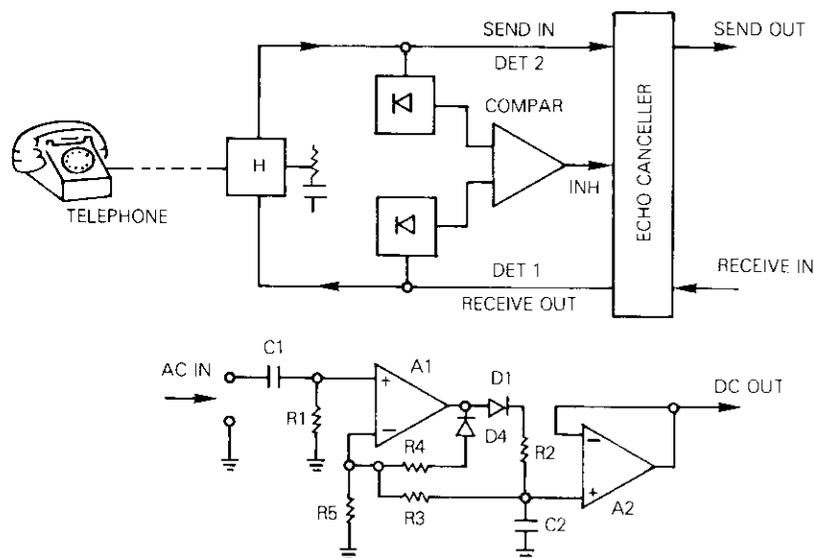


Figure 9. Block Diagram of the Double Talk Detector and the Peak Detector Circuit with Independent Adjustment of Rise Time and Release Time

Both receive-in and send-out rectified signals are compared in a voltage comparator. When the send-in signal is 0.5 dB stronger than the strongest possible echo signal at hybrid H , the output of the comparator inhibits the cross correlator's adaptation process.

Several design criteria must be met. The rise time of the inhibit signal must be substantially shorter than the convergence time to protect im-

pulse response $H(i)$ before it can be substantially distorted. The rise time ($\approx R_2 C_2$) of the output of peak detector DET1 must be longer than that of the DET2 output to give preference to the send-out signal under marginal conditions. The release time ($\approx R_3 C_2$) of the detector's output voltage must be between 50 and 200 ms, *i.e.*, long enough to hold the signal during short pauses, but short enough to permit "rebuilding" of the impulse response during the pauses of near-end speech.

Adaptive center clipper

The DEC described above can enhance the ERL by approximately 20 to 30 dB. With ERL in the range of 6–20 dB, only a very weak, mostly incoherent echo can be heard in the circuit during loud speech passages. To eliminate this residual echo signal, an adaptive center clipper (CC) has been designed for the send-out path of the DEC, as shown in Figure 4.

The nonlinear transfer characteristic of the CC permits the device's amplification to be greater than zero only when the instantaneous amplitude exceeds the threshold, d . If the threshold is proportional to the receive-in signal strength, the residual echo can be suppressed by another 8 to 50 dB in the whole dynamic range of the telephone signal without noticeably affecting the direct signal in the send-out path, which is usually much stronger than the residual echo. Tests have shown that the best results can be obtained if the clipping level $|d|$ is held approximately proportional to the rms value X_i of the speech integrated over a "window" of 50 to 150 ms with a "release" time constant of 100 to 240 ms, which depends on the round trip delay in the 4-wire section.

Only subjective tests can determine the effect of the center clipper distortion on speech, but an analysis based on the sine wave input can be used for design and evaluation. The distortion waveform (see Figure 10) approaches a symmetric square wave and the analysis is straightforward.

Theoretically, distortion is caused only by odd harmonics which are in the frequency band above 800 Hz, where the speech power is falling more than 12 dB/octave. The signal can therefore be pre-emphasized by 6 dB/octave before clipping and de-emphasized by 6 dB/octave after clipping (Figure 4) to attenuate the harmonics. The subjective test has shown that, for continuous speech with the same rms value as a 600-Hz sine wave used for calibration, a clipping level, $d \leq 0.2$ cannot be detected even by immediate comparison of the clipped and unclipped signals when pre-emphasis/de-emphasis is used.

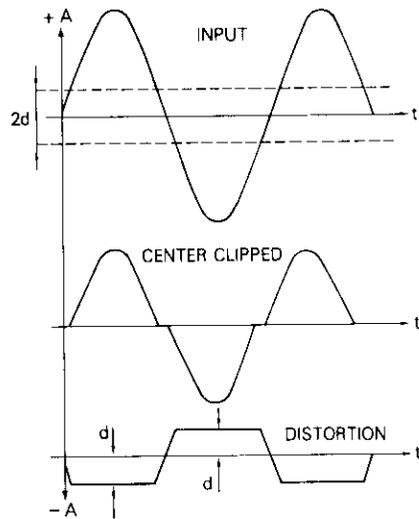


Figure 10. Effect of Center Clipping on a Sinusoidal Waveform

Measurements

The DEC underwent the following series of subjective and objective tests:

a. ERLE TEST. White noise, frequency limited by a 300-Hz to 3.2-kHz bandpass filter, was used as the receive-in signal. An attenuator and a variable delay line simulated the ERL and the delay, $2t_E$, in the 4-wire section. The output was measured by an AC voltmeter with C-message frequency weighting. First the send-out level was measured with the H-register cleared and center clipper disabled. Then the correlation processor was enabled and the output was again measured. The attenuation of the echo signal for different receive-in signal levels, with $ERL = 6$ dB and 3-ms end delay, is shown in Figure 11. For signal levels > -25 dBm0, ERLE is > 20 dB and averages around 22 dB. The center clipper suppresses the residual echo in this range below the system's noise level.

b. CROSS-CORRELATION CONVERGENCE TIME, τ_{CN} . The cross-correlation time convergence time was measured in the setup shown in Figure 12. A pulse generator was used to synchronize the scope and periodically clear and enable the H-register. The noise signal was observed on a storage scope to determine the approximate time, τ_{CN} , between the arrival of the enable pulse and the instant at which the noise ampli-

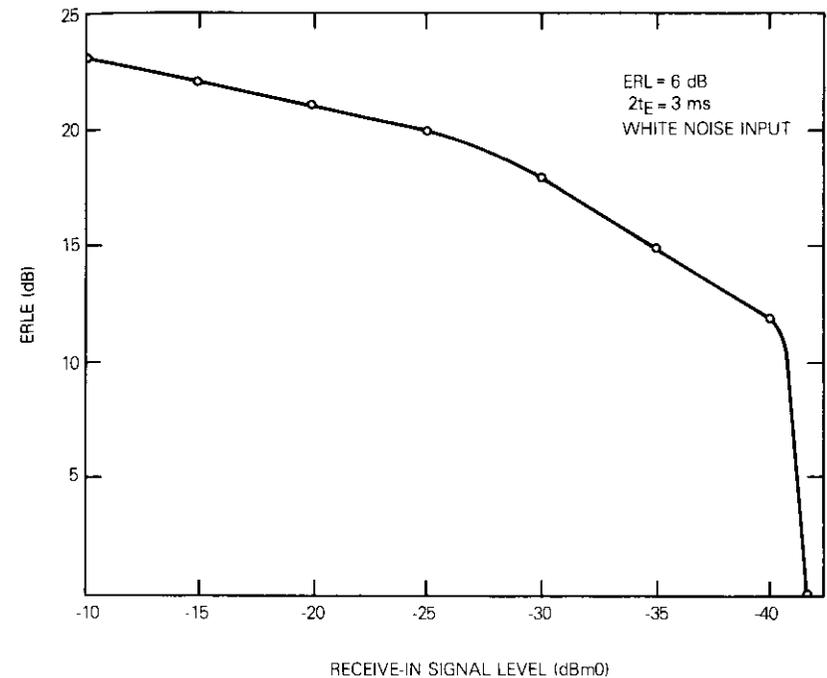


Figure 11. ERLE as a Function of the Receive-in Signal Level

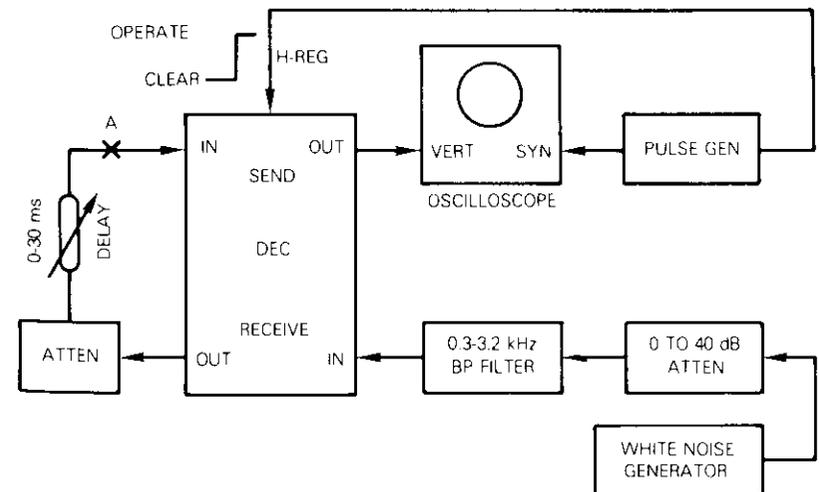


Figure 12. Test Circuit for Cross-Correlation Convergence Time (τ_{CN}) Measurement (phase shifting network is inserted at point A)

tude reached a point 10 percent above the final steady-state value. For $ERL = 6$ dB and a noise signal of -10 dBm0, $\tau_{CN} = 130$ ms with a disabled clipper and 70 ms with center clipping. With increasing ERL , τ_{CN} decreases; with a normal speech signal, the convergence time can always be expected to be shorter than 250 ms.

c. DEC PERFORMANCE WITH PHASE ROLL. This performance was measured with a phase roll unit inserted at point A (Figure 12). The results with a disabled center clipper are shown in Figure 13. With only 6-dB impairment in $ERLE$, the unit can work in a system with a phase roll, $\alpha(t)$, of 2π rad/s. With center clipping, the 6-dB impairment is not detectable for speech signals.

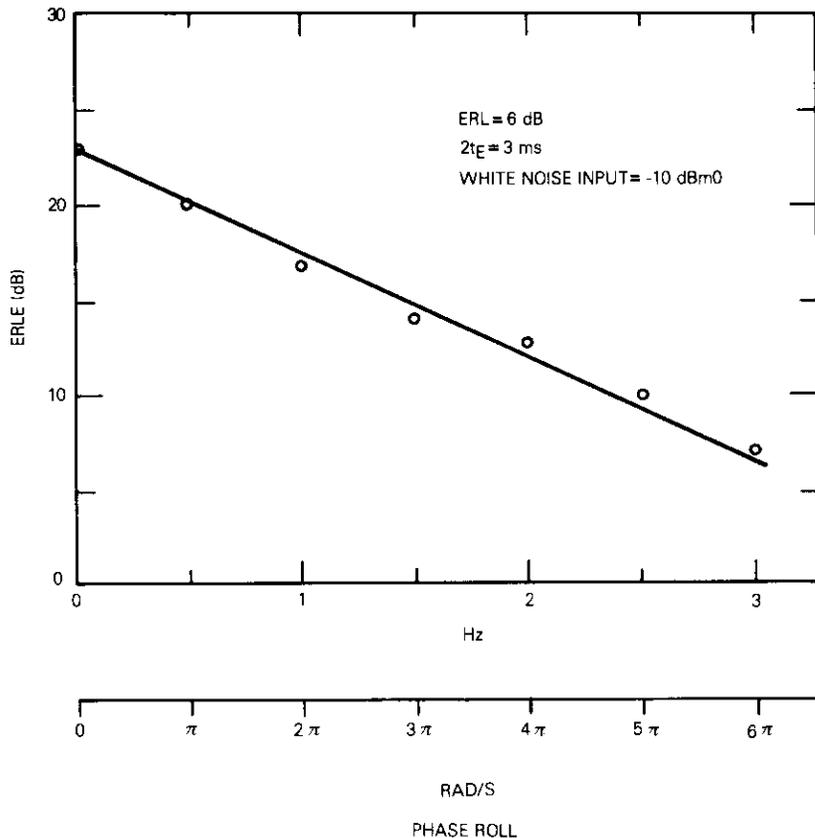


Figure 13. DEC Performance with Different Degrees of Phase Roll in the Echo circuit

d. SUBJECTIVE TESTS. These tests were conducted using a simulated satellite circuit in which the DEC was compared with two commercially available echo suppressors (A and B) [21]. Sixteen different listeners in two different experiments were asked to indicate their relative preference for the three echo controlling devices (DEC, A, and B). The results with statistical evaluations are shown in Table 1. The laboratory model DEC was considered superior to both echo suppressors with a confidence level of $1 - p(x, m) \geq 97.5$ percent, where $p(x, m)$ is the probability of x successes within m trials and each outcome has a probability of 50 percent.

DEC performance was also tested in a circuit used for data transmission at a 2400-bit/s rate [22] to verify that the device does not interfere with this use of the telephone circuit. As expected, no measurable changes in data transmission performance were detected.

TABLE 1. SUBJECTIVE COMPARISON OF THE DEC WITH TWO COMMERCIALY AVAILABLE ECHO SUPPRESSORS

Comparison	Preference Ratio	Confidence Level, $[1 - p(x)] \times 100\%$
DEC to Echo Suppressor A	22:10	97.9
DEC to Echo Suppressor B	28:4	99.999
Echo Suppressor A to Echo Suppressor B	22:10	97.5

Finally, the DEC was used very successfully in a limited experiment to prevent acoustical echo occurring during a phone conversation. Since the amount of end delay can be large even in a room of moderate size with peculiar acoustics, satisfactory performance requires a rather large amount of X and H storage.

Conclusions

The results described in this paper indicate that the use of a sophisticated convergence algorithm and logarithmic coding greatly improves the performance and reduces the complexity of digital adaptive echo cancellers. Furthermore, the economics of fabrication may well permit this canceller to compete with present-day echo suppressors even without introducing large-scale integration.

A greater theoretical understanding of the adaptive process and convergence criteria may lead to still faster convergence under adverse con-

ditions. Moreover, when time sharing can advantageously be applied, further economies will be realized, whereby echo cancellation will replace echo suppression in the vast number of cases in which it is desirable.

Addendum

Recent field tests performed by AT&T [23] have shown that the quality of satellite circuits equipped with COMSAT echo cancellers similar to those described in this paper equals that of domestic terrestrial circuits.

Acknowledgment

The author wishes to thank Dr. Campanella, H. G. Suyderhoud, and M. Onufry for their many valuable suggestions concerning the work and the manuscript, and P. R. Culley, who carried out most of the experiments, measurements, and implementations of the digital echo canceller.

References

- [1] L. B. Gogan, "Impedance and Return Loss Performance of Telephone Plant in Metropolitan Areas," *AIEE Transactions*, Vol. 77, Pt. 1, July 1958, pp. 257-261.
- [2] F. P. Duffy et al., "Echo Performance of Toll Telephone Connections in the United States," *Bell System Technical Journal*, Vol. 54, February 1975, pp. 209-243.
- [3] *Transmission Systems for Communications*, E. F. O'Neill, ed., Fourth Edition, Winston-Salem, N.C.: Bell Telephone Laboratories, 1970, Chapter 4.
- [4] H. R. Huntley, "Transmission Design of Intertoll Telephone Trunks," *Bell System Technical Journal*, Vol. 52, September 1953, pp. 1019-1036.
- [5] P. T. Brady and G. K. Helder, "Echo Suppressor Design in Telephone Communications," *Bell System Technical Journal*, Vol. 62, November 1963, pp. 2893-2917.
- [6] S. J. Campanella, M. Onufry, and H. G. Suyderhoud, "Digital Controlled Echo Suppressor," U.S. Patent 3,979,701, September 7, 1976.
- [7] M. M. Sondhi, "An Adaptive Echo Canceller," *Bell System Technical Journal*, Vol. 46, March 1967, pp. 497-511.
- [8] S. J. Campanella, H. G. Suyderhoud, and M. Onufry, "Analysis of an Adaptive Impulse Response Echo Canceller," *COMSAT Technical Review*, Vol. 2, No. 1, Spring 1972, pp. 1-36.
- [9] H. G. Suyderhoud, S. Campanella, and M. Onufry, "Results and Analysis

- of Worldwide Echo Canceller Field Trial," *COMSAT Technical Review*, Vol. 5, No. 2, Fall 1975, pp. 253-273.
- [10] B. Widrow et al., "Adaptive Noise Cancelling," *Proc. IEEE*, Vol. 63, No. 12, December 1975, pp. 1692-1716.
- [11] N. Demytko and L. Mackechnie, "A High Speed Digital Adaptive Echo Canceller," *ART*, Vol. 7, No. 1, May 1973, pp. 20-28.
- [12] N. Demytko et al., "Study of Echo Canceller Under the Conditions of Phase Roll," INTELSAT Contract IS-668, Phase II.b Telecom Australia, November 1975.
- [13] G. D. Dill, Personal Correspondence
- [14] M. M. Sondhi and D. Mitra, "New Results on the Performance of a Well-Known Class of Adaptive Filters," *Proc. IEEE*, Vol. 64, No. 11, November 1976, pp. 1586-1597.
- [15] K. Virupaksha, Personal Correspondence
- [16] H. Berndt, "Correlation Function Estimation by a Polarity Method Using Stochastic Reference Signals," *IEEE Transactions on Information Theory*, IT-14, November 1968, pp. 796-801.
- [17] J. L. Flanagan, *Speech Analysis, Synthesis, and Perception*, New York: Academic Press, 1965.
- [18] A. Svoboda, "Linear Analyser in Czechoslovakia" (in Czech), *Czechoslovak Journal for Physics*, Vol. 1, No. 10, October 1951.
- [19] O. Pokorna, "Solution of a System of Algebraic Equations by Minimization of Squares of Residues" (in Czech), *Information Processing Machines*, Prague: Academia, 1954, Vol. 2, pp. 111-116.
- [20] V. Cerny and A. Marek, "Application of the Method of Minimization of the Sum of Absolute Values . . ." (in Czech), *Information Processing Machines*, Prague: Academia, 1958, Vol. VI, pp. 209-225.
- [21] M. Onufry, Personal Correspondence
- [22] Rixon, Inc., Sangamo, Codec Mark DS-2401, Personal Correspondence.
- [23] G. K. Helder and P. C. Lopiparo, "Improving Transmission on Domestic Satellite Circuits," *Bell Laboratories Record*, Vol. 55, No. 8, October 1977, pp. 202-207.

Appendix A. Convolution processor

The function of the convolution processor (Figure 5) is shown schematically in Figure A-1. The n samples h_i of the echo response $H(t)$ of the Dirac impulse $\delta(t)$ are stored in the H -register. The X -registers store samples x_{i+j} of signal waveform $X(t)$. The number of stored samples, n , must be greater than

$$n > \frac{2t_E}{T_S} \quad (A1)$$

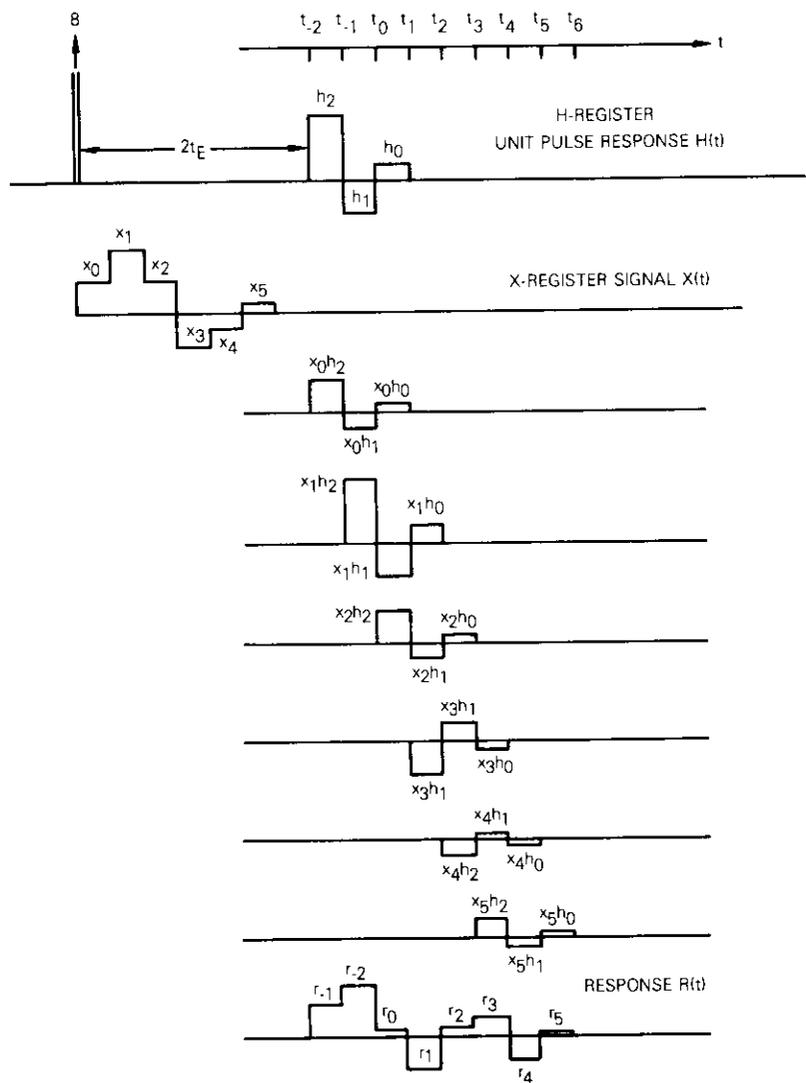


Figure A-1. Schematic Presentation of Waveforms \$X(t)\$ and \$H(t)\$ Entering the Convolution Processor and the Formation of Resulting Response \$R(t)\$ as the Sum of Partial Responses \$x_i H(t)\$

where \$t_E\$ is the so-called flat delay and \$T_S\$ is the sampling frequency. Because every sample \$x_{i+j}\$ can be considered to be a Dirac pulse with amplitude \$x_{i+j}\$ at time \$(i + j)T_S\$, the circuit response to every \$x_{i+j}\$ is the product \$x_{i+j} \times H(t)\$, as shown in Figure A-1 for samples \$x_0\$ to \$x_6\$. At every instant \$t_j\$, according to the superposition theorem, response \$r_j\$ is the sum of the responses to individual samples \$x_{i+j}\$:

$$r_j = \sum_{i=0}^m h_i \times x_{i+j} = H(t) * X(t) \quad (A2)$$

For example, at time \$t_1\$,

$$r_1 = x_3 h_2 + x_2 h_1 + x_1 h_0 \quad (A3)$$

The sequence of samples \$r_j\$ represents the resulting response \$R(t)\$, i.e., the simulated echo of signal \$X(t)\$ in the circuit with echo impulse response \$H(t)\$. Functions \$H(t)\$, \$X(t)\$, and \$R(t)\$ in Figure A-1 are typical of those which can actually be on an oscilloscope.

For an adaptive filter design, the most important parameter is the number, \$m\$, of samples \$h_i\$ necessary to emulate the response \$R(t)\$ with the required accuracy. The following is an estimate made under the assumption that the echo path approaches the ideal bandpass filter (320 Hz to 3.2 kHz) with linear phase characteristic and with frequency characteristic \$A(\omega)\$ as follows:

$$A(\omega) = \begin{cases} 0, & \omega_2 > \omega > \omega_1 \\ 1, & \omega_2 < \omega < \omega_1 \end{cases} \quad (A4)$$

For analysis, it is convenient to consider this characteristic as a difference of two ideal low-pass filters, one with maximum frequency \$\omega_m = \omega_1\$, and the other with \$\omega_m = \omega_2 < \omega_1\$. The impulse response of an ideal low-pass filter is

$$H(t) = \frac{\omega_m}{\pi} \frac{\sin \omega_m t'}{\omega_m t'} \quad (A5)$$

where by definition \$t' = 0\$ for the maximum peak \$h_0 = \omega_1/\pi\$. The response has local extremes at \$|\omega_m t'| = (1 + 2k) \pi/2\$, where \$k\$ is an integer \$k \ge 1\$ with amplitudes

$$|h_k| = \frac{h_0}{\pi} \frac{2}{1 + 2k} \quad (A6)$$

That is, the relative amplitude of peaks \$|h_k|/h_0\$ is a monotonic function of normalized time \$\omega_1 t'\$

$$\frac{|h_k|}{h_o} = \pm \frac{1}{\omega_m t'} \rightarrow 2t' = \frac{2}{\omega_m} \times \frac{h_o}{|h_k|} \quad (A7)$$

If this response is sampled during the interval $2t'$, the number of samples m is

$$m \doteq 2t' f_s + 1 \doteq \frac{f_s}{\pi f_m} \times \frac{h_o}{|h_k|} + 1 \quad (A8)$$

For echo cancellation, the accuracy of the echo impulse response emulation is ≤ 0.05 ; therefore, it appears to be sufficient to retain only the portion of response $H(t)$ with peaks $h_m \leq h_o/40$. If $f_m = f_1 = 3.2 \times 10^3$ and the sampling rate is $f_s = 8$ kHz, the number of samples according to equation (A-8) is $m \geq 33$. This agrees surprisingly well with COMSAT experiments and measurements on actual and simulated telephone circuits in which the echo path characteristics differ substantially from the ideal bandpass mainly because of the nonlinear phase characteristic. Only after the number of h_i samples was reduced below $m < 25$ did the echo cancellation show measurable change. Since the low-frequency limit of the echo circuit is $\omega_2 \doteq \omega_1/10$, the maximum peak h'_o of the transient response is $h'_o = h_o/10$, but of the opposite sign. Therefore, only the portion of the response with $h'_m \leq h'_o/4$ must be preserved; i.e., $m \geq 33$ also satisfies the requirement of the low-frequency limit of the filter.

The number of samples, m , needed to simulate the echo response $R(t)$ always occupies only a small fraction of the whole memory because the number, n , of stored samples h_i is usually $n \geq 250$ to cover the whole range of t_B .

Appendix B. A-law encoding

A 7-bit 13-segment pseudo-logarithmic code, which is similar to the A-law code used for encoding the speech signal, has been adopted as the internal code for the present DEC. As shown in Figure B-1, the signal-to-quantization noise ratio, S/N_q , for this code is > 28 dB for the speech power range of -10 to -40 dB. The relationship between the 11-bit linear format and the 7-bit code is given in Table B-1 [B-1] from which properties of the code can be derived. In every segment with $e > 0$, only the $1 + b$ most significant bits can be determined from the code; the resolution is therefore equivalent to a linear code with

$$B_{SN} = Sg + 1 + b = 2 + b \text{ bits} \quad (B1)$$

where b is the number of bits in the mantissa, which is by definition $0 \leq m \leq 1$. Therefore, for the case described herein, $B_{SN} = 5$. According to Table B-1, the dynamic range is equivalent to that of a linear code with

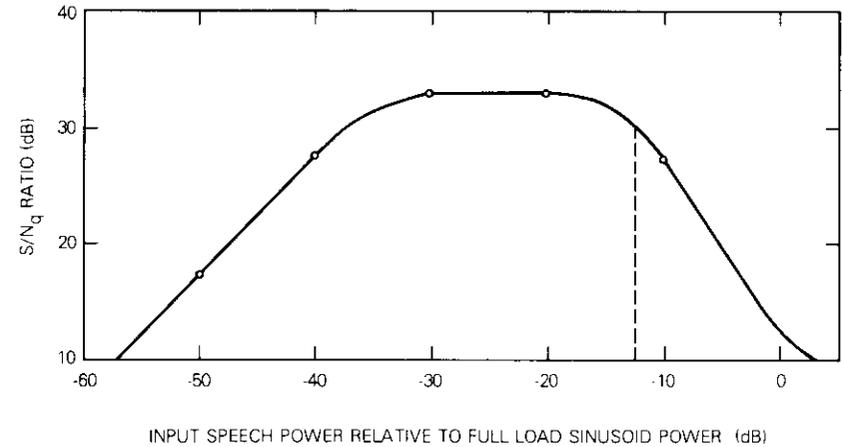


Figure B-1. S/N_q Performance of the 13-Segment 7-bit A Law Code with Speech Signal

$$B_L = 1 + e + b \text{ bits} \quad (B2)$$

i.e., $B_L = 11$ for this code with the maximum exponent $e = 7$.

The log-to-lin conversion algorithm is given by equations (6) and (8); the unknown digits (X in Table B-1) can be filled in by a pattern ϵ_M of 0's and 1's

TABLE B-1. RELATIONSHIP BETWEEN THE 11-BIT LINEAR CODE AND A-LAW FORMAT (x digits in the linear code are ignored when transforming in the A-law format)

MSB	Linear Code		A-Law Code		Relative Quant. Step	Segment No.	
	MSB	LSB	MSB	Exp Mantissa			
SG	0 0 0 0 0 0 0	a b c	SG	0 0 0	a b c	Δx	L
SG	0 0 0 0 0 0 1	a b c	SG	0 0 1	a b c	Δx	1
SG	0 0 0 0 0 1	a b c x	SG	0 1 0	a b c	$2\Delta x$	2
SG	0 0 0 0 1	a b c x x	SG	0 1 1	a b c	$4\Delta x$	3
SG	0 0 0 1	a b c x x x	SG	1 0 0	a b c	$8\Delta x$	4
SG	0 0 1	a b c x x x x	SG	1 0 1	a b c	$16\Delta x$	5
SG	0 1	a b c x x x x x	SG	1 1 0	a b c	$32\Delta x$	6
SG	1	a b c x x x x x x	SG	1 1 1	a b c	$64\Delta x$	7

according to equation (28). The code used for impulse response $H(i)$ has similar properties, but its maximum S/N_q ratio is 38.9 dB.

Reference

- [B-1] *Transmission Systems for Communications*, E. F. O'Neill, ed., Fourth Edition, Winston-Salem, N.C.: Bell Telephone Laboratories, 1970, Chapter 4.

Appendix C. Cross-correlator time constants

Two indirectly related time constants can be considered to be figures of merit of the correction process, *i.e.*, the convergence time, τ_{CN} , and the correction time constant, τ_{CR} .

The convergence time is by definition [C-1], [C-2] the time necessary to build up pulse response $H(i)$ so that ERLE reaches a value which is 1 dB below its final state. Since this also depends on input signal amplitudes and its statistical distribution, the true value of τ_{CN} can be established only experimentally. However, the number of correction steps, C_r , which are needed to generate the same response $H(i)$ can be used to compare the relative speed of different correction algorithms. To cover the whole dynamic range from zero to the maximum possible value of $|h_i|$ by adding one LSB every convolution cycle, the number of steps is

$$C_r = 2^B \quad (C1)$$

where B is the number of bits in the code (without the sign bit). The dynamic range of the pseudo-logarithmic code is equivalent to B_L bits (see Appendix B):

$$B_L = e + b \quad (C2)$$

where $e = 6$ and $b = 4$ for the format used to encode the response $H(i)$, *i.e.*, $B_L = 10$. The number of steps for the linear code is $C_{ra} = 2^{10}$; the number of steps for the equivalent 8-bit logarithmic code (Sg + 7 bits) is $C_{rb} = 2^7$; and the ratio is

$$\frac{C_{rb}}{C_{ra}} = \frac{2^7}{2^{10}} = \frac{1}{8} \quad (C3)$$

Therefore, for the pseudo-logarithmic code the convergence time for ERL < 10 dB can be expected to be eight times shorter than that with the equivalent linear code under the same condition.

The correction time constant, τ_{CR} , defined as the time needed to reduce the amplitude of sample h_i to $1/e$ of its original value, indicates the speed with which the response $H(i)$ can follow the changing circuit parameters, *e.g.*, with phase roll. Since the value $1/e$ is approximately

$$\frac{1}{e} = 1 - 5 \times 2^{-3} \quad (C4)$$

$5 \times 2^{b-3}$ correction steps $\Delta h_{i,j}$ equivalent to the least significant bit of mantissa m_n [equation (32)] are needed to reduce the amplitude of $|h_i|$ by 63 percent. If T_S is the sampling period and p_e is the probability of $F(\epsilon_j) = \Phi_{i+j} = 1$, then the time constant is

$$\tau_{CR} = 5 \times 2^{b-3} \frac{T_S}{p_e} \quad (C5)$$

For $b = 4$, $T_S = 125 \mu s$, and $0.01 < p_e < 0.1$ under near-steady-state conditions, the constant is $12.5 < \tau_{CR} < 125 ms$, which explains the relative immunity of the echo canceller to phase roll (see Figure 13).

References

- [C-1] M. M. Sondhi, "An Adaptive Echo Canceller," *Bell System Technical Journal*, Vol. 46, March 1967, pp. 497-511.
 [C-2] S. J. Campanella, H. G. Suyderhoud, and M. Onufry, "Analysis of an Adaptive Impulse Response Echo Canceller," *COMSAT Technical Review*, Vol. 2, No. 1, Spring 1972, pp. 1-36.



Otakar A. Horna (A'49-M'56-SM'69) studied at various European Universities and received an M.S. degree in electrical engineering from the Czech Institute of Technology, a Ph.D. in electronics from the Institute of Radio Technique and Electronics, and a State Diploma in Mathematical Logic from Charles University in Prague.

He was Senior Scientist and Secretary of the Scientific Council of the Research Institute for Mathematical Machines in Prague, in charge of the Computer Technology Division. Later he became Senior Engineer with Multronics Corporation, Rockville, Md., where he worked in military electronics. He joined COMSAT Laboratories, Clarks-

burg, Md., in 1969 and is presently a member of the technical staff of the Signal Processing Branch.

He is the author of five books, more than 30 U.S. and foreign patents, and numerous scientific and technical articles. His work was awarded a Gold Medal at the Brussels World Fair, in 1958, and a Gold Medal at the Brno Fair in 1967. Recently he received the first COMSAT Research Award for his echo canceller design. He is a Senior Member of IEEE.

Index: differential pulse code modulation, predictor

DPCM prediction for NTSC composite signals

I. DINSTEIN

(Manuscript received April 20, 1977)

Abstract

Two 2-dimensional DPCM predictors for NTSC composite signals have been proposed and evaluated. These predictors can use any sampling frequency which is a rational multiple of the color subcarrier frequency. The proposed predictors and three other predictors (previous third element, H-O-DPCM, and a predictor using three samples of the previous line) have been simulated and applied to a set of five digitized pictures. The parameters computed for evaluating and comparing these predictors are the entropy, variance, and highest absolute values of the differences between the predictors and the corresponding video values, and the correlation between the predictions and the corresponding video values. The proposed 2-D-II predictor was found to be the best of the five predictors.

Introduction

The selection of a predictor is an important factor in the design of a DPCM encoder for NTSC color composite signals. It has been shown [1] that the predictor yielding the minimum mean square error (for any signal) is a function of the autocorrelation of the signal. Because of the limitations of presently available hardware, such an approach is impractical for real-time encoding of television signals. Most DPCM predictors for composite signals require certain phase relationships between the samples and the

color subcarrier [2], [3]; therefore, they are dependent on the sampling rate. For NTSC composite video, a sampling rate of 10.7 MHz (three times the color subcarrier frequency) is most commonly used because it is the smallest integral multiple of the subcarrier frequency which is above the required Nyquist rate (8.4 MHz for NTSC signals). Almost all the proposed DPCM predictors for NTSC composite signals are limited to data sampled at 10.7 MHz. However, the 10.7-MHz rate is not very efficient, since it is about 20 percent higher than the required 8.4-MHz rate.

Two 2-dimensional DPCM predictors for NTSC composite signals have been proposed and evaluated.* These predictors, which may be used in conjunction with any sampling frequency of the form $(n/m) \times f_{sc}$, where n and m are integers, n is even, and f_{sc} is the color subcarrier frequency, require the phase of the samples to be adjusted between lines. A procedure for performing this adjustment at any sampling rate is outlined in Appendix A.

The proposed predictors were compared with three other DPCM predictors for NTSC composite signals. These predictors, which are described in the following section, were simulated on a PDP 11/10 computer interfaced with an experimental DPCM encoder for TV signals currently being developed at COMSAT Labs. The subsequent sections describe the data acquisition, the parameters and measures used for evaluating the performance of the predictors, and the evaluation results. Of the five predictors, the proposed 2-D-II predictor was found to be the best.

DPCM predictors for NTSC composite signals

Five predictors, defined in the following subsections, were considered in the study reported herein. For practical reasons, the number of previous pels (picture elements) used for each prediction was limited to three. Some of the predictors require phase adjustment of the samples between lines to vertically align the samples, or to minimize the distance between samples of consecutive lines having the same subcarrier phase.

2-D-I predictor

The first 2-dimensional predictor, designated as 2-D-I, requires the samples to be vertically aligned [5]. With reference to Figure 1, y_i is the luminance at pel i , and c_i and α_i are, respectively, the amplitude and

*One of the proposed predictors (2-D-I) was independently proposed in Reference 4.

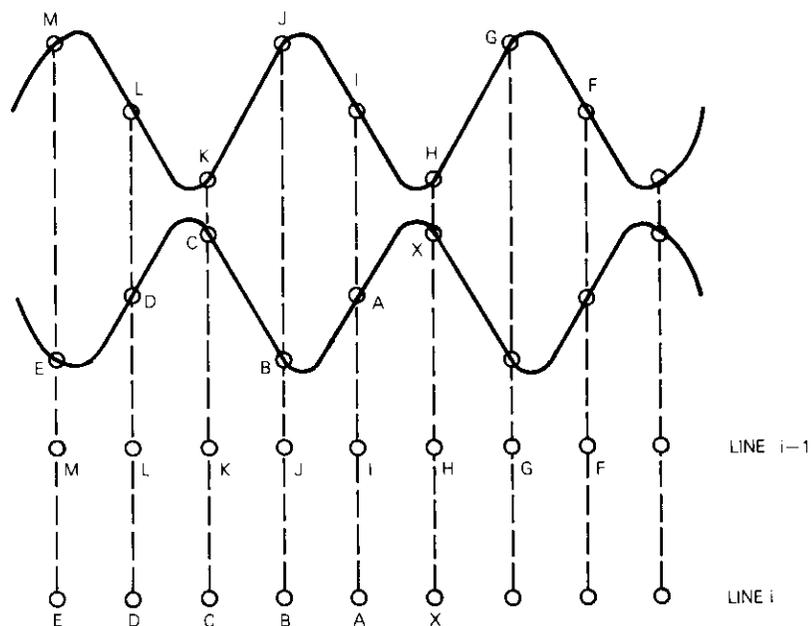


Figure 1. Spatial Distribution and Phase Relationship Between Vertically Aligned Pels (10.7-MHz sampling rate)

phase of the color subcarrier at pel i , where $i = X, A, B, C, \dots, M$. The composite signal at pel i is $y_i + c_i \sin \alpha_i$, and the prediction of pel X is denoted as \hat{X} . The 2-dimensional predictor proposed herein is defined as follows;

$$\hat{X} = A + I - H.$$

When samples of NTSC color composite signals are vertically aligned, the color subcarrier phase difference between any two vertically aligned pels is 180° . The bandwidths of the I and Q chrominance components of the NTSC color composite signals are 1.2 and 0.6 MHz, respectively. For data sampled at rates higher than 8.4 MHz (the Nyquist rate for NTSC composite signals), it is reasonable to assume that changes in the chrominance of adjacent pels are very small. If it is assumed that the chrominance of pels X, A, H , and I is the same, then

$$c_x = c_A = c_H = c_I$$

$$\alpha_x = -\alpha_H = \alpha$$

$$\alpha_A = -\alpha_I = \alpha - \psi$$

where ψ is the subcarrier phase difference between pels X and A . Hence,

$$\hat{X} = A + I - H = y_A + y_I - y_H + c_x \sin \alpha$$

This predictor is attractive because it is not restricted to the sampling rate of three times the subcarrier frequency. Any sampling frequency of the form $n/m \times f_{sc}$, where n and m are integers, n is even, and f_{sc} is the frequency of the color subcarrier, can be used.

2-D-II predictor

The second 2-dimensional predictor, designated as 2-D-II, requires the phase relationship between samples of consecutive lines shown in Figure 2. The required phase relationship, which is a function of the sampling frequency, can be obtained for any sampling frequency of the form $n/m \times f_{sc}$ by adjusting the phase of the samples at the beginning of each line (see Appendix A).

Predictor 2-D-II is defined as

$$\hat{X} = H + \frac{1}{2}(A - I)$$

where \hat{X} is the prediction for pel x , and A , H , and I are previous pels, as shown in Figure 2. With the notation defined in the previous subsection, it is assumed that

$$c_x = c_H \quad c_A = c_I$$

$$\alpha_x = \alpha_H \quad \alpha_A = \alpha_I$$

Then

$$\hat{X} = Y_H + c_H \sin \alpha_H + \frac{1}{2}(y_A - y_I)$$

Pel X is predicted by pel H with a correction due to the slope between pels A and I . Note that the distance between pel X and pel H is not a function of the sampling frequency. Instead, it consists of the distance between lines in the vertical direction and the distance equivalent to half a subcarrier period in the horizontal direction. The performance of this predictor is therefore expected to degrade very little as the sampling frequency is decreased.

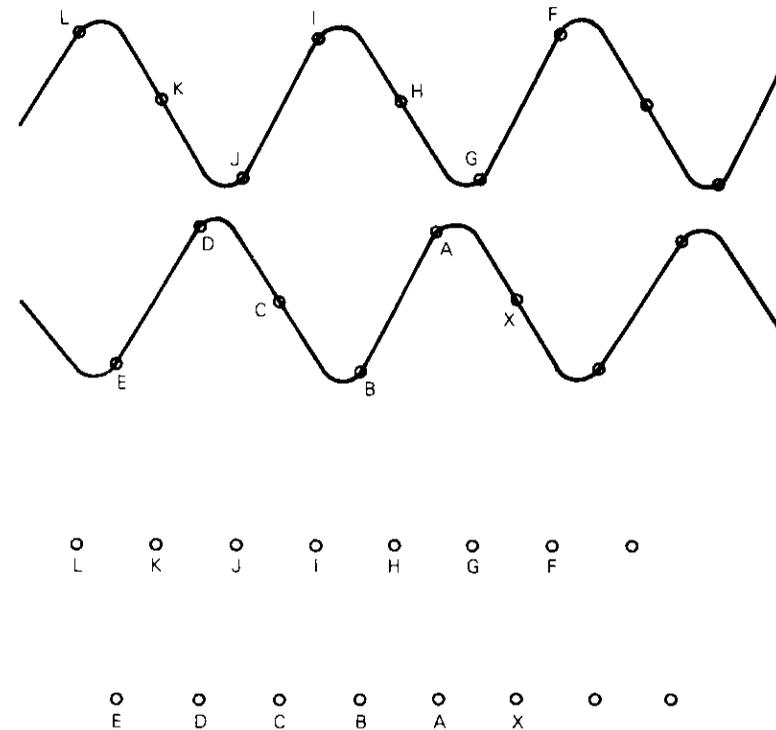


Figure 2. Spatial Distribution and Phase Relationship Between Samples (10.7-MHz sampling rate, samples X and H have the same subcarrier phase)

Third element predictor

The third element predictor, which is the simplest predictor for composite signals, is restricted to a sampling rate of $3 \times f_{sc}$ (about 10.7 MHz). It is defined as

$$\hat{X} = C$$

where pels X and C are as shown in Figure 1. If it is assumed that there is no change in the chrominance between pels X and C , both have the same subcarrier phase and amplitude.

H-O-DPCM predictor

The H-O-DPCM predictor, developed at NEC [6], is part of the H-O-DPCM (high-order DPCM) predictor family. It is defined as

$$\hat{X} = C + \frac{1}{2}(A - D)$$

where pels X , C , A , and D are as shown in Figure 1. The predictor given here is restricted to data sampled at 10.7 MHz. Intuitively, this predictor is similar to the third-element predictor, but it also incorporates a correction for the slope in the luminance. Other members of the H-O-DPCM predictor family allow different sampling rates, but require four previous samples to calculate the prediction and are not considered in this paper.

Jaeger predictor

The Jaeger predictor [7] is restricted to a sampling rate of 10.7 MHz and requires the samples to be vertically aligned. It is defined as

$$\hat{X} = \frac{2}{3}(I) - \frac{1}{3}(H) + \frac{2}{3}(G)$$

where pels X , I , H , and G are as shown in Figure 1. With Y_i , C_i , and α_i as defined before,

$$\begin{aligned} c_I &= c_H = c_G \\ \alpha_I &= \alpha_H - 120^\circ \\ \alpha_G &= \alpha_H + 120^\circ \\ \alpha_X &= -\alpha_H = \alpha \end{aligned}$$

With some trigonometric manipulation, it can be shown that

$$\hat{X} = \frac{1}{3}(2Y_I - Y_H + 2Y_G) + c_H \sin \alpha$$

A disadvantage of this predictor is the complexity of multiplication by a weight of 1/3.

These predictors were simulated in computer programs implemented on a PDP 11/10 computer interfaced to an experimental digital television system. Data acquisition for the experiment is described in the following section; evaluation and comparison results will be summarized in subsequent sections.

Data acquisition

Figure 3 is a simplified block diagram of an experimental DPCM encoder for television signals which is being developed at COMSAT Labs. The

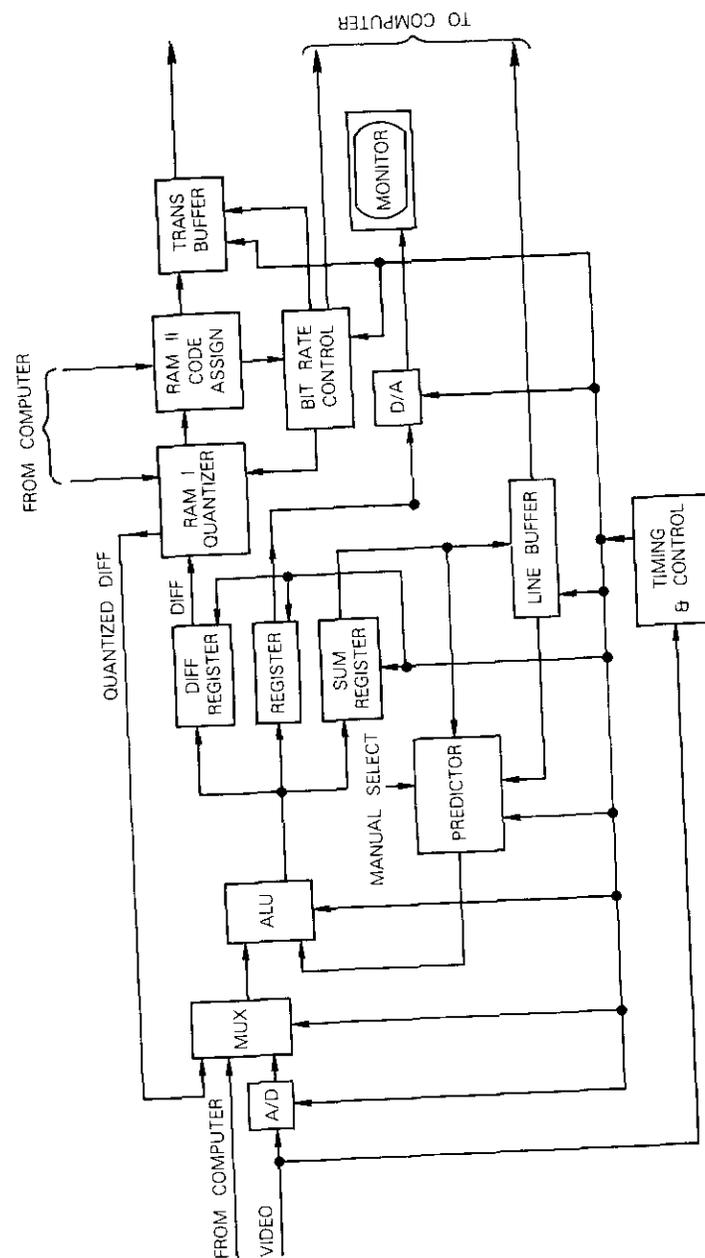


Figure 3. Experimental DPCM Encoder for Television Block Diagram

differences are quantized and the codes are assigned by means of look-up tables stored in random access memories (RAMs). The RAMs are loaded by a PDP 11/10 computer interfaced to the system.

All the timing signals are derived from a clock which is phase locked to the color subcarrier. The frequency of this clock is $n \times f_{sc}$, and the sampling frequency is $n/m \times f_{sc}$. The limits for n and m are $n \leq 24$ and $m \geq 8$. The highest possible sampling frequency is $24/8 \times f_{sc} = 10.7$ MHz, which is three times the frequency of the color subcarrier. For any selected frequency, the phase of the samples can be adjusted with respect to the color subcarrier as described in Appendix A. The predictor contains the logic required to enable manual selection of three previous pels (from the current and/or previous line) to be used for the prediction. Still pictures may be transferred to the computer two lines at a time.

Five slides, shown in Figure 4, were digitized, transferred to the computer, and stored on a magnetic tape. Three sampling rates were used:

- a. $24/8 \times f_{sc} = 10.7$ MHz,
- b. $24/9 \times f_{sc} = 9.54$ MHz,
- c. $24/10 \times f_{sc} = 8.59$ MHz.

For all the digitized slides, the samples were vertically aligned or phase adjusted as required by the predictors. (Phase adjustment of the samples is required by some of the predictors and does not affect the others.) Only the active portion of each slide, which included 130,500, 123,750, and 112,500 pels for 10.7-, 9.54-, and 8.59-MHz sampling rates, respectively, was used for each slide. The slides were digitized using eight bits per sample. The video signal (including the synchronization pulses portion) was set to cover the entire range of the analog-to-digital converter. Results of the experiment are summarized in the next two sections.

Predictor evaluation measures

The following sections describe the measures computed for the predictor evaluation.

Entropy of differences

For any given picture and predictor, the entropy of the differences is computed as follows. Assume that $PEL(i, j)$ is the j th pel in the i th line, $V(i, j)$ is the digitized video value of $PEL(i, j)$, $P(i, j)$ is the prediction of



Figure 4. Digitized Slides Used in the Experiment (10.7-, 9.54-, and 8.59-MHz sampling rates)

$PEL(i, j)$, and M is the total number of pels in the picture being considered. Define $N(d)$ as the number of pels in the picture for which

$$V(i, j) - P(i, j) = d \quad .$$

Note that, when the picture is digitized with eight bits, the range of the video and the prediction is (0,255), and the range of the differences is (-255, +255).

The value $N(d)/M$ is the relative frequency of occurrence of the difference, d . The entropy defined as [8]

$$\sum_{d=-255}^{+255} \frac{N(d)}{M} \times \log_2 \frac{N(d)}{M}$$

indicates the minimum number of bits required for the transmission of the coded differences.

Variance of differences

The variance was computed according to

$$\frac{1}{M} \sum_i \sum_j [V(i, j) - P(i, j)]^2$$

where M , $V(i, j)$, and $P(i, j)$ are as defined previously. The differences have zero mean; therefore, the mean does not appear in the preceding expression. Again, the variance is an indication of the expected quantizing noise; that is, an increase in the variance indicates increasing quantizing noise.

Correlation between prediction and video values

The correlation between the prediction and the corresponding video values was computed as follows:

$$\text{correlation} = \frac{\sum_i \sum_j [V(i, j) - \mu_V][P(i, j) - \mu_P]}{\left(\sum_i \sum_j [V(i, j) - \mu_V]^2\right)^{1/2} \left(\sum_i \sum_j [P(i, j) - \mu_P]^2\right)^{1/2}}$$

where $V(i, j)$ and $P(i, j)$ are as defined in the previous subsection, and μ_V and μ_P are the mean values of the video and the predictions, respectively.

The correlation is 1 when $P(i, j) = V(i, j)$ for all (i, j) , and less than 1 otherwise.

Peak differential values

The peak differential values are differences in the range $\pm(d_p, 255)$, where d_p is the smallest absolute difference which satisfies

$$\frac{1}{M} \sum_{d=-d_p}^{d_p} N(d) \geq 0.99$$

and $N(d)$ and M are as defined previously. For any given picture, 99 percent of the absolute values of the differences are smaller than the corresponding d_p . Since the differences are being quantized, the value of d_p is an indicator of the quantizing noise; that is, an increase in d_p indicates increasing quantizing noise.

Results

Figure 5 shows the entropies computed for the various predictors and pictures. With the exception of one picture (TOYS), the 2-D-II predictor yielded differences with the lowest entropy when applied to data sampled at 10.7 and 9.5 MHz. The H-O-DPCM predictor yielded entropies lower than those obtained with the 2-D-I predictor. This result is somewhat surprising, since 2-dimensional predictors would be expected to perform better than 1-dimensional predictors. This poor performance is attributed to the effect of changes in the luminance on the 2-D-I predictor. Instead of correcting the prediction according to the slope in the luminance signal (similar to the 2-D-II and the H-O-DPCM predictors), the 2-D-I amplifies such changes, thus yielding higher differences between predictions and actual values.

The theoretical minimum bit rate required for transmission of coded differences can be estimated from the product of the entropy of the differences and the sampling rate. The estimated minimum bit rates given in Figure 6 decrease with decreasing sampling rate in an obvious pattern. This indicates that decreasing the sampling rate by a certain factor causes the respective entropy to be increased by a smaller factor.

Figure 7 shows the correlations between the predictions of the various predictors and the respective video values. It can be seen that the 2-D-II predictor yielded predictions with the highest correlation with the respec-

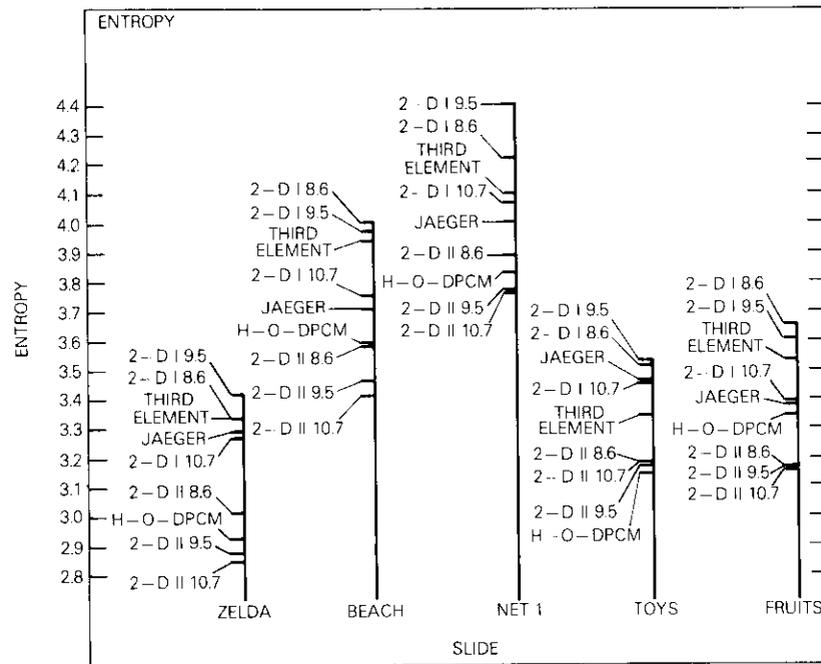


Figure 5. Entropies of Differences

tive video value. As shown in Figure 8, the 2-D-II predictor also yielded differences with the smallest variances.

As defined previously, peak differential values are differences within the range $\pm(d_p, 255)$, where d_p is the smallest absolute difference satisfying

$$\frac{1}{M} \sum_{d=-d_p}^{d_p} N(d) \geq 0.99$$

and M and $N(d)$ are as defined above. Values of d_p for the various pictures and predictors are given in Figure 9. Here, again, the 2-D II predictor yielded the smallest peak difference values.

Conclusions

Two new 2-dimensional DPCM predictors for NTSC TV composite signals were proposed and evaluated together with three predictors previously proposed in the literature. The measures used for the evaluations were the

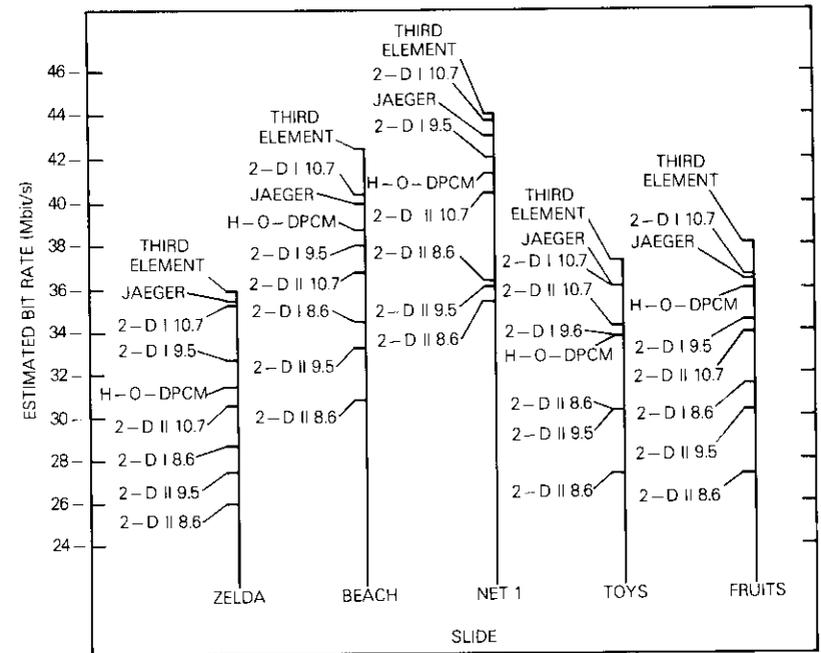


Figure 6. Estimated Minimum Bit Rate (product of entropy and sampling rate)

entropy, variance, and peak value of the differences between predictions and corresponding video values, and the correlation between the predictions and the corresponding video values. Average values of the results are given in Figure 10. Based on these measures, the proposed 2-D-II predictor was found to be the best of the five predictors tested. This predictor is applicable to NTSC composite signals sampled at any frequency of the form $n/m \times f_{sc}$, where n and m are integers, n is even, and f_{sc} is the color subcarrier frequency. Results show that the 2-D-II predictor maintains good performance even when the sampling frequency is decreased to 8.6 MHz.

The lower bound for the sampling frequency depends on the ability to filter out aliasing frequencies. This problem should be studied carefully. The use of lower sampling rates permits either a decrease in the required transmission bit rate for a given picture quality, or an improvement in the picture quality for a given transmission bit rate.

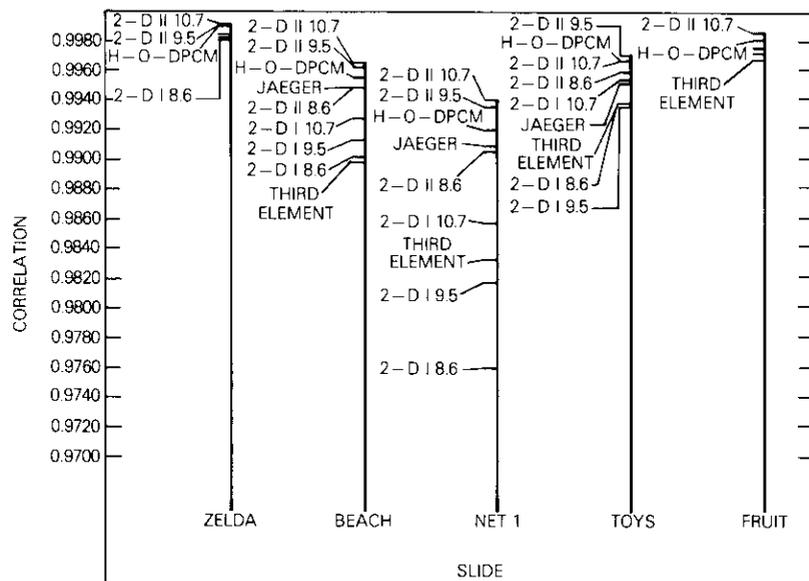


Figure 7. Correlation Between Predictions and Video Values

References

- [1] J. B. O'Neal, "Predictive Quantizing Systems (DPCM) for the Transmission of Television Signals," *Bell System Technical Journal*, May/June 1966, pp. 689-721.
- [2] J. E. Thompson, "Differential Encoding of Composite TV Signals Using Chrominance-Corrected Prediction," *IEEE Transactions on Communications*, COM-22, No. 8, August 1974, pp. 1106-1113.
- [3] J. E. Thompson, "Data Compression of Composite Color Television Signals Using Planar Prediction," *Proc. Third International Digital Satellite Communications Conference*, Kyoto, Japan, November 1975, pp. 315-321.
- [4] Y. Hatori and H. Yamamoto, "Direct Predictive Coding for NTSC Color Signals," *IECE (Japan) 75th National Conference*, Tokyo, Japan, Paper No. 983, March 1975 (in Japanese).
- [5] J. P. Rossi, "Color Decoding a PCM NTSC TV Signal," *Journal of SMPTE*, Vol. 83, No. 6, June 1974, pp. 489-495.
- [6] Y. Iijima and T. Ishiguro, "DPCM of NTSC Color Television Signals," *IECE (Japan) Communications Group Meeting Record*, No. CS73-44, July 1973 (in Japanese).

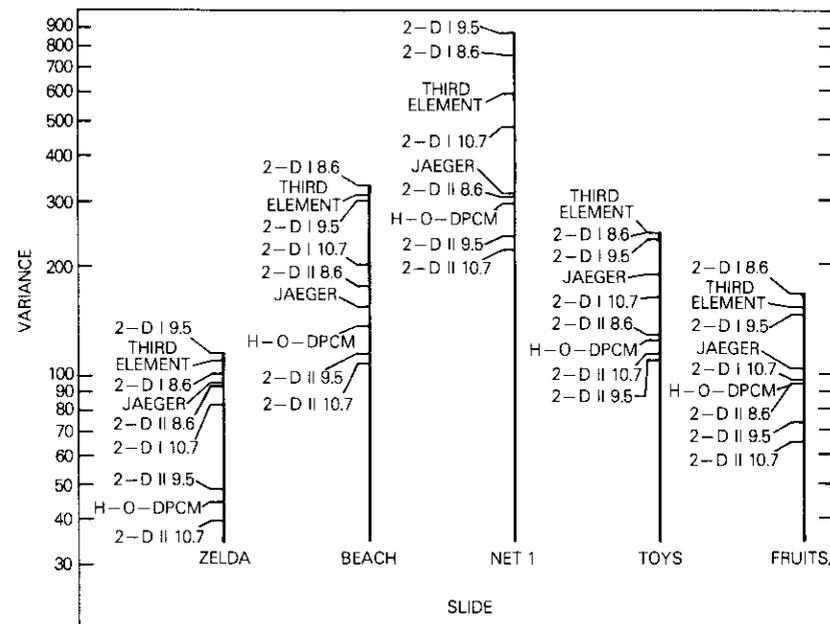


Figure 8. Variance of the Differences

- [7] P. A. Jaeger, D. J. Connor, and O. G. Deczky, "Correlations in Composite NTSC Color Television Signals and Optimum-Predictor Digital Picture Coding," *Picture Coding Symposium*, Asilomar, California, January 1976.
- [8] A. B. Carlson, *Communications Systems*, New York; McGraw-Hill, 1968.

Appendix A. Phase adjustment of samples of NTSC composite signals

Vertical alignment of samples

Assume that NTSC color composite signals are sampled at a rate of $n/m \times f_{sc}$, where n and m are integers, n is even, and f_{sc} is the color subcarrier frequency:

$$f_{sc} = \frac{455}{2} \times f_H \tag{A-1}$$

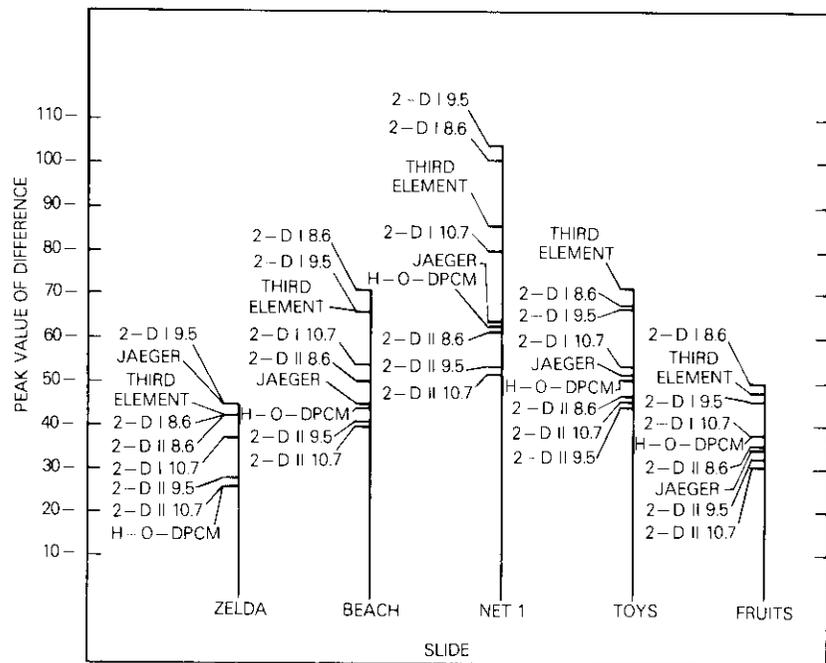


Figure 9. Peak Values of the Differences

where f_H is the line frequency. The number of samples per line is

$$SL = \frac{455}{2} \times \frac{n}{m} \quad (A-2)$$

Samples are vertically aligned when SL is an integer, but this is not generally the case.

If ISL is defined as the largest integer which is smaller than or equal to SL , then

$$SL = ISL + \frac{k}{m} \quad (A-3)$$

where k is an integer, $0 \leq k < m$, and k/m is the horizontal displacement between respective samples of adjacent lines expressed as a fraction of a sample period. To vertically align the samples, the phase of the samples must be adjusted between lines so that this displacement will be cancelled. This is achieved by the circuit shown in Figure A-1. A master clock of frequency $n \times f_{sc}$ is phase locked to the color subcarrier. The master clock output is divided by m , yielding the sampling frequency $n/m \times f_{sc}$. The circuit enclosed in the dashed lines per-

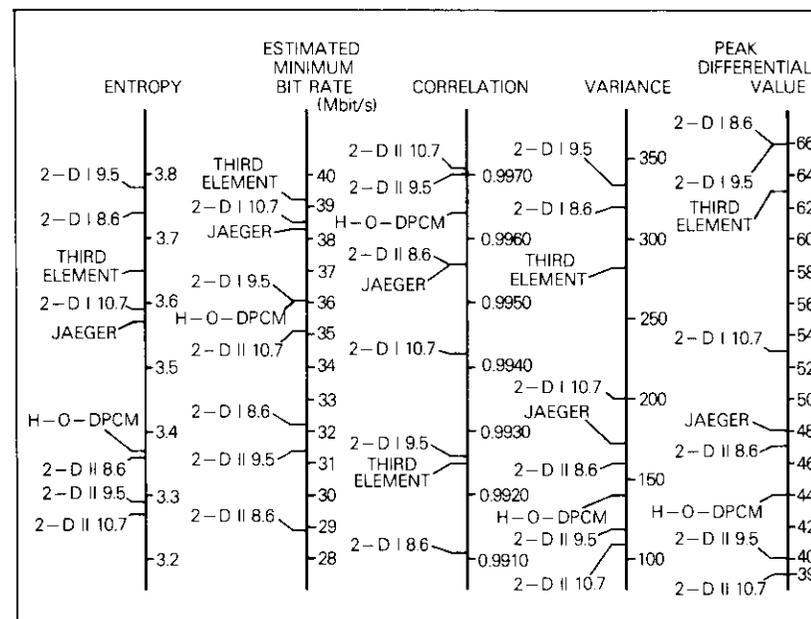


Figure 10. Average Values of Entropies, Estimated Bit Rates, Correlations, Variances, and Peak Differential Values

forms the phase adjustment. Once every horizontal blanking period, the decoder resets the "divide by m " counter upon decoding the number k , derived from equation (A-3):

$$k = m \times (SL - ISL) \quad (A-4)$$

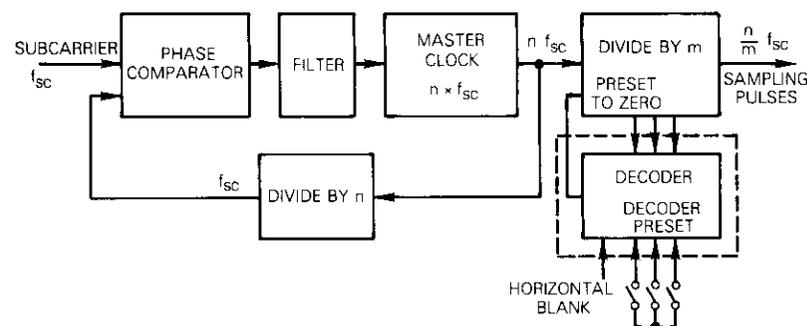


Figure A-1. Phase Adjustment of Sampling Pulses

Minimizing the distance between pairs of samples of adjacent lines having the same subcarrier phase

The minimum horizontal displacement between samples of adjacent lines having the same subcarrier phase corresponds to half a subcarrier period, since there is a 180° phase difference between the subcarrier of those lines. Following the notation of the preceding section, the horizontal displacement between respective samples of adjacent lines expressed with the color subcarrier phase difference is

$$\Delta = k \times \frac{360}{n} \quad (\text{A-5})$$

The required phase relationship between samples is achieved by adjusting the phase of the samples between lines by $l \times 360/n$ so that

$$(k + l + m) \frac{360}{n} = 180 \quad (\text{A-6a})$$

or

$$\frac{k + l + m}{n} = \frac{1}{2} \quad (\text{A-6b})$$

The phase adjustment is performed by the circuit shown in Figure A-1. Once every horizontal blanking period the "divide by m " counter is reset upon detection of the number l , derived from equation (A-6):

$$l = \frac{n}{2} - m - k \quad (\text{A-7})$$



Its'hak Dinstein received a B.S.E.E. from the Technion, Israel, in 1965, and an M.S.E.E. (1969) and a Ph.D. (1974) from the University of Kansas. From 1965 through 1967, he was with the digital instrumentation group of Elron-Elbit, Israel, and from 1967 to 1974, he was a senior research engineer in the Remote Sensing Laboratory of the University of Kansas, working on pattern recognition and image processing. He joined COMSAT Laboratories in 1974, where he was a member of the technical staff in the Image

Processing Department of the Communications Processing Laboratory. Dr. Dinstein is on leave of absence from the Laboratories and is teaching and conducting research on image processing and its applications to medicine and photo interpretation at the University of the Negev in Be'er-Sheva, Israel.

Index: coherent phase shift keying, simulation, repeater, interference, thermal noise, time division multiplexing

Simulation analysis for differentially coherent quaternary PSK regenerative repeater

Y. S. LEE

(Manuscript received March 28, 1976)

Abstract

An onboard regenerative repeater can provide significant improvements in error rate performance of digital satellite communications. This paper presents simulation analysis results for the evaluation of symbol error rate performance with coherent quaternary phase shift keying (CQPSK) and differentially coherent quaternary phase shift keying (DQPSK) detections. Signal degradation mechanisms due to overall intersymbol interference and additive thermal noise are included. Certain critical design parameters have been obtained for the implementation of an up-link DQPSK demodulator-regenerator. System considerations for onboard regenerative repeater applications are discussed with reference to future SS-TDMA systems in which minimum acquisition time is important.

Introduction

Future digital satellite communications systems may employ onboard regenerative repeaters for high-data-rate burst mode transmissions, e.g., time-domain multiple access (TDMA) or satellite-switched TDMA (SS-TDMA) [1], [2]. It has been pointed out that, while Gaussian noise is additive in the up- and down-links in conventional transponders, only the symbol error rates (BERs) are additive when regenerative repeaters are used [3].

The advantage of the regenerative repeater approach is shown in Figure 1, where the three curves apply to ideal channels with an equi-symbol error rate of 1×10^{-4} at the ground receiver. Curve I applies to coherent

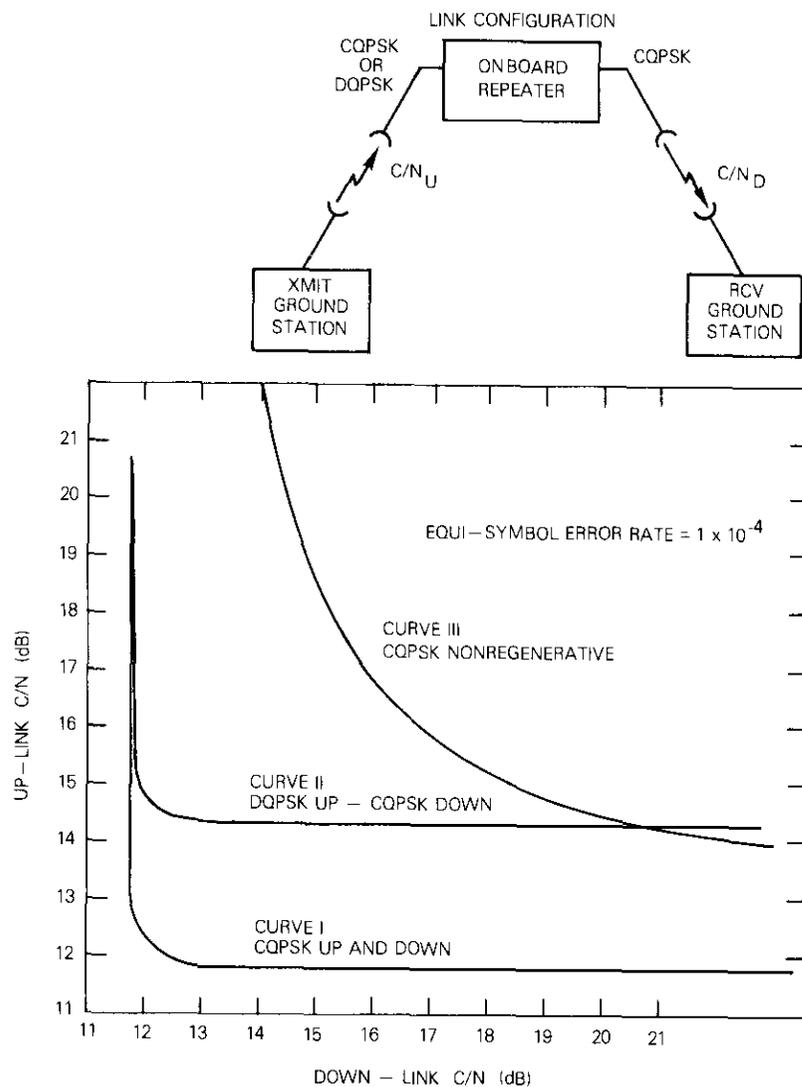


Figure 1. Equi-symbol Error Rate Curves of 4-phase PSK Systems Using a Regenerative Repeater vs a Conventional Transponder

quaternary phase shift keying (CQPSK) detection and onboard regeneration in the up-link and CQPSK detection in the down-link. Curve II applies to differentially coherent quaternary phase shift keying (DQPSK) detection and onboard regeneration in the up-link and CQPSK detection in the down-link, and curve III applies to a conventional transponder for CQPSK transmission including only 1.6-dB carrier-to-noise-power (C/N) ratio degradation [4] in a saturated traveling wave tube amplifier (TWTA). The regenerative repeater approach provides a significant improvement, particularly at low C/N, relative to the conventional transponders.

Theoretically optimum CQPSK detection requires both coherent carrier recovery and symbol timing extraction from the random-access incoming burst signals for which minimum acquisition time is the major concern [5]-[7]. On the other hand, DQPSK detection requires only symbol timing extraction, and symbol synchronizer implementation should be simpler. It is conceivable that the DQPSK demodulator can be implemented directly at the RF up-link carrier frequency, resulting in less complex circuits which can be fabricated in integrated form (MIC).

A potential advantage of the DQPSK detection process would be the removal of phase ambiguity. For a given symbol error rate, the DQPSK up-link requires higher e.i.r.p. than CQPSK, but this higher e.i.r.p. will be available in future satellites using spot beams.

Although the feasibility of an MIC design for a biphase DPSK implementation has been demonstrated [8], no DQPSK design has been reported. It should be noted that the design of a DQPSK channel must include a careful evaluation of the intersymbol interference effects (see Appendix A). Extensive theoretical performance analysis for biphase DPSK systems has been documented [9], [10]. However, since rigorous performance analyses for a DQPSK system are much more complicated, only ideal models of the transmission channel imperfections or interference has been considered [11]-[13].

This paper presents a computer simulation technique for calculating the symbol error rate for DQPSK detection. In addition, it provides computed error rate performance results for typical transmission channels, circuit design criteria for the DQPSK demodulator-regenerator, and systems considerations for the DQPSK regenerative repeater. It should be emphasized that the regenerative repeater can provide onboard signal processing and message rerouting functions for SS-TDMA systems.*

*The companion paper in this issue by L. Palmer and S. Lebowitz discusses the effects of synchronization errors in time-domain simulation.

Simulation technique

A modular transmission channel analysis program (CHAMP) has been developed at COMSAT Laboratories [14]. If ideal PCM-PSK modulation is assumed, the time-domain signals are represented by sampled values of a pair of quadrature envelopes which are sequentially stored on disk files for processing through the simulated channel modules of the transmission system.

The performance measure of the digital message is defined by the probability of error of the received symbols through the overall transmission channel. In CHAMP, the additive white Gaussian noise (WGN) is simulated by the theoretically computed equivalent noise power in the transmission channel noise bandwidth.

The symbol error rate for DQPSK detection is computed as follows. The differentially decoded signal output in the DQPSK demodulator (shown in Figure 2) is obtained from the product of the incoming signal and the signal preceding it by one symbol. As shown in Figure 3, the information is contained in the difference phase, $\Delta\theta$, of the two signals.

If it is assumed that the white Gaussian noise samples in the two successive symbol intervals are independent in a relatively broad bandwidth channel, the probability density function (pdf) of the phase angle difference of the two successive time intervals can be obtained by the convolution integral [15]

$$f_{\Delta\theta}(\Delta\theta) = \int_{-\pi}^{\pi} f_{\theta}(\theta)f_{\theta}(\Delta\theta + \theta) d\theta \quad (1)$$

where $f_{\theta}(\theta)$ is the probability density function of the phase angle associated with each incoming signal, which is [16]

$$f_{\theta}(\theta) = \frac{e^{-\rho}}{2\pi} + \frac{1}{2} \sqrt{\frac{\rho}{\pi}} \cos \theta e^{-\rho \sin^2 \theta} [1 + \operatorname{erf}(\sqrt{\rho} \cos \theta)] \quad (2)$$

where ρ is the C/N ratio, and

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du \quad (3)$$

The probability density function of equation (1) was computed numerically and tabulated for a number of C/N values; its accuracy was checked by computing symbol error rates vs C/N for ideal CPSK and DPSK detection

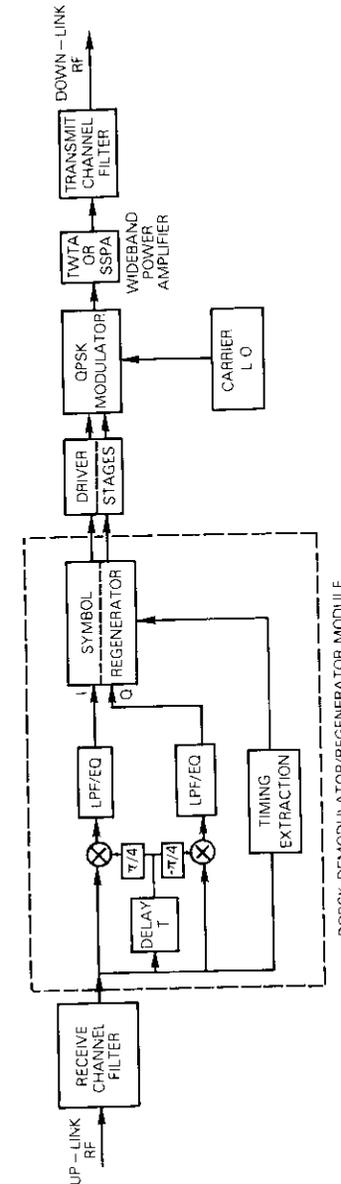


Figure 2. DQPSK Regenerative Repeater Scheme

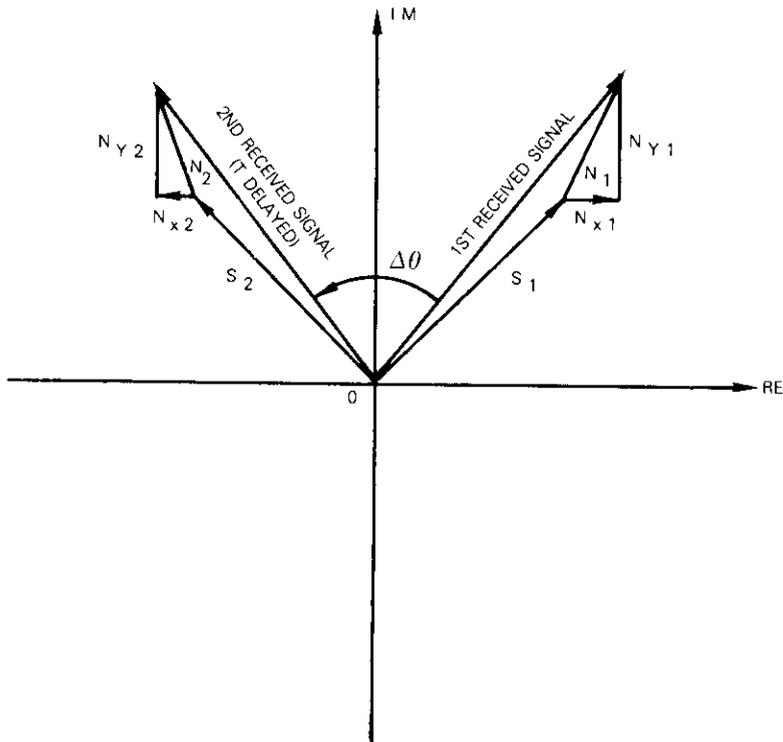


Figure 3. DQPSK Phasor Diagram

with white Gaussian noise only, as shown in Figure 4. For practical purposes, subdivision of the angle 2π into 256 parts for the probability density function computation is sufficient to achieve a C/N accuracy better than 0.4 dB for BER $< 10^{-1}$ considering the actual computational speed.

When the transmission channel is band limited and nonsymmetrical and delay distortion is present, intersymbol interference further degrades the BER. Intersymbol interference shifts the mean value of $f_{\Delta\theta}(\Delta\theta)$ by $\Delta\phi_I$ for each decoded symbol sampling instant. This is equivalent to shifting the decision line by $\Delta\phi_I$ with a fixed mean of $f_{\Delta\theta}(\Delta\theta)$, as shown in Figure 5. Maximum likelihood detection based on equal *a priori* symbol probabilities is then used to compute the symbol error rate as the shaded area in Figure 5, which is the integration of the $f_{\Delta\theta}(\Delta\theta)$ curve outside the decision region determined by

$$\left(-\frac{\pi}{M} + \Delta\phi_I\right) \leq \Delta\theta \leq \left(\frac{\pi}{M} + \Delta\phi_I\right) \tag{4}$$

where $M = 4$ for DQPSK.

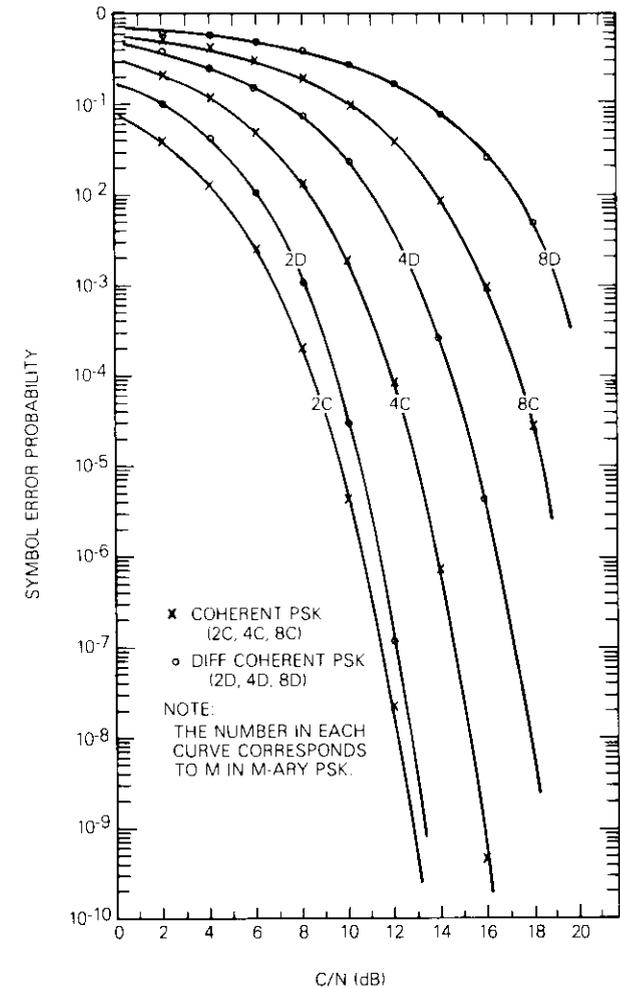


Figure 4. Probability of Symbol Error in Ideal CPSK and DPSK Detections with White Gaussian Noise Only

The average probability of symbol error is

$$P_s(\epsilon) = \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^M P_k(S_i) P_k(\epsilon | S_i) \quad (5)$$

where $P_k(S_i)$ = probability of occurrence of the symbol S_i , which is actually the differentially encoded M -ary discrete phase
 $P_k(\epsilon | S_i)$ = error probability for the given S_i at the k th decision instant
 N = total number of simulated symbols.

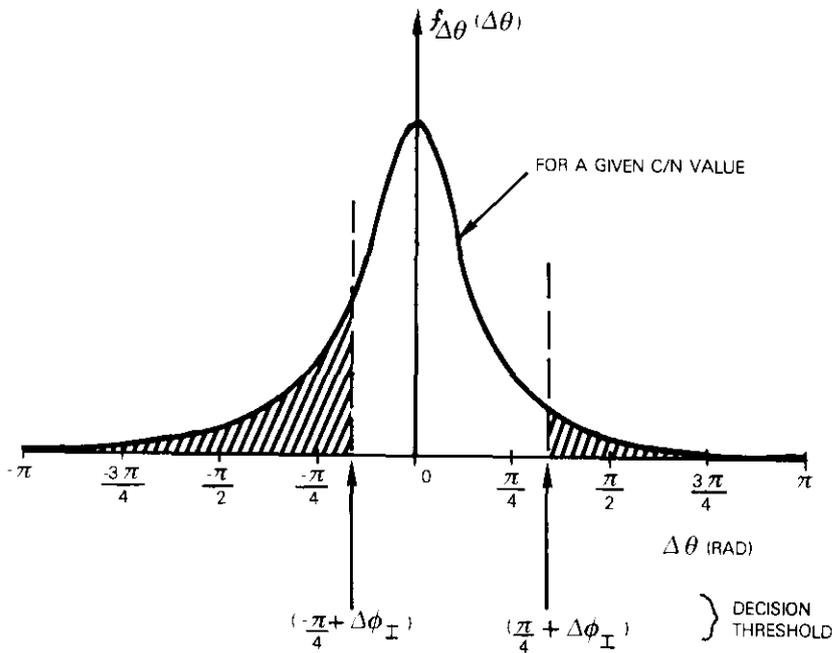


Figure 5. Intersymbol Interference Effect on DQPSK Symbol Error Probability Computation

The error probability was computed for each of the sequences of N simulated symbols, and the average probability of symbol error was plotted vs C/N in dB. The simulated results were obtained using an input

symbol sequence typically consisting of 128 pseudo-random bits with 16 sample points per bit. Overall intersymbol interference effects due to all previous symbols were included in the simulated transmission channel models by computing $\Delta\phi_I$ due to intersymbol interference for each differentially decoded symbol decision instant and using it for the BER calculation as described above.

Transmission channel

CQPSK and DQPSK detection systems were compared by evaluating the symbol error rate performance characteristics in non-ideal (non-Nyquist) transmission channels. In a transmission channel, a theoretically minimum bandwidth Nyquist filter with a constant group delay can be approximated by using various kinds of channel equalizers, and hence the intersymbol interference can be minimized. However, in a hardware implementation a certain amount of degradation must be allowed for fabrication tolerances in the composite transmission channel responses. This degradation may be due to restrictions imposed by non-ideal realization and/or to the changes in operating environmental conditions such as ambient temperature. More importantly, interference encountered in real communications channels cannot be easily taken into account.

The simulation analysis of single-carrier QPSK transmission reveals the dependence of BER performance with CQPSK and DQPSK detection on channel imperfections. The analysis results can lead to specifications for hardware implementation and expected performance capabilities of a DQPSK regenerative repeater.

Figure 6 shows the C/N required for $BER = 1 \times 10^{-4}$ as the number of poles increases for two types of channel filters (Butterworth and Chebyshev) with a constant bandpass 3-dB bandwidth-time product (BT) of 1.5 with ideal CQPSK and DQPSK detection. The simulated Chebyshev channel filter has 0.01-dB passband ripples. As the number of poles is increased for both filters, the channel impairment in DQPSK is greater than that in CQPSK. As shown in Figure 6, higher order Chebyshev filters introduce more C/N degradation than Butterworth filters of the same order. This channel impairment mechanism can be understood by comparing the peak delay distortion in the two types of the filter, as shown in Figure 7. In DQPSK the dependence of the impairment on the parabolic group-delay distortion in the channel filter is greater than that in CQPSK.

Figure 8 shows C/N for $BER = 1 \times 10^{-4}$ as the filter bandwidth is increased from $BT = 1.2$ to $BT = 6.0$ in the two types of 10-pole filters. The

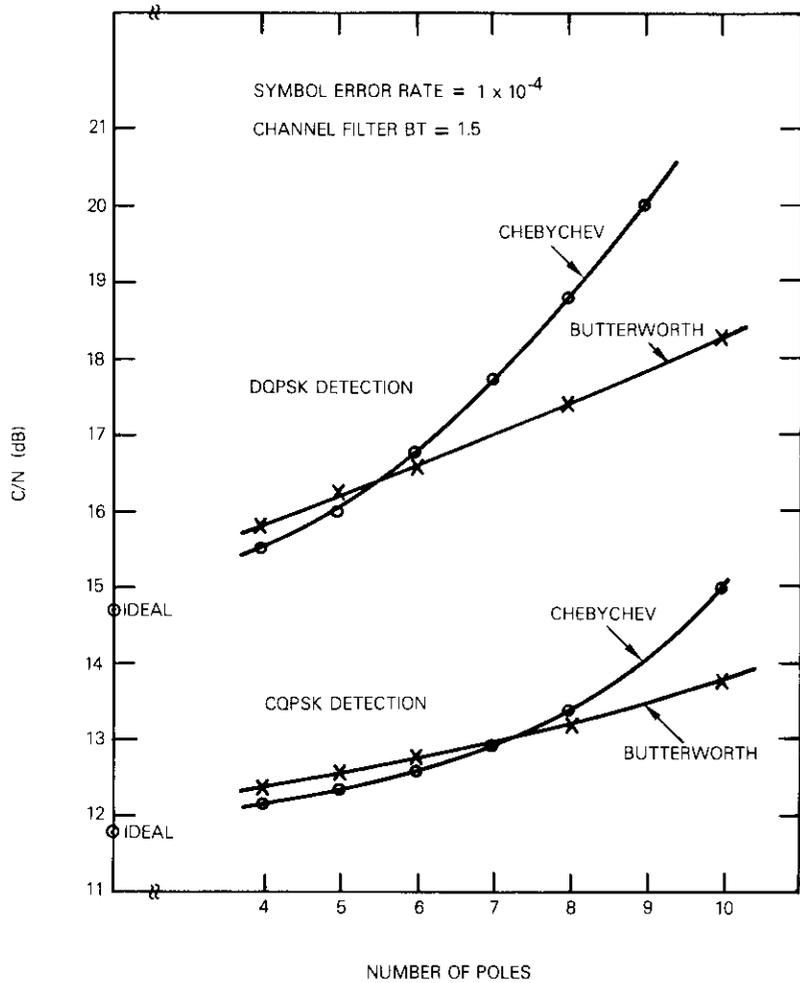


Figure 6. C/N Degradation Dependence on the Number of Poles in Butterworth and Chebychev Channel Filters in CQPSK and DQPSK Detections

obvious result is that for any given filter the channel impairments decrease as the channel bandwidth is increased.

If the transmission channel bandwidth is relatively unrestricted, the simulation results show that, because of their maximally flat group-delay

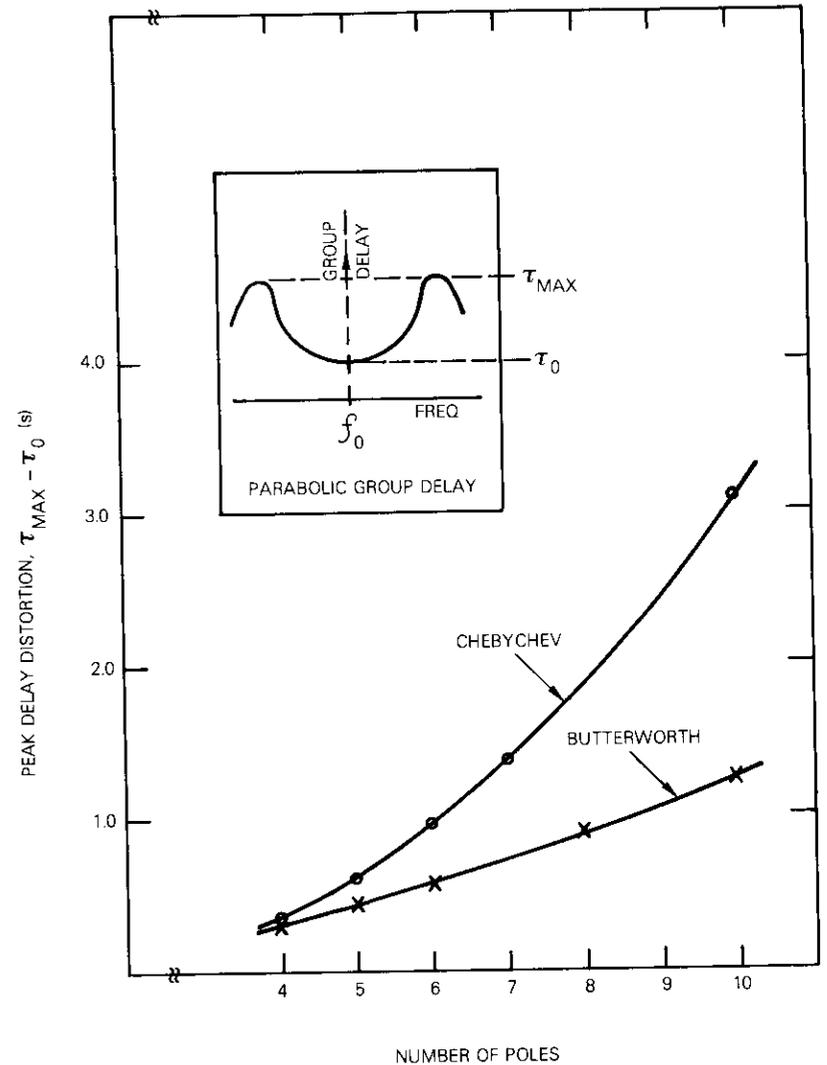


Figure 7. Peak Group-Delay Distortion vs Number of Poles for Chebychev and Butterworth Filters

responses near the band center frequency, higher order ($N > 4$) Bessel-type filters provide transmission characteristics which are as good as those of a raised cosine Nyquist channel filter. However, when the min-

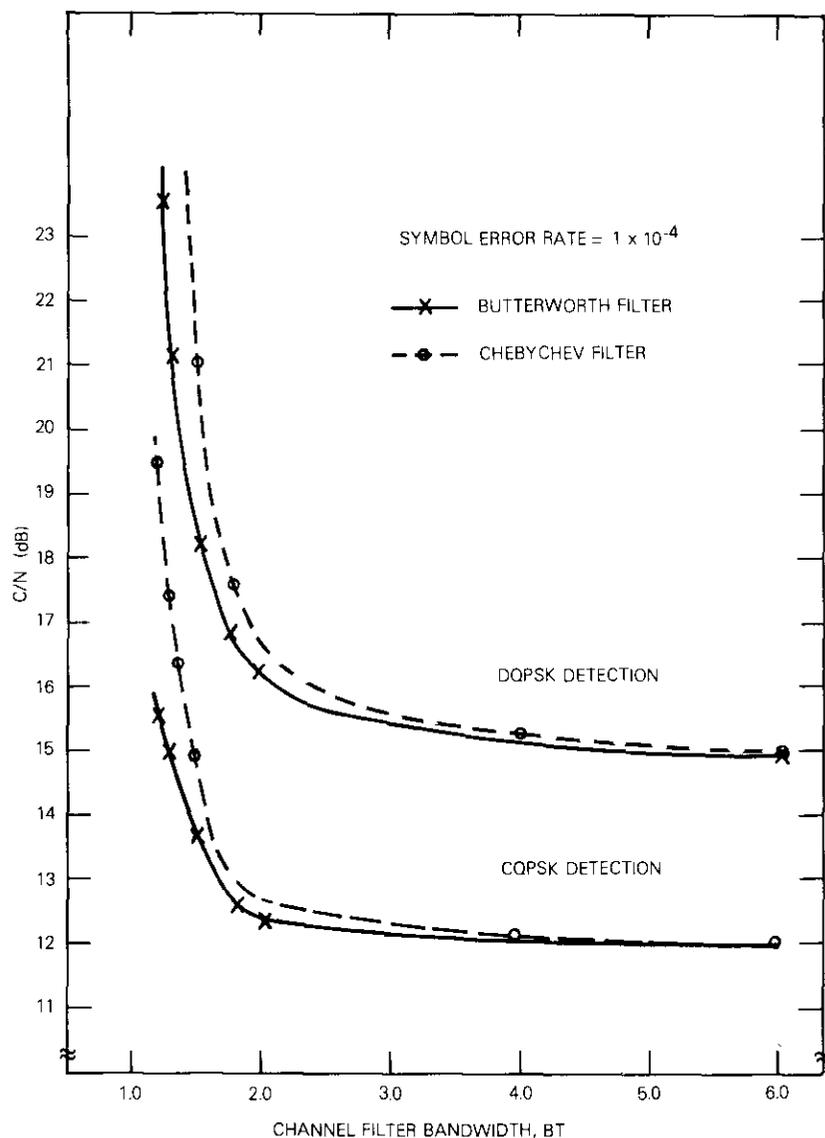


Figure 8. *C/N Degradation Dependence on the Channel Filter Bandwidth of 10-Pole Chebychev and Butterworth Filters*

imum channel bandwidth is allocated and adjacent channel interference is present in a multichannel environment, composite Nyquist channel filter design techniques are necessary for minimum bandwidth occupancy [17].

It has been observed that non-ideal transmission channel characteristics introduce significant effects in the overall system BER performance when the QPSK demodulators are not ideal due to symbol timing offsets and non-ideal 1-symbol delay elements in the DQPSK demodulator, for example. This will be discussed in more detail in the following section.

DQPSK demodulator-regenerator

The DQPSK demodulator-symbol regenerator module is shown in Figure 2. Assuming that the RF (or IF) signal phase shift through the 1-symbol delay element is an integer multiple N of 2π , if ideal product-type phase detectors are used, the most critical component is the delay element, which affects the 1-symbol-delayed reference carrier phase. Exact phase relationships must be maintained between the I and Q channels in the demodulator circuits.

Stability and distortion in delay elements

Analog delay elements, such as electromagnetic transmission lines or very wideband delay filter circuits, are suitable for a very-high-data-rate (≥ 120 -Mbit/s) DQPSK demodulation. In this simulator program the 1-symbol-delay RF phase stability requirement was determined, and the BER degradation effects due to parabolic group-delay distortion in the delay filter were analyzed to obtain DQPSK demodulator design criteria.

Figure 9 shows the computed symbol error rate vs C/N with the RF phase deviation (from $2\pi N$) as a parameter. The result was computed for DQPSK detection with mid-bit sampling for the pseudo-random PSK inputs. The accuracy of the RF phase offset was $\pm 0.7^\circ$ in the simulation. For a constant BER of 1×10^{-4} , the equivalent C/N degradation vs phase deviation is plotted as curve II in Figure 10 for the limit case of infinite bandwidth.

Figure 10 also shows that the DQPSK detection impairment due to small delay phase offsets ($< 15^\circ$) is not dependent on the Nyquist channel filter roll-off. Curve I is the computed result for a cosine roll-off Nyquist channel filter with a roll-off factor (ρ) of 1.0, and curve III is the result for Assal's near-optimum composite channel Nyquist filter [17] ($1 + \rho \approx 1.2$).

Figure 10 shows that about 6° of delay phase deviation introduces an

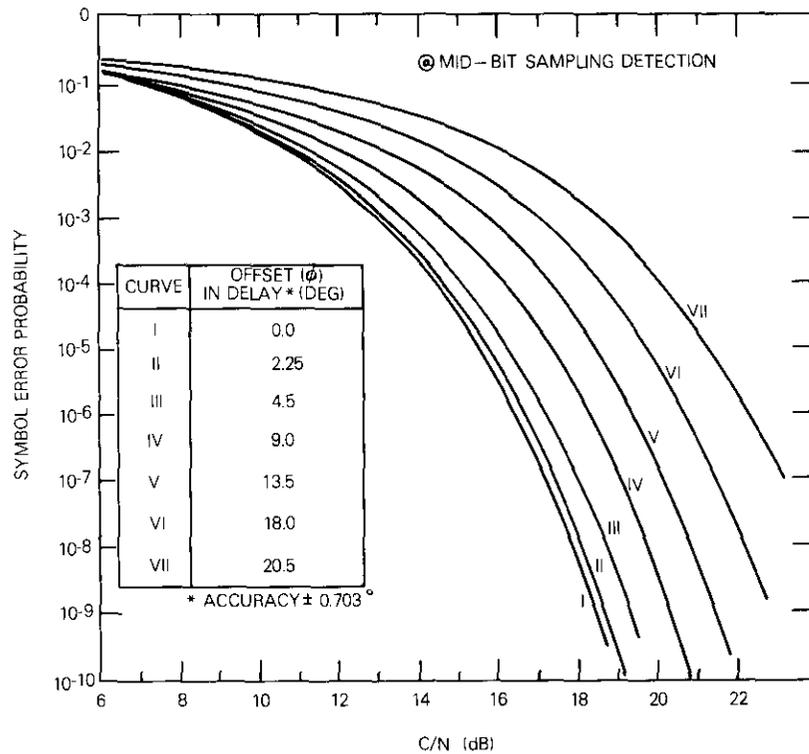


Figure 9. BER Curves for Small Phase Offsets in the 1-Symbol Delay Element in the DQPSK Demodulator

equivalent C/N degradation of 1 dB at BER = 1×10^{-4} . As a design requirement, the maximum phase offset must be restricted to 3° , which is equivalent to 0.3-dB C/N degradation in DQPSK detection.

The highest possible RF for the DQPSK demodulation can therefore be determined in conjunction with the actual realizable delay element phase stability characteristics for a given data rate. For instance, a phase offset of less than 6° at 14 GHz for DQPSK at a data rate of 120 Mbit/s requires a delay element stability smaller than 70 parts per million (ppm) over the operating environment variations. Delay stability can be reduced for a lower demodulating RF at a given baud rate, or for higher baud rates at a fixed RF.

The effects of channel group-delay distortion on data transmission, analyzed by Sunde [12] for the ideal DQPSK detection system, assume some

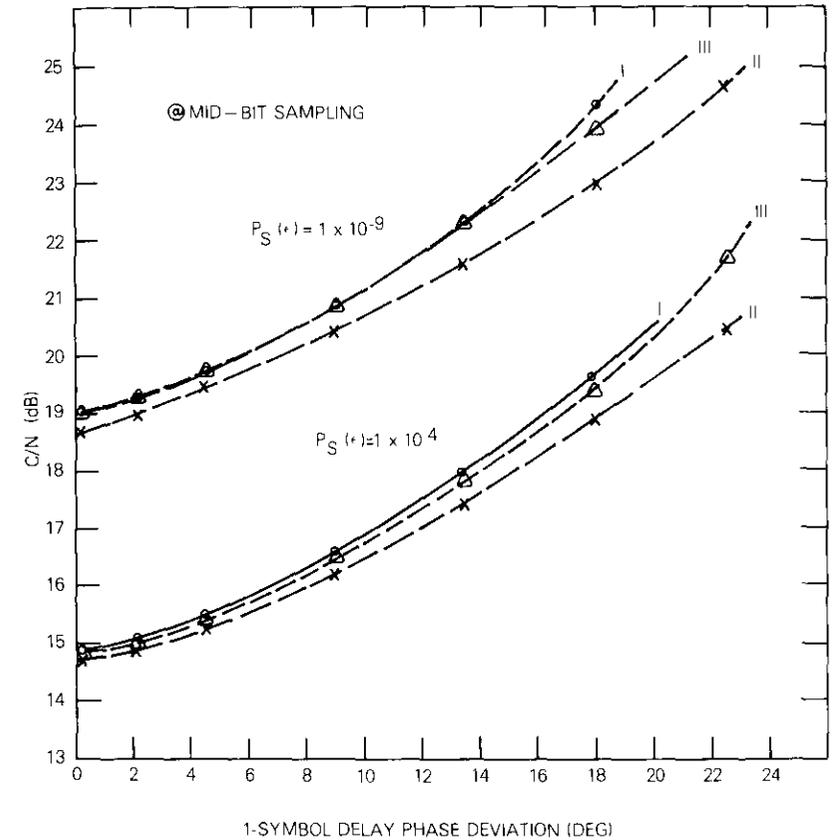


Figure 10. C/N Impairment Dependence on the 1-Symbol Delay Phase Stability in Two Types of Nyquist Channels

particular group-delay characteristics. The present simulation analysis was performed to further evaluate the effects of parabolic group-delay distortion in both the 1-symbol delay element and the actual channel filter on BER performance for DQPSK detection.

The parabolic group delay was defined as $T_D(F) = AF^2 + 1.0$, where T_D is normalized as a 1-symbol interval at the normalized band center frequency, $F = 0$. Figure 11 shows the computed group-delay response of a 10-pole Chebychev filter (0.01-dB ripple) with a fractional bandwidth of 3.5 percent. [$T_D(0) = 1.0$ for a rate of 208 Mbauds and $A = 0.34$.]

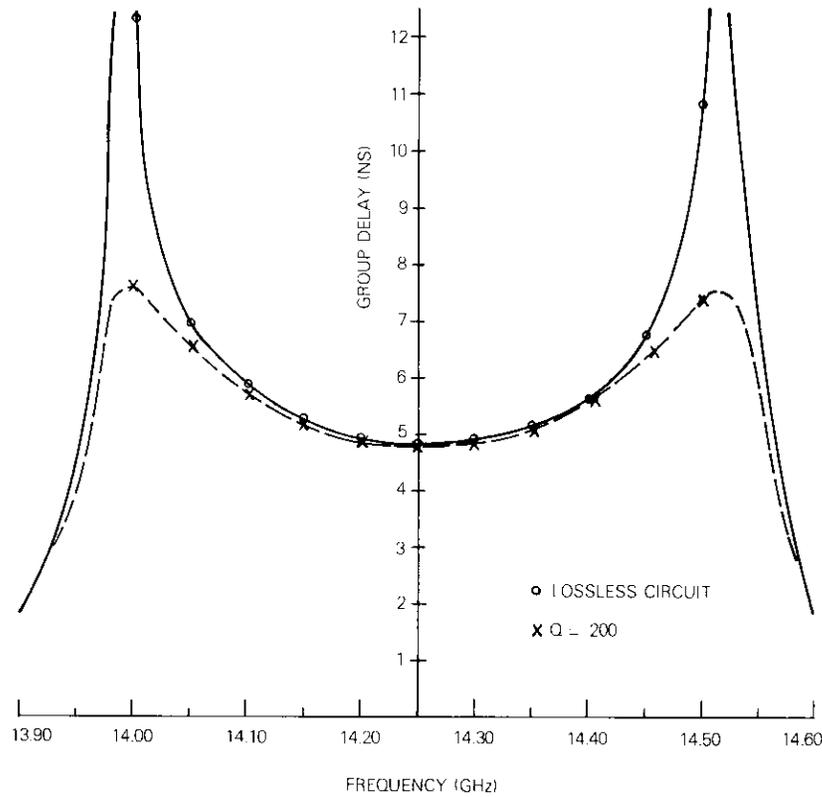
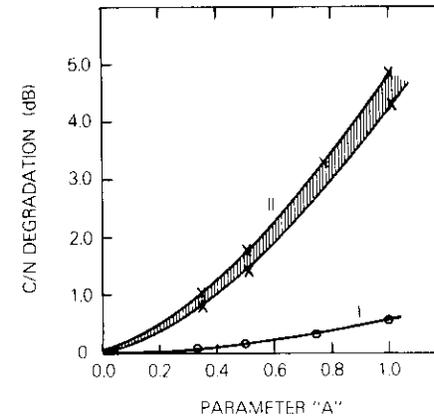


Figure 11. Group Delay of a 10-Pole 14-GHz Bandpass Filter with a Bandwidth of 3.5 percent and Unloaded Q's of 200 and ∞

The channel characteristic was simulated as a raised cosine (unity roll-off factor) Nyquist filter ($T_D = 1.0$), and the constant-amplitude parabolic group-delay element was synthesized by the FLT6 command in CHAMP using a finite-duration impulse response filter simulation technique. The computed C/N degradation vs the parameter A is shown in Figure 12 for a constant BER of 1×10^{-4} . Curve I of Figure 12 was obtained with an additional parabolic group-delay filter, FLT6, in the transmission channel followed by ideal DQPSK detection. Curve II was obtained with FLT6 as a 1-symbol delay element in the DQPSK demodulator, as shown by the block diagram in Figure 12.

Curve II in Figure 12 shows that the delay distortion in the 1-symbol



SIMULATION MODEL:

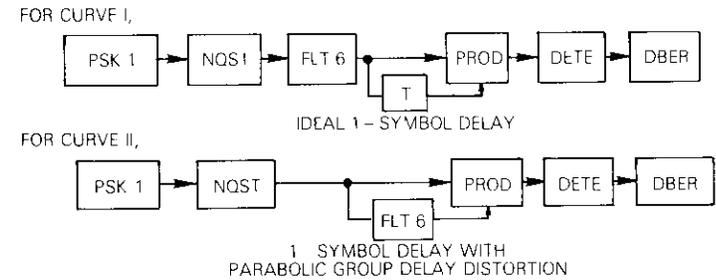
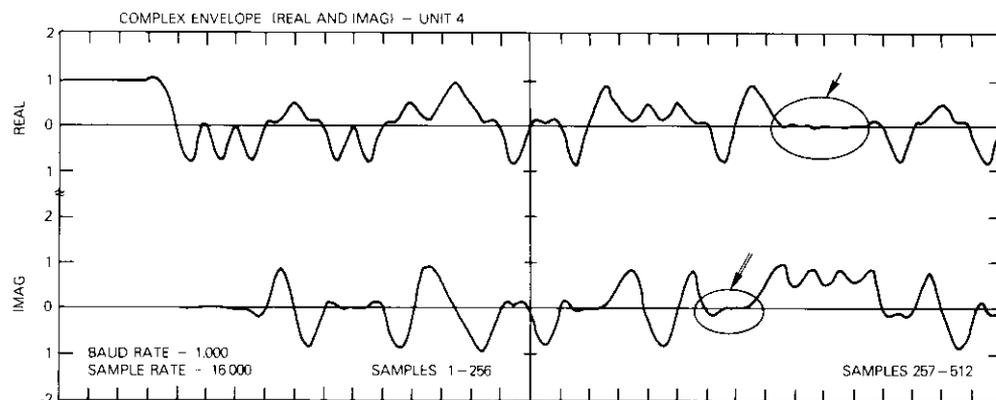


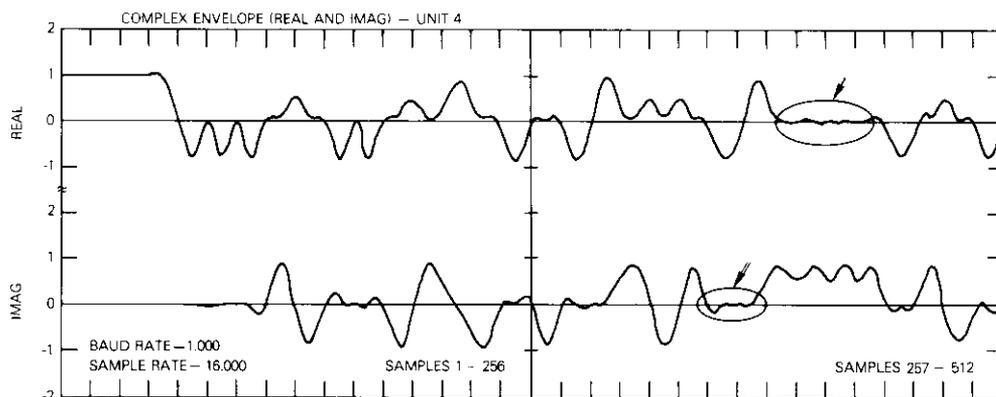
Figure 12. C/N Degradation Dependence on Group-Delay Distortion Introduced in the Transmission Channel vs That Introduced in the 1-Symbol Delay Element for DQPSK Detection

delay element introduces a large impairment in DQPSK detection. The BER degradation mechanism can be identified in the simulated product (difference phase) symbol streams of the first two frames, as shown in Figure 13, particularly in those bits indicated by arrows. Figures 13a, 13b, and 13c were obtained with the parameter $A = 0, 0.5, \text{ and } 1.0$, respectively, in the simulated FLT6 as the delay element. If used, the parabolic delay filter must be restricted so that $A < 0.1$; the delay filter bandwidth must be very broad ($BT \approx 10$) for the delay element application.

a.



b.



c.

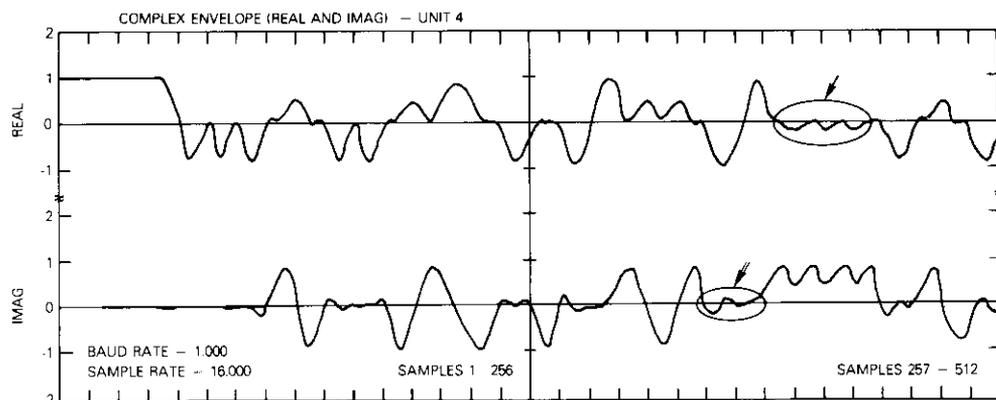


Figure 13. Product Bit Streams as the Output of Phase Detectors in a Raised Cosine (roll-off factor = 1.0) Nyquist Channel

- a. With no Delay Distortion ($A = 0$) in the Parabolic Group Delay of the 1-Symbol Delay Element in the DQPSK Demodulator
- b. With $A = 0.5$ in the Parabolic Group Delay of the 1-Symbol Delay Element in the DQPSK Demodulator
- c. With $A = 1.0$ in the Parabolic Group Delay of the 1-Symbol Delay Element in the DQPSK Demodulator

Symbol timing and regenerator threshold width

The sampling instant for symbol regeneration in the DQPSK regenerative repeater must be set at the maximum eye opening instant to obtain optimum BER performance in a bandwidth-limited transmission channel. The symbol timing clock can be extracted from the received bit streams with a symbol timing extraction circuit having a minimum acquisition characteristic. The effect of symbol timing offset (or rms phase jitter) in the timing clock on BER performance is strongly dependent on the characteristics of the transmission channel.

The BER curves in Figure 14 are the computed results for a number of small symbol timing offsets (in percent) from the optimum mid-bit sampling in DQPSK detection. The simulated channel is equivalent to Assal's composite Nyquist filter [17]. Figure 15 compares the equivalent C/N

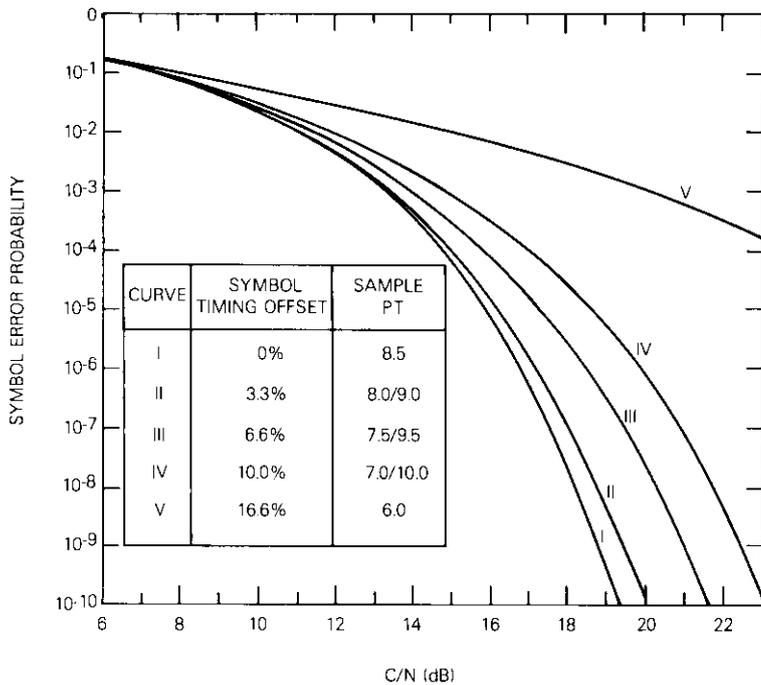


Figure 14. BER Curves with Symbol Timing Offsets for DQPSK Detection in a Composite Nyquist Channel

degradation due to symbol timing offsets in DQPSK and CQPSK detection with two types of Nyquist channel filters in the transmission channel. The curves designated by I in Figure 15 are with a raised cosine Nyquist filter

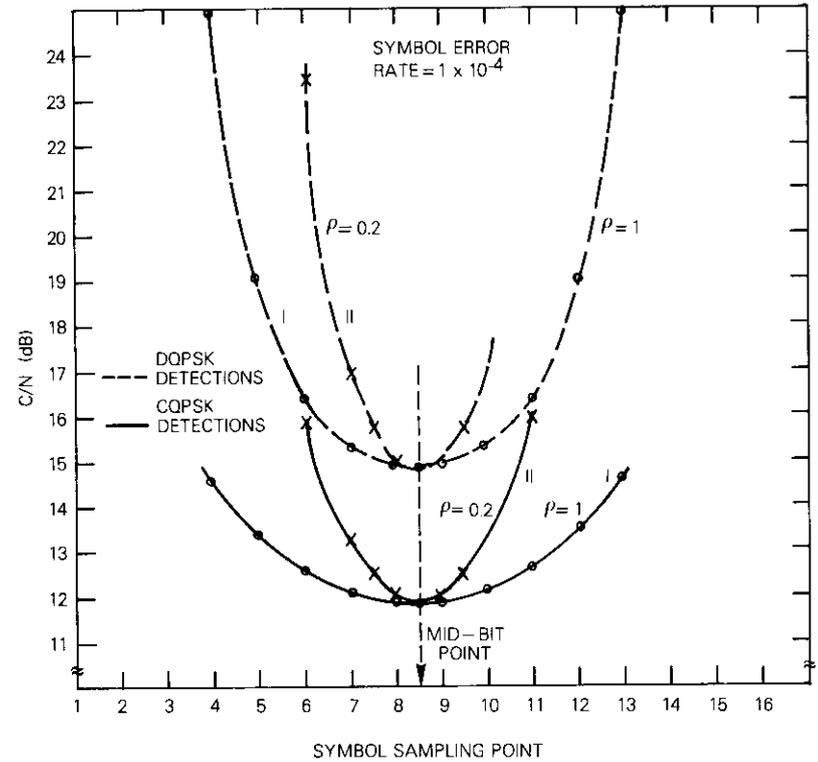


Figure 15. C/N Degradation Dependence on Symbol Timing Offsets in CQPSK vs DQPSK Detection in Raised Cosine Nyquist (roll-off factor = 1.0, curve I) and Composite Nyquist (curve II) channels

(roll-off factor of 1.0); those designated by II are with the composite Nyquist filter, which has a narrower bandwidth (roll-off factor of 0.2).

For DQPSK detection with the composite filter ($\rho = 0.2$) in the channel, the expected C/N degradation is 1 dB at BER = 1×10^{-4} when the symbol timing is offset by 6.6 percent from the optimum mid-bit point. For CQPSK detection, an 8.3-percent timing offset can cause the same amount of

BER degradation. Figure 16 shows the BER performance curves for a number of symbol timing offsets in CQPSK detection in the composite Nyquist channel filter.

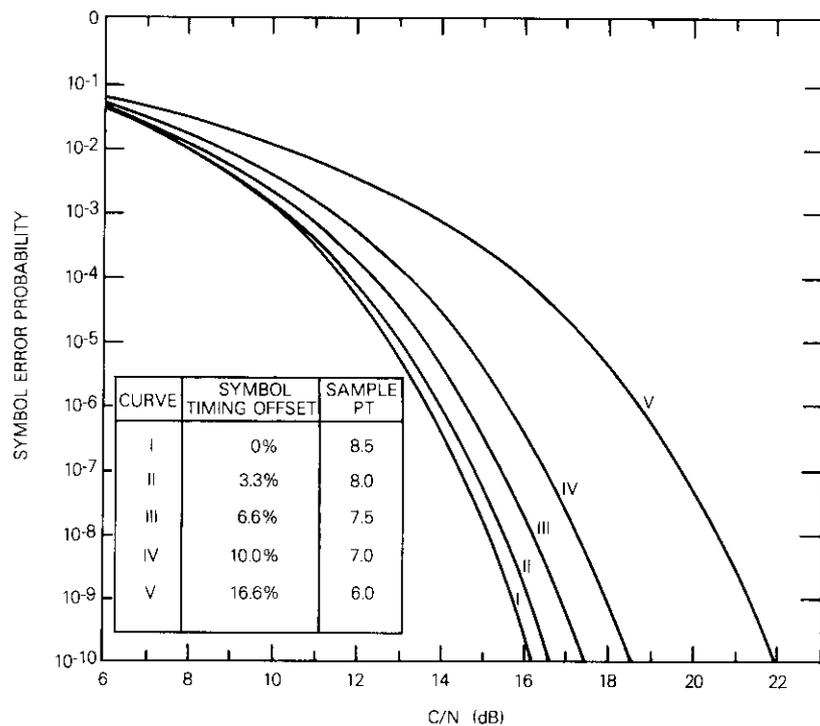


Figure 16. BER Curves with Symbol Timing Offsets for CQPSK Detection in a Composite Nyquist Channel

The curves in Figures 14-16 show that, with the same channel transfer characteristic, the symbol timing offset in DQPSK more severely degrades the BER performance than that in CQPSK. For a channel with a wider bandwidth, the BER performance is less dependent on symbol timing offset in both DQPSK and CQPSK. For a given transmission channel, the minimum allowable timing offset must be evaluated carefully in the implementation of the actual symbol timing circuitry.

BER degradation can also be caused by a finite decision width in the symbol regenerator because it is equivalent to the non-zero width in the

decision threshold in the phasor diagram for the symbol error rate computation. Simulation results show that a finite decision threshold corresponding to 9° can degrade the C/N by about 2 dB at $BER = 1 \times 10^{-4}$.

QPSK regenerative repeater

An onboard regenerative repeater for PCM-QPSK gives an increased margin for low C/N values in the up- and down-links, as shown in the equi-BER curves in Figure 1. One advantage of the DQPSK onboard regenerator up-link is the simpler DQPSK demodulator implementation in a microwave integrated circuit (MIC) configuration for SS-TDMA applications. The down-link can employ a CQPSK system.

The regenerated symbols can be rerouted through onboard switch matrices, and the rerouted bit streams can be directly CQPSK modulated with a common down-link RF carrier. Each CQPSK modulated signal goes through a broadband ($BT > 10$) TWTA. In a band-limited non-regenerative transponder, the C/N degradation in the saturated TWT is greater than 1.5 dB at $BER = 1 \times 10^{-4}$ due to AM/PM and AM/AM conversion. These effects can be almost completely eliminated in the regenerative repeater because the pure phase modulated RF carrier maintains a constant amplitude envelope and encounters no AM/PM conversion in the wideband TWT nonlinearity.

In the regenerative repeater, transmit channel filters following TWTA's or broadband solid-state power amplifiers will shape the down-link spectrum and prevent any inter-channel interference in a multiple-channel application. A single-carrier QPSK system may be useful at gigabit rates for which logic elements such as transferred-electron logic devices are being developed.

MIC configurations for the DQPSK demodulator shown in Figure 2 can be implemented for direct detection at the up-link carrier frequency without down-converting to IF. The critical element is the 1-symbol delay circuit, which is required to maintain a phase stability less than 3° over the operating temperature range. The delay stability of MIC delay elements, such as delay bandpass filters or transmission lines on temperature stable dielectric substrates, must be experimentally evaluated to determine the highest practical RF for demodulation. Temperature control or a temperature proportional microwave phase shifter may be necessary to obtain the RF delay phase stability.

A precision symbol timing clock signal can be extracted for symbol regeneration at the optimum sampling instant. A number of nonlinear

elements followed by tuned circuits can be used for clock extraction. A high-speed automatic phase control circuit with a stable, non-slave oscillator in the repeater is also suitable for fast clock acquisitions (less than 50 ns) using additional burst detection circuitry. However, PLL circuits are not desirable for this purpose due to the hang-up phenomena in acquisition.

The present simulation study provided some of the critical design criteria for the DQPSK demodulator-symbol regenerator subsystem. Further design optimization of the overall DQPSK regenerative repeater system can be obtained by using the DQPSK BER module in CHAMP for given transmission channel characteristics.

Conclusions

The DQPSK simulation capability in CHAMP provides a versatile tool for system design and development of an onboard regenerative repeater employing DQPSK in the up-link and CQPSK in the down-link for future SS-TDMA applications.

The BER computation for the overall transmission channel included signal degradation due to white Gaussian noise and intersymbol interference. Some of the critical design parameters were obtained through simulation analysis of the DQPSK demodulator-regenerator. The 1-symbol delay element must be very stable to <35 ppm, corresponding to an RF phase stability $<3^\circ$, over the operating environmental variations for a 14-GHz up-link frequency DQPSK detection of 120-Mbit/s data, for example. If present, parabolic group delay in the delay element must be restricted so that the parameter A is less than 0.1. The tolerable symbol timing offset in the regenerator is quite dependent on the actual transmission channel characteristics.

The onboard regenerative repeater can provide an increased C/N margin in the overall system link budget by eliminating the serious degradation occurring in the saturated high-power amplifier. In addition, it is capable of performing the signal processing and message rerouting functions for future SS-TDMA systems.

Acknowledgment

The author would like to thank S. Lebowitz for his CHAMP simulation program and F. Assal for helpful technical discussions.

References

- [1] W. G. Schmidt, "Satellite Time-Division Multiple Access Systems: Past, Present and Future," *Telecommunications International*, Vol. 7, No. 8, August 1973, pp. 21-24.
- [2] G. D. Dill, Y. Tsuji, and T. Muratani, "Applications of SS-TDMA in a Channelized Satellite," *ICC'76 Conference Record*, Vol. 3, pp. 51-1-51-5.
- [3] S. J. Campanella, F. Assal, and A. Berman, "Onboard Regenerative Repeater," *ICC'77 Conference Record*, Vol. 1, pp. 6.2-121-6.2-125.
- [4] F. T. Assal and K. Betaharon, "In-Band Degradations for Single-Carrier Transmission in Nonlinear Satellite Channels," *New Directions in Signal Processing in Communication and Control*, NATO Advanced Study Institute, Series E, 1975, pp. 347-358.
- [5] L. Lundquist, M. Lopriore, and F. M. Gardner, "Transmission of 4ϕ -Phase-Shift Keyed Time-Division Multiple Access over Satellite Channels," *IEEE Transactions on Communications*, COM-22, No. 9, September 1974, pp. 1354-1360.
- [6] F. M. Gardner, "Carrier and Clock Synchronization for TDMA Digital Communications," ESTEC Preliminary Working Paper (Rev. 2), June 1974.
- [7] K. Nosaka and T. Muratani, "New Satellite Transponder with Regenerative Repeater and Higher Order DPSK for TDMA/SDMA System," *INTELSAT/IEE Second International Conference on Digital Satellite Communications*, *Proc.*, pp. 460-468.
- [8] H. F. Lenzing, "An L-band High Baud-rate DCPSK Detector/AFC Discriminator in Microstrip," *1972 International Microwave Symposium Digest*, pp. 201-203.
- [9] O. Shimbo, M. I. Celebiler, and R. J. Fang, "Performance Analysis of DPSK Systems in Both Thermal Noise and Intersymbol Interference," *IEEE Transactions on Communication Technology*, COM-19, December 1971, pp. 1179-1188.
- [10] W. M. Hubbard, "The Effect of Intersymbol Interference on Error Rate in Binary Differentially-Coherent Phase-Shift-Keyed Systems," *Bell System Technical Journal*, July-August 1967, pp. 1149-1172.
- [11] A. S. Rosenbaum, "PSK Error Performance with Gaussian Noise and Interference," *Bell System Technical Journal*, February 1969, pp. 413-442.
- [12] E. D. Sunde, *Communication Systems Engineering Theory*, New York: John Wiley & Sons, Inc., 1969, Chapter 7.
- [13] V. K. Prabhu, "Error Rate Considerations for Digital Phase-Modulation Systems," *IEEE Transactions on Communication Technology*, COM-17, February 1969, pp. 33-42.
- [14] W. L. Cook, "Interactive Computer Simulation of Satellite Transmission Systems," *Proc. Fifth Annual Pittsburgh Conference, Modeling and Simulation*, Vol. 5, Pt. 2, 1974, pp. 867-872.

- [15] R. W. Lucky, J. Salz, and E. J. Weldon, Jr., *Principles of Data Communications*, New York: McGraw-Hill Book Co., 1968.
- [16] D. Middleton, *An Introduction to Statistical Communication Theory*, New York: McGraw-Hill Book Co., 1960, p. 417.
- [17] F. Assal, "Approach to a Near-Optimum Transmitter-Receiver Filter Design for Data Transmission Pulse-Shaping Networks," *COMSAT Technical Review*, Vol. 3, No. 2, Fall 1973, pp. 301-322.

Appendix A. DQPSK detection and intersymbol interference

A differentially encoded M-phase PSK modulated signal can be written as

$$s(t) = \text{Re} \sum_{k=-\infty}^{\infty} w(t - kT) e^{i(\omega_0 t + \phi_k)} \quad (\text{A-1})$$

where ϕ_k is the differentially encoded M-ary phase variable, which can be given by

$$\phi_k = 2\pi \frac{i-1}{M} \quad (\text{A-2})$$

In equation (A-2), i is an integer chosen from the set $\{1, 2, \dots, M\}$. In equation (A-1), $w(t)$ represents the window function, and T is the 1-symbol duration.

Suppose that the original information is mapped into a signal set whose phase variable can be written as follows for each time interval of T :

$$\dots, \phi_{k-1}^{\circ}, \phi_k^{\circ}, \phi_{k+1}^{\circ}, \dots \quad (\text{A-3})$$

Then its differentially encoded phase variable becomes

$$\dots, (\phi_{k-1}^{\circ} + \phi_{k-2}), (\phi_k^{\circ} + \phi_{k-1}), (\phi_{k+1}^{\circ} + \phi_k), \dots \equiv \{\phi_k\} \quad (\text{A-4})$$

where the symbol $\{ \}$ indicates "set of." The variable ϕ_k in equation (A-1) is defined by equation (A-4).

When the signal of equation (A-1) is transmitted through an ideal channel, the received signal phase, $\phi_{\gamma,k}$, can contain an ambiguous phase component, ϕ_a in each ϕ_k so that

$$\{\phi_{\gamma,k} = \phi_k + \phi_a\} \quad (\text{A-5})$$

The output of the ideal differentially coherent M-ary PSK demodulator can be given by the decision variables as the following sequential set

$$\{\hat{\phi}_k = \phi_{\gamma,k} - \phi_{\gamma,k-1}\} \quad (\text{A-6})$$

If it is assumed that ϕ_a is constant, at least for two symbol intervals ($2T$), equation (A-6) becomes

$$\{\hat{\phi}_k = \phi_k^{\circ}\} \quad (\text{A-7})$$

using equations (A-5) and (A-4). Equation (A-7) shows that the original phase information is recovered without ambiguity.

When the transmission channel is non-ideal, the signal complex envelope arriving at the demodulator through the channel at instant $t = kT$ is given by

$$\bar{S}_o = p(0) e^{i\phi_k} + \sum'_{i=-\infty}^{\infty} p[(k-i)T] e^{i\phi_i} \quad (\text{A-8})$$

where $p(kT)$ is determined from the impulse response of the overall transmission channel transfer characteristic, and the prime in the summation indicates that i excludes k . A complex envelope of a signal delayed by one symbol is given by

$$\bar{S}_{-1} = p(-1) e^{i\phi_{k-1}} + \sum'_{i=-\infty}^{\infty} p[(k-i-1)T] e^{i\phi_i} \quad (\text{A-9})$$

where the prime indicates that i excludes $k-1$.

From equations (A-8) and (A-9), if there are ideal product phase detectors in the differentially coherent 4-phase PSK demodulator, the demodulated output can be written as follows at sampling instant $t = kT$:

$$\begin{aligned} S_{PD} &= S_o S_{-1}^* \\ &= p(0) p(-1) e^{i(\phi_k - \phi_{k-1})} \\ &\quad + \left\{ \sum'_{\substack{i=-\infty \\ (i \neq k)}}^{\infty} p(-1) p[(k-i)T] e^{i(\phi_i - \phi_{k-1})} \right. \\ &\quad + \sum'_{\substack{i=-\infty \\ (i \neq k-1)}}^{\infty} p(0) p[(k-i-1)T] e^{i(\phi_k - \phi_i)} \\ &\quad + \left. \sum'_{\substack{j=-\infty \\ (j \neq k)}}^{\infty} \sum'_{\substack{i=-\infty \\ (i \neq k-1)}}^{\infty} p[(k-j)T] \right. \\ &\quad \cdot \left. p[(k-i-1)T] e^{i(\phi_j - \phi_i)} \right\} \quad (\text{A-10}) \end{aligned}$$

The term in braces in equation (A-10) represents intersymbol interference due to all the previous symbols. The simulation results have shown that the intersymbol interference effects for DQPSK detection in a non-ideal channel are greater than those for CQPSK detection as shown in Figure 6.



Young Soo Lee received a B.S. (1963) and an M.S. (1967) in electronics engineering from Seoul National University and S.M. (E.E.) and Electrical Engineer degrees from M.I.T. (1969). Since September 1969, he has been a member of the technical staff at COMSAT Laboratories, specializing in microwave active and passive circuits and subsystems as applied to communications satellites. He is currently engaged in the development of an onboard regenerative repeater for future digital satellite communications systems. He is a member of IEEE.

Index: computer simulation, phase shift keying, probability theory, error, synchronization

Including synchronization in time-domain channel simulations

L. C. PALMER AND S. LEBOWITZ

(Manuscript received May 17, 1977)

Abstract

Computer simulations have been performed to determine the degradation of error probability due to errors in estimating carrier phase and symbol epoch for quaternary phase shift keyed (QPSK) modulation. A simulation program (RECEIVER) has been developed to model specific receive modem implementations. This program utilizes data generated by COMSAT's CHAMP (*channel modeling program*) and processes these data to determine the sensitivity of error probability to phase and timing errors; the statistical distribution of these errors, $P(\theta_e, \tau_e)$; and the average probability of bit error, $\bar{P}_b(\bar{\epsilon})$, including the effects of synchronization errors and the noise and distortion introduced by the channel.

The model is applied to a particular receiver implementation: a quadrupling circuit and tuned filter which estimates carrier phase, and a one-half-symbol delay, multiply, and filter combination which recovers a symbol clock from the modulated signal. The effect of synchronization errors is determined for typical satellite channel configurations including low bandwidth-time product (BT_b) filtering and nonlinearities which produce AM/AM and AM/PM distortions.

Introduction

Time-domain simulation has become a useful analysis tool for predicting the performance of digital transmission channels. The computer simulation approach supplements analytical and hardware simulation modeling.

It provides a flexible and efficient means of analyzing complex channel configurations containing filters, nonlinearities, and interfering signals. Flexible and realistic simulation programs have been developed, for example, to model the communications satellite [1]-[3]. These programs give insight into the sources of transmission impairment encountered on satellite channels.

When the results of these computer simulations are compared to hardware simulator results, it can be seen that the former are sometimes optimistic in that the software models do not predict all of the losses that are encountered in practice. A large part of the discrepancy seems to be due to synchronization errors, *i.e.*, errors in estimating carrier phase and symbol timing which exist in actual modems but are difficult to include in computer simulation models. Although there are good reasons for omitting synchronization errors in channel simulations—notably the need to minimize computer running time—it is desirable to have the option to include these effects for certain simulation experiments.

This paper describes a technique that has been used to include the effects of synchronization errors in simulations of quaternary PSK (QPSK) transmission over the communications satellite channel [4]. The concept used in the channel modeling program (CHAMP) is reviewed briefly and the limitations of the current simulations are discussed. Next, a more complete model of the coherent demodulation process is presented, along with a brief description of the concept used in the RECEIVER program.

Two examples which are related to QPSK transmission over the satellite channel are described. The first was selected because extensive laboratory simulation data were available for this channel configuration. The goal was to duplicate the hardware channel characteristics as closely as possible in the software simulation so that the results [probability of bit error $P_b(\epsilon)$ vs E_b/N_o] could be compared on an absolute basis. The second example includes a more complicated model of the satellite channel and a different concept for filtering the QPSK signals in the modem. Conventional and offset QPSK transmissions are compared for this second channel configuration.

Time-domain simulation concept

Description of CHAMP

Time-domain simulations of communications channels utilize sampled or discrete-time representations of the modulated signals which are passed over the channel. The models use the complex envelope representation of

signals. For example, an unfiltered QPSK signal with power P , phase angle θ , and random modulation $\{A_i, B_i\}$ applied to the in-phase and quadrature components, respectively, would be generated as

$$S(k\Delta) = A_i\sqrt{P} \exp \left\{ j2\pi k \frac{f_c}{f_s} + j\theta \right\} + jB_i\sqrt{P} \exp \left\{ j2\pi k \frac{f_c}{f_s} + j\theta \right\} \quad (1)$$

where f_c is the carrier frequency and f_s is the arbitrary sampling frequency. The time index k in equation (1) denotes uniform time increments $\Delta \doteq 1/f_s$. The symbol duration, T_s , corresponds to an integer number M of these uniform intervals so that $T_s = M\Delta$.

Algorithms that simulate the effect of various sources of distortion in the real channel operate on the sampled signal. The effect of filtering, for example, can be simulated in three ways: by passing the signal through a complex digital filter algorithm, by convolving it with the impulse response of the filter, or by using the discrete Fourier transform (DFT). Generally, simulation programs employ more than one of these filtering techniques, depending on the description of the channel filters.

Channel nonlinearities are simulated by storing two mapping functions—input envelope to output envelope (AM/AM), and input envelope to output phase shift (AM/PM)—and using these functions to transform the input signal into the distorted output signal. Other sources of impairment, such as adjacent or cochannel interference, are easily included in the simulation by generating and filtering parallel signal sources and then adding these interfering signals at the proper point in the model.

The concept used in CHAMP [3] is typical of the approach used in these channel simulation programs. CHAMP deals with samples of the complex envelope of signals and uses a basic sampling increment, Δ . These samples are $Z^i \doteq X_k^i, Y_k^i, \dots$, where $k = 0, 1, \dots, N-1$, and i denotes a particular point in the simulation model. A fixed number (M) of these samples make up one transmitted symbol. For example, with QPSK transmission, each symbol might typically be represented by $M = 16$ complex samples. A given simulation run would consist of L symbols, where $L = 2,048, 4,096, \text{ or } 8,192$. With $M = 16$, these runs would consist of 32,768, 65,536, or 131,072 complex samples, respectively.

The channel simulation concept is shown in Figure 1. The first step generates N samples of the randomly modulated signal Z_1 . This signal is then passed through channel element C_{12} to produce signal Z_2 . If the channel element is a filter, for example, output Z_2 would be the filtered version of Z_1 . The original signal could be saved or its storage positions

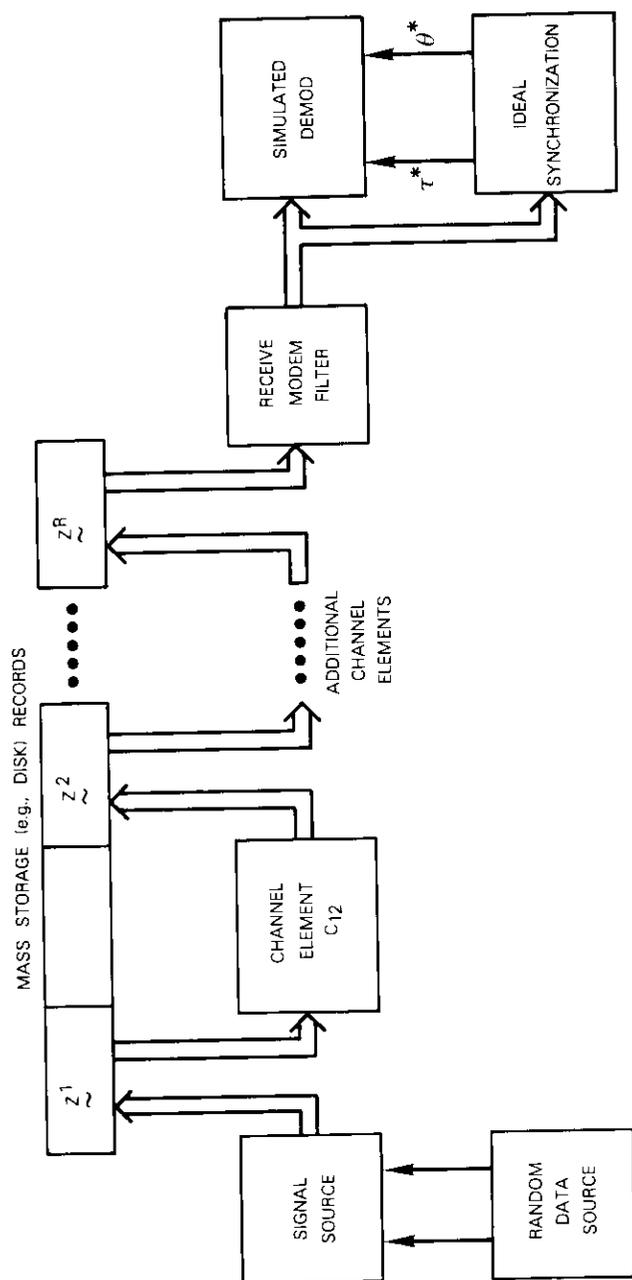


Figure 1. General Simulation Concept

(on a disk file) could be reused by signals generated later in the processing sequence.

The processing continues step by step through all channel elements in the particular model. A typical set might include a transmitter filter, a transmitter HPA nonlinearity, a satellite input filter, a satellite TWTA nonlinearity, a satellite output filter, and finally a receive modem filter.

Determination of average error rate

The technique used to determine the average error rate, $\overline{P_b(\epsilon)}$, in CHAMP will be reviewed briefly, since an extension of this technique is used to include the effect of synchronization errors. The total data record, containing N complex samples, is generated at the transmitter in Figure 1 and passed through the channel elements until it reaches the input to the receive modem filter. At this point in the real channel, the down-link noise, with density N_n W/Hz, would be added to the received signal. The received signal has also experienced distortion and envelope delay due to the channel nonlinearities and the elements with memory (filters). Two approaches can be used to simulate the coherent demodulation of the received signal:

- Direct simulation approach.* Gaussian random numbers are added to the signal; the signal is demodulated, and bit errors are counted.
- Hybrid simulation analysis approach.* The noise is omitted and the signal is demodulated to observe the statistics of the X and Y samples due to distortion and intersymbol interference only.

The second approach is used in CHAMP. This technique is valid for the class of coherent receivers for which the demodulation can be modeled as a linear process. When the detection is linear, it is valid to omit one of the noise sources (thermal noise) and to then superimpose this random variable on the statistical results obtained from the simulation. This technique results in a considerable saving in computer time, since an entire curve of $P_b(\epsilon)$ vs E_s/N_n can be obtained from a single simulation run. With the error counting technique, different runs must be made for each value of E_s/N_n . In addition, this approach becomes impractical at low error rates.*

*To obtain reasonably accurate results at a particular value of $P_b(\epsilon)$, at least 100 errors should be observed, requiring a run length of $100/P_b(\epsilon)$ bits. At $P_b(\epsilon) = 10^{-4}$, runs of 10^6 bits (500,000 symbols) are required; these runs could require several hours on a moderately fast computer.

The hybrid simulation analysis approach results in more efficient use of computer time relative to the direct simulation approach in which noise is added to the received waveforms and decision errors are counted. To perform these hybrid simulations, estimates of symbol timing and carrier phase must be established at the receiver to determine the exact point at which the complex waveforms entering the receiver should be sampled, and the orientation (phase shift) of the complex sample pair with respect to the decision axes. Given the phase of the received carrier and the timing epoch within each symbol, the waveforms out of the receiver matched filter can be processed to generate noise-free eye diagrams as shown in Figure 2a. Successive pairs of samples from the in-phase and quadrature channels of the demodulator (X and Y , respectively) can be plotted as a scatter diagram as shown in Figure 2b by sampling the received waveforms at the timing epoch.

The eye diagrams in Figure 2a are simply a superposition of the waveforms out of the I and Q demodulator channels on the interval $0-T_s$. Due to filtering and other sources of distortion in the channel, these waveforms are spread somewhat and do not produce a single overlapping trace. However, as shown in Figure 2a for the negative values on the X (in-phase) channel, with good control of intersymbol interference, the waveforms can exhibit very small variation at one particular time value, ϕ' . In this case, it can be seen that the spread in the sampling values is minimized if sampling occurs precisely at ϕ' . The spread, or distortion, is defined as either the difference between the maximum and minimum sample values:

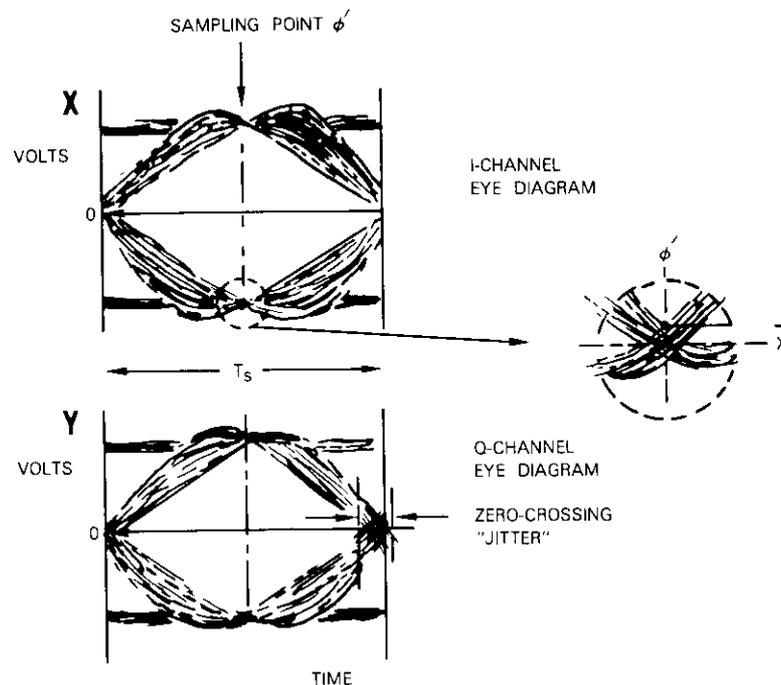
$$S_1(X, \phi') \doteq |X_{\max}(\phi')| - |X_{\min}(\phi')| \quad (2)$$

or the variance of the sample values:

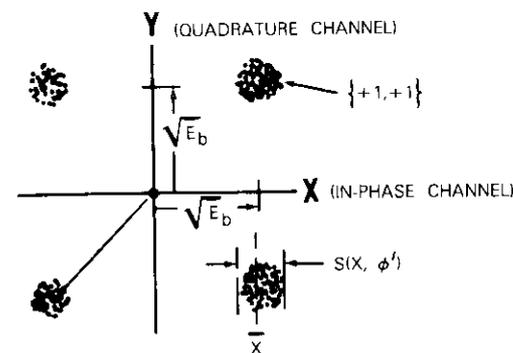
$$S_2(X, \phi') \doteq \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2 \quad (3)$$

where \bar{X} denotes the mean or expected value of the samples. For either definition of distortion, a symmetrical situation should exist in which the spread is the same at the positive and negative sampling instants of both eye diagrams.

Figure 2b, which is a plot of the scatter of the X and Y values of each symbol at a particular sampling instant ϕ' , provides a useful condensation of the information in the eye diagram. In this figure the $[X_i, Y_i]$ pairs are plotted for a large number of symbols, $i = 1, 2, \dots, L$. The scatter points



a. Eye Diagram



b. Scatter Diagram

Figure 2. Qualitative Indicators of Channel Performance

pair together samples from both the X and Y channels on a given symbol, information which is not easily obtained from the eye diagrams. Although the scatter diagrams do not include the detailed time history or the "trajectories" of the waveforms through the sampling points, they are extremely useful as qualitative indicators of channel performance. Note that much of the information in the eye diagrams, such as the mean of the sample values, \bar{X} and \bar{Y} , and the distortions, is retained in the scatter diagrams.

To obtain the eye and scatter diagrams from the simulated demodulator, the data Z_R in Figure 1 must be passed through the receive modem filter and then processed to determine two synchronization values: an average signal phase, θ' , over the total record; and an average signal zero-crossing time, ϕ' , over the total data record. The procedure used in CHAMP is one means of obtaining the necessary synchronization. First, the complex waveforms out of the receiver matched filter are observed to sense zero crossing, and an average zero-crossing epoch is established by observing over the entire record of N samples. An idealized† sampling point (τ^*), which is determined midway between the average zero crossings, is used to extract L complex samples of the waveforms, one for each symbol. These L complex samples are further processed to determine an average phase angle (θ^*) over the total record, which is then removed from each sample. This procedure attempts to center the scatter diagram in the decision region; with QPSK, for example, it should result in four symmetrical scatter diagrams located at 45° , 135° , 225° , and 315° in the decision region.

With the received waveforms reduced to a set of L complex samples, and these sample pairs rotated to give a symmetrical scatter diagram in the four quadrants of the decision region, an average error probability, $\bar{P}_b(\epsilon)$, must be determined based on the scatter of the received points. Specifically, the effect of thermal noise is introduced analytically by numerically integrating over the scatter diagram to determine an average error probability over all received symbols. This procedure assumes that the perturbations in the decision region are due to two independent causes: the signal distortion caused by the channel, and the thermal noise introduced at the input to the receiver. The known statistical fluctuations due

†Idealized in the sense that it is the best estimate of zero crossings given the limited set of discrete data points. The sampling point derived from this estimate may not minimize the error probability however.

to thermal noise are convolved with the unknown statistics due to channel impairments reflected in the scatter diagram by computing an average error probability as

$$\bar{P}_b(\epsilon) = \frac{1}{L} \sum_{l=1}^L P_b(\epsilon | X_l, Y_l) \quad (4)$$

The conditional error probability in equation (4) is computed for each X_l, Y_l pair as

$$P_b(\epsilon | X_l, Y_l) = \frac{1}{2} \left\{ \frac{1}{2} \operatorname{erfc}(p_x) + \frac{1}{2} \operatorname{erfc}(p_y) \right\} \quad (5)$$

$$\text{where } p_x = \frac{X_l}{\sigma}$$

$$p_y = \frac{Y_l}{\sigma}$$

$$\sigma = \left[\frac{P_R}{E_s/N_o} \right]^{1/2}$$

$$P_R = \bar{X}^2 + \bar{Y}^2 \text{ (average "received energy" per symbol) .}$$

The normalization described above is only one of several possible definitions that can be used to display the simulation results. Normalization and correction of the results are discussed in more detail for each example in a subsequent section of this paper. At this point, it is sufficient to note that equations (4) and (5) can be used together to compute $\bar{P}_b(\epsilon)$ for specific values of E_s/N_o . The calculation described above scales itself to the condition in which the average received energy per symbol is unity.* With a perfect channel (absolutely no scatter) and $P_R = 1$, equation (5) reduces to

$$P_b \left(\epsilon \left| \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right. \right) = \frac{1}{2} \operatorname{erfc}(E_b/N_o) \quad (6)$$

which is the theoretical performance of QPSK.

*Although this may seem to be an unnecessary detail, scaling, defining received signal levels, and "correcting" the simulation results are actually some of the more troublesome areas encountered in simulations of this type. More detailed definitions will be given later in this paper.

This calculation of average error rate is based on the following assumptions:

- a. all points in the scatter diagram are equally likely;
- b. the diagrams contain enough points, *i.e.*, the simulation runs are long enough that the scatter is a valid representation of the statistics of the distortion;
- c. Gaussian noise with zero mean and variance $N_o/2E_b$ appears independently on the X and Y symbols and is independent from symbol to symbol;
- d. synchronization errors are negligibly small.

Assumptions *a* and *b* can be satisfied in almost all practical situations encountered on satellite channels. In some situations in which intersymbol interference extends over long intervals (*e.g.*, eight to ten symbols) it may be necessary to use P_N sequences as a data source to ensure a balanced set of run lengths in the transmitted data. Observation of the buildup of the scatter diagram on the CRT is usually convincing proof that the diagram is in fact converging as required by assumption *b*.

Assumption *c* is not strictly true when the noise entering the receiver is filtered; when the filter bandwidth is fairly narrow, there will be some noise correlation from one sampling point to the next. This phenomenon could have a slight effect on the clustering of errors, but would have a negligible effect on the long-term average error probability, $\overline{P_b(\epsilon)}$.

Assumption *d*, which is generally not true, will be discussed further in the following subsection.

Limitations of the hybrid simulation approach

The general simulation procedure described in the preceding subsection does not represent the true operation of a coherent receiver since synchronization errors are not included. In the real receiver, the local estimates of symbol timing (sampling epoch) and carrier phase would be random variables which would fluctuate slowly from symbol to symbol. These fluctuations would be due to the thermal noise, which is present at the input to the receiver, and additional fluctuations could be caused by the distortion of the channel. These latter effects would be highly dependent upon the exact implementation of the receiver. For example, certain carrier tracking implementations could be sensitive to the AM/PM conversion experienced in the channel, and symbol synchronizers, particularly those that extract information from the zero crossings in the received waveforms, could be highly sensitive to intersymbol interference. To include

these effects, the particular receiver implementation must be modeled in detail.

As discussed in the following, there are good reasons for implementing the existing simulation program with idealized static synchronization. However, the omission of synchronization errors does limit the present program in that the results are optimistic. Thus, the simulation results must always be qualified by clearly stating that synchronization impairments are not included.

There is some evidence that certain synchronizer implementations may be quite sensitive to the signal distortion encountered in the nonlinear satellite channel. This sensitivity depends on the degree of filtering in the channel where tight, (*i.e.*, low- BT_s -product) filtering gives rise to severe envelope fluctuations on the signal. When these envelope fluctuations encounter the AM/AM and AM/PM nonlinearities in the channel, the total distortion can cause additional errors in the synchronization process. These additional errors can be caused by the pattern-dependent distortion which perturbs the error signals generated at the input to the synchronizer filters. Since these phenomena are known only qualitatively, detailed modeling of the coherent demodulation process is necessary to accurately determine the additional performance degradation due to these effects.

Before such a model is described, it is useful to summarize the reasons why it is difficult to include synchronizer performance directly in a simulation structure similar to that used in CHAMP. First, the present simulation program is used almost exclusively in a mode in which thermal noise is not added at the input to the receiver. Error probability is calculated "after the fact" by computing $P_b(\epsilon)$ vs E_s/N_o given the scatter of the detected symbols due to distortion and intersymbol interference. Since most synchronization techniques involve nonlinear operations, the superposition used in the hybrid approach will not be valid. Thus, a direct simulation (adding noise numbers) must be used to accurately simulate synchronizer performance.

Secondly, for almost all cases of interest, the time constants of the synchronizers are much longer than a symbol duration. For QPSK operation, averaging times of 50 to 200 symbols might typically be used in estimating symbol epoch and carrier phase. Direct simulation of the synchronization process could therefore require very long simulation runs to obtain statistically significant results. Finally, synchronizer performance could be strongly dependent on the actual receiver implementation that is to be investigated. To obtain results which would be valid for a particular hardware implementation, that implementation would have to be modeled in

detail. This level of detail was not included in the initial development of CHAMP.

More complete model of the coherent demodulation process

Model

Inclusion of the synchronization functions in time-domain simulations is complicated by the need to model the functions of the particular receiver implementation in detail. This requires detailed knowledge of the receiver, or at least knowledge of the concept used in the receiver. Carrier phase estimators for QPSK signals, for example, can use a quadrupling circuit followed by either a tuned filter or a phase-locked loop. Various configurations which employ data decisions to remove the modulation from the phase error signal are also possible; these configurations are included in the general category of decision-directed feedback techniques.

Similarly, there are many alternatives for symbol timing recovery, ranging from the delay-and-multiply techniques, which are relatively easy to implement, to the more efficient early-late tracking schemes, which may also, in some cases, utilize data decisions for modulation removal. In some implementations, the symbol timing and carrier phase estimation operations may be interrelated; this is particularly true of their acquisition behavior, which becomes important in TDMA applications. Therefore, to be generally useful, a simulation structure must be capable of handling different specific implementations. Such a structure has been developed, and has been used to examine the behavior of one particular implementation.

The general problem being considered can be described in terms of the receiver model shown in Figure 3. This description will be restricted to QPSK signals although the concepts could be extended to other techniques.

The three different operations shown in Figure 3 are described as follows:

- carrier phase recovery establishes a continuously varying estimate, $\hat{\theta}$, of the incoming signal phase, θ ;
- symbol timing recovery likewise extracts an estimate, $\hat{\phi}$, of the symbol clock phase, ϕ , which modulates the incoming signal;
- coherent demodulation includes multiplication of the incoming signal by sine and cosine components of the recovered carrier, sampling of the resulting baseband waveforms, and comparison of

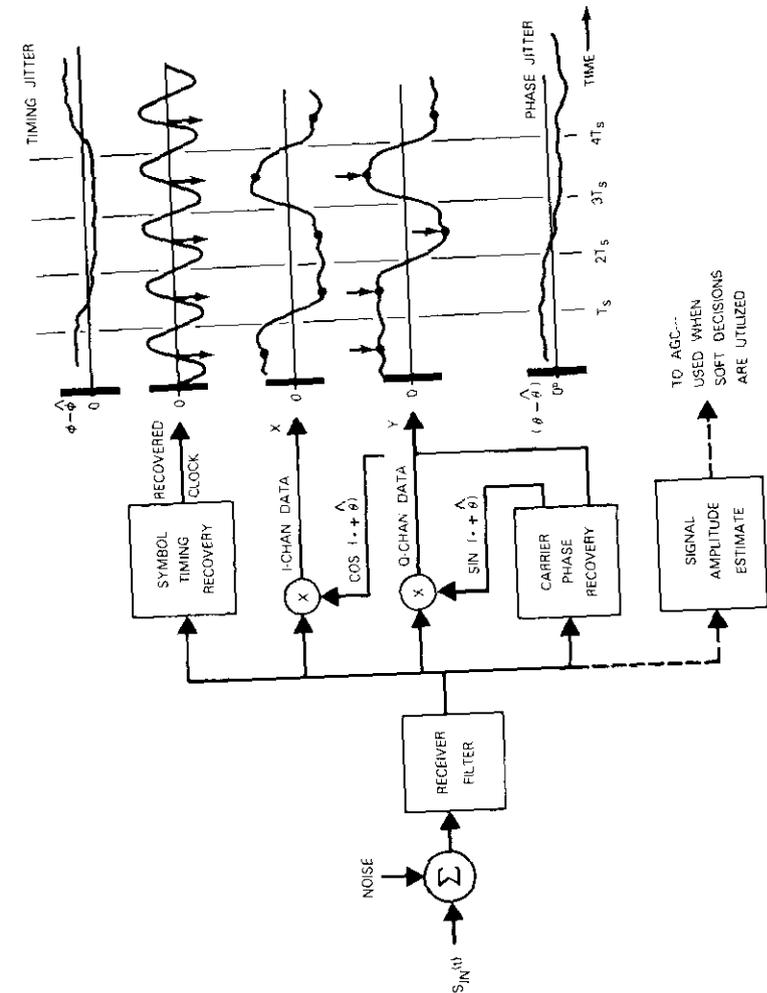


Figure 3. Receiver Model

these samples with thresholds to make decisions on the received symbols.

Figure 3 implies a specific method of symbol timing recovery in which a sine wave clock is recovered from the modulated signal; the positive-going zero crossings in this waveform are used to designate the point at which the data waveforms should be sampled. Conceptually, all forms of timing recovery must produce this same information, although different implementations may vary considerably from this one.

The waveforms shown in Figure 3 contain most of the information required to compute $\overline{P_b(\epsilon)}$. Although data waveforms X and Y are shown as noise free, the noise at the input to the receiver appears as an additive perturbation to the waveform which is inversely proportional to E_s/N_o .

Effect of phase errors

The carrier phase recovery operation shown in Figure 3 utilizes the distorted signal plus noise to establish an estimate, $\hat{\theta}$, of the phase of the incoming signal. To establish this estimate with the small error required for demodulation of QPSK signals, the averaging time of the phase estimator would be much longer than the symbol duration, T_s . Equivalently, the bandwidth of the phase-locked loop or the tuned filter used to estimate phase would be selected to be much less than the symbol rate. With QPSK, averaging times of the synchronizer would typically be in the range of 50 to 200 times T_s . Consequently, the random process $(\theta - \hat{\theta})$, where θ is the true carrier phase and $\hat{\theta}$ the estimate of θ produced by averaging the phase estimate over many symbols, would fluctuate slowly relative to T_s . This is shown in Figure 3 where the phase error varies slowly relative to the symbol duration. As a result, $\theta - \hat{\theta}$ is highly correlated from symbol to symbol; when the error is small, it would tend to remain small for many symbols, or if the error became large, many symbols could be affected by increased error probabilities. As argued earlier,* however, this phenomenon should affect only the slight clustering of errors and have a negligible effect on long-term average error probability.

Consider the effect of a phase error $(\theta - \hat{\theta})$ on the detection of a single data symbol, which can be expressed as the complex number

$$Z_l = \sqrt{E_b} A_l + j\sqrt{E_b} B_l \quad (7)$$

*Cycle skips would require a modification of these procedures.

where $\{A_l, B_l\}$ are the random data, with each component randomly assuming values of ± 1 . The phase error effectively rotates the received vector by $\theta - \hat{\theta}$ degrees to produce the vector

$$\begin{aligned} Z_l &= Z_l \exp \{j(\theta - \hat{\theta})\} \\ &= \sqrt{E_b} [A_l \cos(\theta - \hat{\theta}) - B_l \sin(\theta - \hat{\theta})] \\ &\quad + j\sqrt{E_b} [B_l \cos(\theta - \hat{\theta}) + A_l \sin(\theta - \hat{\theta})] \end{aligned} \quad (8)$$

This rotation is shown in Figure 4a for the case in which $\theta - \hat{\theta}$ is a small positive value of 10° . When the mean phase error is zero and the data are random, the symmetry of the problem makes it possible to deal only with the first quadrant and to compute an average probability as

$$\overline{P_b(\epsilon)} = \int_{-\pi}^{\pi} \overline{P_b(\epsilon|\theta_e)} P(\theta_e) d(\theta_e) \quad (9)$$

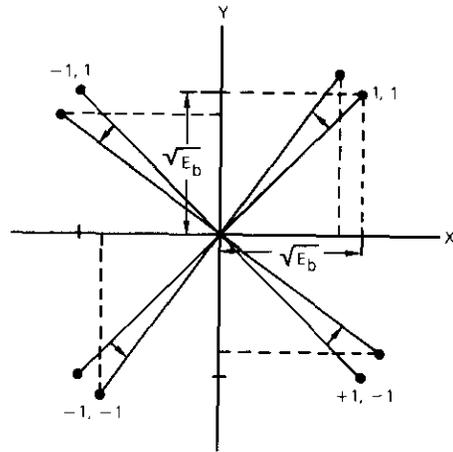
where $\overline{P_b(\epsilon|\theta_e)}$ = probability of bit error given a particular value of phase error $\theta_e \doteq \theta - \hat{\theta}$

$P(\theta_e)$ = probability density function for the phase error,
 $-\pi < \theta_e \leq +\pi$

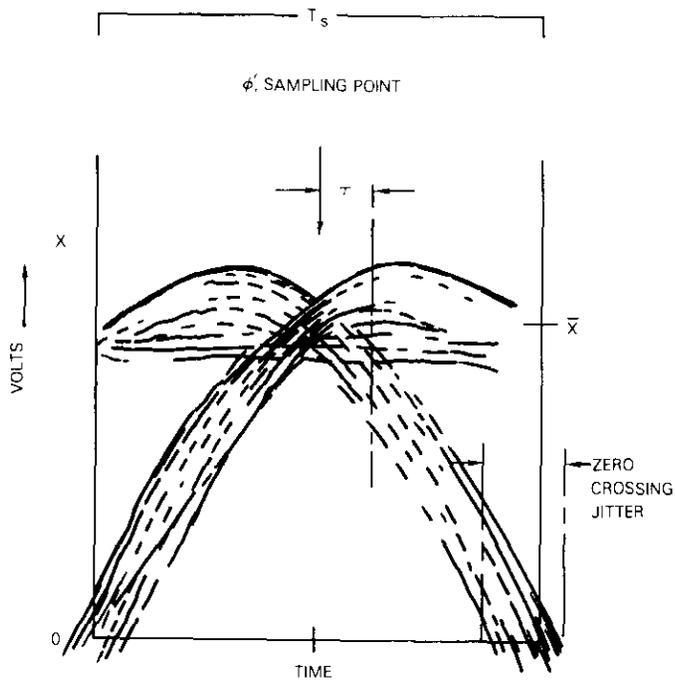
Equation (9) is a commonly used procedure* for determining the effect of an imperfect phase reference on PSK transmission. The integration must be performed numerically and a probability density function must be assumed for the phase error [4], [5].

The RECEIVER simulation program is structured to model a particular phase estimator and to collect statistics to experimentally determine $P(\theta_e)$. These phase errors consist of two components, one due to the thermal noise at the input to the receiver, and a second due to the jitter on the detected symbols due to distortion and intersymbol interference. The second component is difficult to describe, since the effect of the data-dependent distortion and interference on a phase estimator must be determined experimentally. Qualitative evidence of this source of error is observed in all of the CHAMP simulations as a scatter of the detected symbols in the radial direction. This pattern-dependent phase jitter, as observed

*Again, the effect of cycle skipping is neglected in this procedure.



a. Effect of Phase Errors on QPSK Detection



b. Effect of Symbol Timing Errors
Figure 4. Errors in Synchronization

in the scatter diagrams, represents an effective phase error on each symbol. The effect of these individual errors is reduced by the averaging time used in the phase estimator.

Effect of timing errors

The effect of carrier phase errors is fairly easy to visualize from the scatter diagrams as a random, slowly-varying rotation of the received vector as shown in Figure 4a. Timing errors are somewhat more difficult to visualize; hence, it is necessary to return to the eye diagram shown in Figure 2a to describe this effect. One such diagram is expanded as Figure 4b, which is a symmetrical diagram for the positive voltage values of the X detector channel. In this case, the symmetry holds about the sampling point ϕ' so that positive and negative time errors with respect to ϕ' should have the same effect. Examination of the effect of a time error on any given waveform reveals that the waveform can increase, decrease, or remain approximately the same relative to the sample value at time ϕ' .

A straightforward approach to determining error rate for a fixed time error is simply to sample the waveforms on both the X and Y channels with a fixed offset τ to obtain a scatter diagram that, in the case of Figure 4b, would exhibit much more scatter than is experienced at the ideal sampling point, ϕ' . This capability exists in CHAMP; that is, the error rate can be computed for a single fixed offset timing point to determine $P_b(\epsilon|\phi' + \tau)$, which is an average error rate given a fixed timing error, τ . Without phase errors, it is reasonable to extend this technique to obtain an average error probability as follows:

$$\overline{P_b(\epsilon)} = \int_{-T_s/2}^{T_s/2} \overline{P_b(\epsilon|\phi_\epsilon)} P(\phi_\epsilon) d\phi_\epsilon \quad (10)$$

where $\overline{P_b(\epsilon|\phi_\epsilon)}$ = average bit-error rate given the timing error $\phi_\epsilon \triangleq (\phi' - \tau)$

$P(\phi_\epsilon)$ = probability density function for the timing error.

Similar to phase errors, the total timing jitter consists of two parts, one due to thermal noise and the other due to signal distortion and intersymbol interference. The signal-dependent timing jitter is evident in the expanded eye diagram of Figure 4b. Even without noise, intersymbol interference and distortion due to AM/PM conversion can cause a noticeable jitter, or fluctuation, in the zero crossings of the superimposed waveforms from symbol to symbol. Since certain types of timing estimators

utilize this zero-crossing information, they must average these fluctuations. Therefore, given the practical limit of finite averaging time, there will always be some error* in timing even without noise. The phenomenon analogous to cycle skipping, in which the symbol epoch estimator adds a false zero crossing or skips a valid crossing to result in a loss of bit count integrity, will be neglected in this treatment. Extremely long simulation runs would be required to observe this phenomenon.

CHAMP allows demodulation with both a single fixed phase error and a single fixed timing offset. Thus, although the approach may be impractical, it is straightforward conceptually to combine equations (9) and (10) to give

$$\overline{P_b(\epsilon)} = \int_{-\pi}^{\pi} \int_{-T_s/2}^{T_s/2} \overline{P_b(\epsilon|\theta_\epsilon, \phi_\epsilon)} P(\theta_\epsilon, \phi_\epsilon) d\theta_\epsilon d\phi_\epsilon \quad (11)$$

where $\overline{P_b(\epsilon|\theta_\epsilon, \phi_\epsilon)}$ is the average error probability given a single fixed phase error, θ_ϵ , and a single fixed timing error, ϕ_ϵ . The average is over the X-Y plane for all points in the scatter diagram using equations (4) and (5). $P(\theta_\epsilon, \phi_\epsilon)$ is the joint density function for the random phase and timing errors, which may be correlated for some implementations.

Equation (11) requires a 4-dimensional numerical integration. Since excessive computation time may be required, some caution is needed in adopting this approach. As a concept, however, equation (11) indicates a general solution to the problem. A practical implementation of this approach is discussed in the next subsection.

Approach used to obtain $\overline{P_b(\epsilon)}$

As an alternative to direct implementation of equation (11), the following procedure has been used to process the simulation data to include the effect of synchronization errors:

a. *Step 1.* A complete channel simulation is run for N symbols with the data saved prior to the receive modem filter. These data will be designated as Z_R ; typically, they will consist of 64,000 complex samples.

b. *Step 2.* Noise-free demodulation is performed by passing data Z_R through the receiver filter. An average error probability is com-

*The possibility of "equalizing" the estimators to remove the effects of distortion and intersymbol interference is not considered herein. The feasibility of this procedure warrants further investigation.

puted for a single value of E_s/N_o with five assumed synchronization conditions: $\{\tau^*, \theta^*\}$, $\{\tau^* + \tau_o, \theta^*\}$, $\{\tau^* - \tau_o, \theta^*\}$, $\{\tau^*, \theta^* + \theta_o\}$, and $\{\tau^*, \theta^* - \theta_o\}$. From these five measurements, a 2-dimensional parabolic function which gives the sensitivity of average probability of error to phase and timing errors is fitted to the data. This empirical function, $P(\epsilon|\tau_\epsilon, \theta_\epsilon)$, is

$$P(\epsilon|\theta_\epsilon, \tau_\epsilon) = 10^{(C_0 + C_1\tau_\epsilon + C_2\theta_\epsilon + C_3\tau_\epsilon^2 + C_4\theta_\epsilon^2)} = 10^{G(\theta_\epsilon, \tau_\epsilon)} \quad (12)$$

$$G(\theta_\epsilon, \tau_\epsilon) \doteq \log_{10} [P(\epsilon|\theta_\epsilon, \tau_\epsilon)]$$

Substituting for the fixed offset, τ_o and θ_o , makes it possible to obtain the constants C_0, C_1, \dots, C_4 :

$$C_0 = G(0, 0)$$

$$C_1 = \frac{G(0, \tau_o) - G(0, -\tau_o)}{2\tau_o}$$

$$C_2 = \frac{G(\theta_o, 0) - G(-\theta_o, 0)}{2\theta_o}$$

$$C_3 = \frac{\frac{1}{2}G(0, \tau_o) + \frac{1}{2}G(0, -\tau_o) - G(0, 0)}{\tau_o^2}$$

$$C_4 = \frac{\frac{1}{2}G(\theta_o, 0) + \frac{1}{2}G(-\theta_o, 0) - G(0, 0)}{\theta_o^2}$$

c. *Step 3.* Simulations are performed by adding noise to the received data, putting both signal and noise through the receiver filter, and then processing these data to estimate symbol epoch and carrier phase. This step results in two random time records, $\hat{\theta}_i$ and $\hat{\tau}_i$, which give the phase and timing estimates for each symbol. Since the averaging times for these two parameters, $1/B_p$ and $1/B_c$, respectively, are long relative to the QPSK symbol duration, estimates vary slowly relative to the symbol duration. These simulations are long enough to ensure good confidence in the measured statistics of these variables and generally require many repetitions of the data sequence Z_R with different Gaussian random sequences added for each repetition. A 2-dimensional histogram formed from these results serves as an estimate of

the joint density function, $P(\theta_\epsilon, \tau_\epsilon)$. One such histogram is shown in Figure 5, where the general tilt indicates some correlation between phase and timing errors.

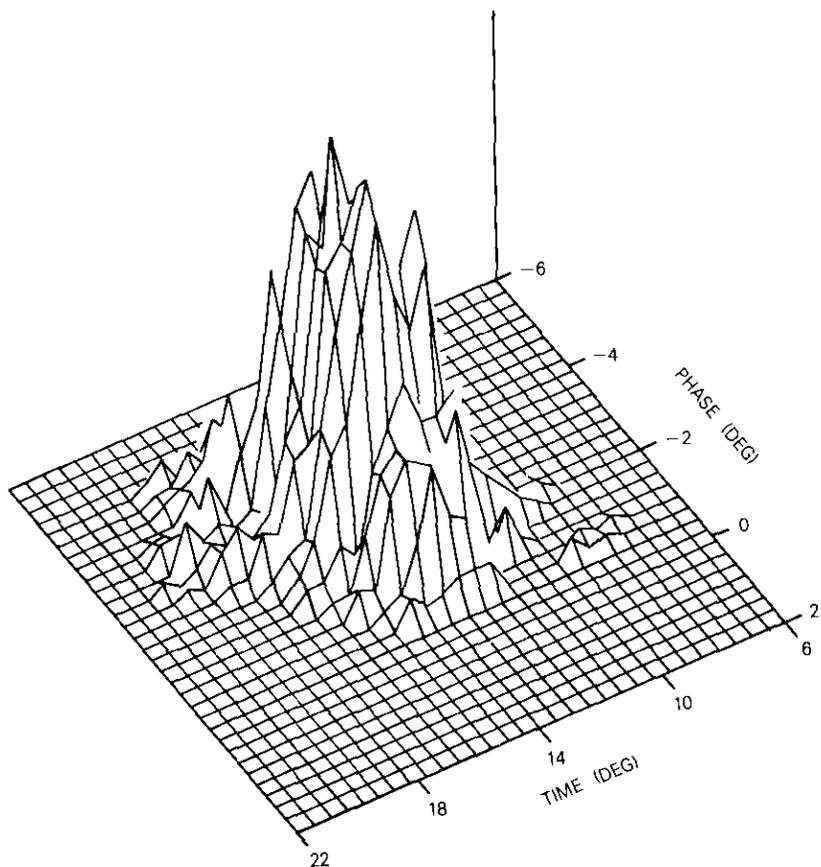


Figure 5. Typical Histogram of Time and Phase Errors $P(\theta_\epsilon, \tau_\epsilon)$

d. Step 4. A numerical integration is performed over each cell of the synchronization error histogram to determine the average error rate as $P_b(\epsilon) = \sum P(\epsilon|\tau_\epsilon, \theta_\epsilon) P(\theta_\epsilon, \tau_\epsilon)$, where the summation is over all cells in the histogram, $P(\theta_\epsilon, \tau_\epsilon)$, and the error probability conditioned on $\tau_\epsilon, \theta_\epsilon$ is obtained from the sensitivity function [equation (12)].

This general procedure is summarized in Figure 6. The error rate sensitivity data (step 2) can be obtained from a single pass through the noise-free channel output Z_R . The synchronization error histograms (step 3) must be obtained for specific values of E_b/N_0 ; each of these histograms generally requires multiple repetitions of the data Z_R with different noise sequences added for each repetition.

Simulation model of a particular modem

A simulation model which contains the functional elements shown in Figure 7 has been developed to accurately model the major functions of a particular modem implementation. In this particular configuration, a clock is recovered from the received signal by first delaying the output of the receive modem filter by one half of the symbol duration ($0.5T_s$). This delayed signal is multiplied by the filter output, and the product is applied to a filter which is tuned to the clock frequency, R_s Hz.* This filter has a bandwidth $B_c \ll R_s$ so that the recovered clock signal is smoothed or averaged over a duration of many symbol intervals. The filter output is processed further to detect positive-going zero crossings in the sinusoidal output; these zero crossings provide timing epochs at which the I and Q channel waveforms are sampled.

The received signal is also processed in a quadrupling circuit which removes the QPSK modulation. This output is applied to a filter tuned to the fourth harmonic of the carrier frequency (in practice, the fourth harmonic of an intermediate frequency). This filter has a bandwidth B_p , which is also much less than the symbol rate. The divide-by-four operation produces an estimate of carrier phase, $\hat{\theta}$.

Figure 8 shows the simulation model of the configuration in Figure 7. The operation of the symbol timing recovery circuit, shown in the upper part of the figure, will be discussed first. It will be assumed that noise and distortion are not present so that the complex signal out of the receive filter can be represented as sampled rectangular waveforms with M samples per symbol. Signal Z_F can be written as

$$Z_F = A_{l+1} \exp(j\theta) + jB_{l+1} \exp(j\theta) \quad (13)$$

for the M samples that constitute the $l+1$ symbol. The elements of the binary pair A_{l+1}, B_{l+1} , which represent random modulation in the I and

*The discussion in this section applies to conventional QPSK. Modifications of the model to simulate clock recovery for offset QPSK are minor.

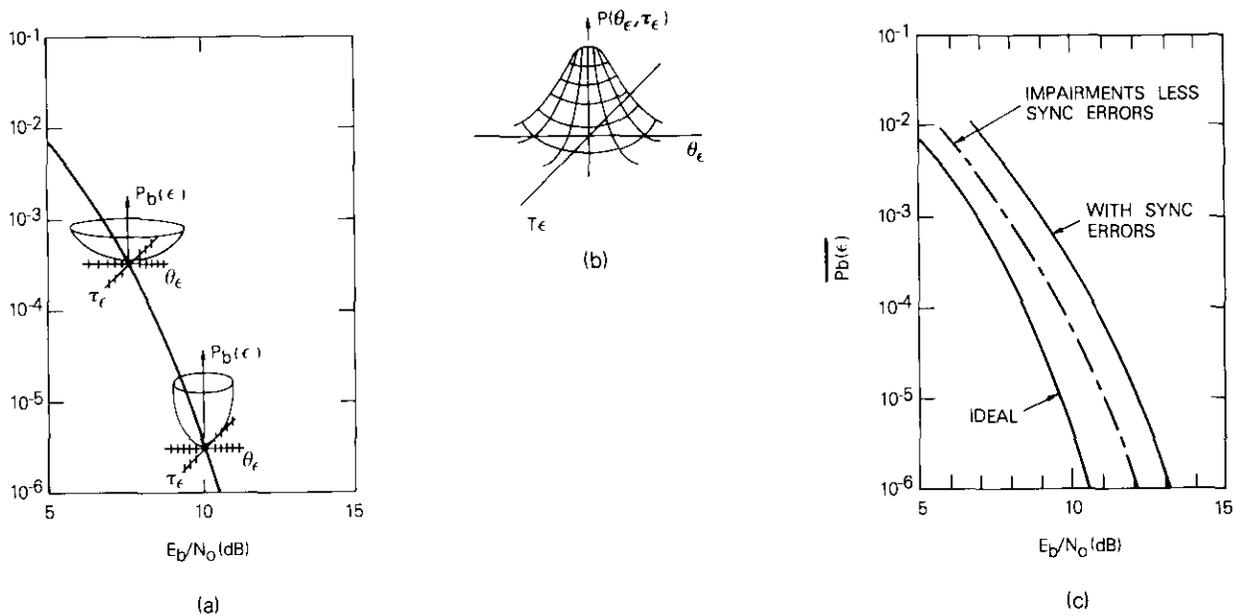
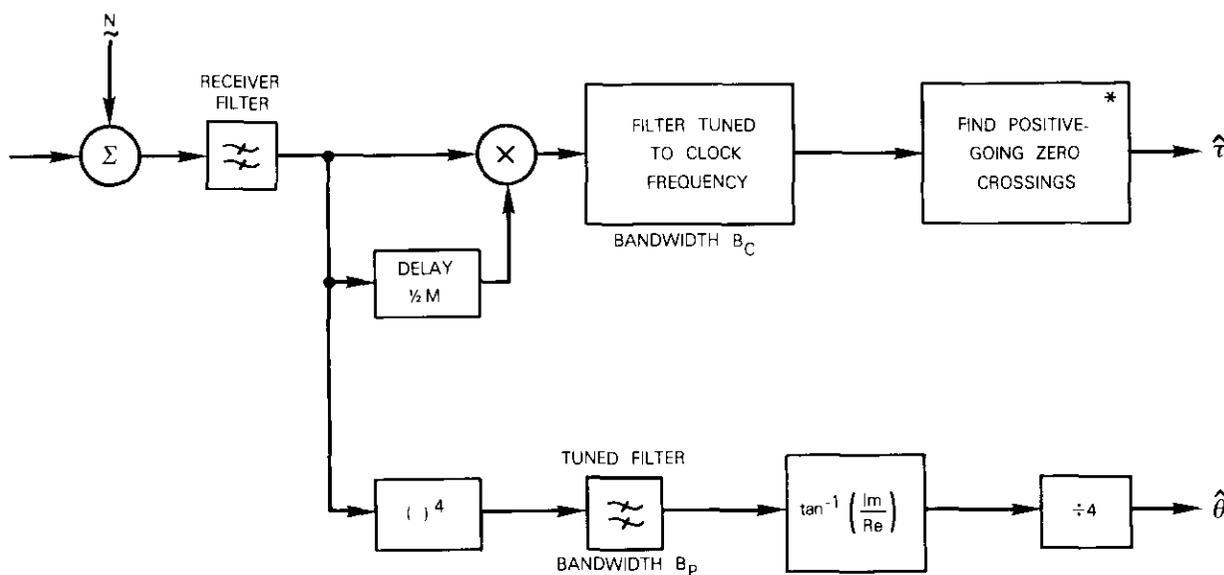
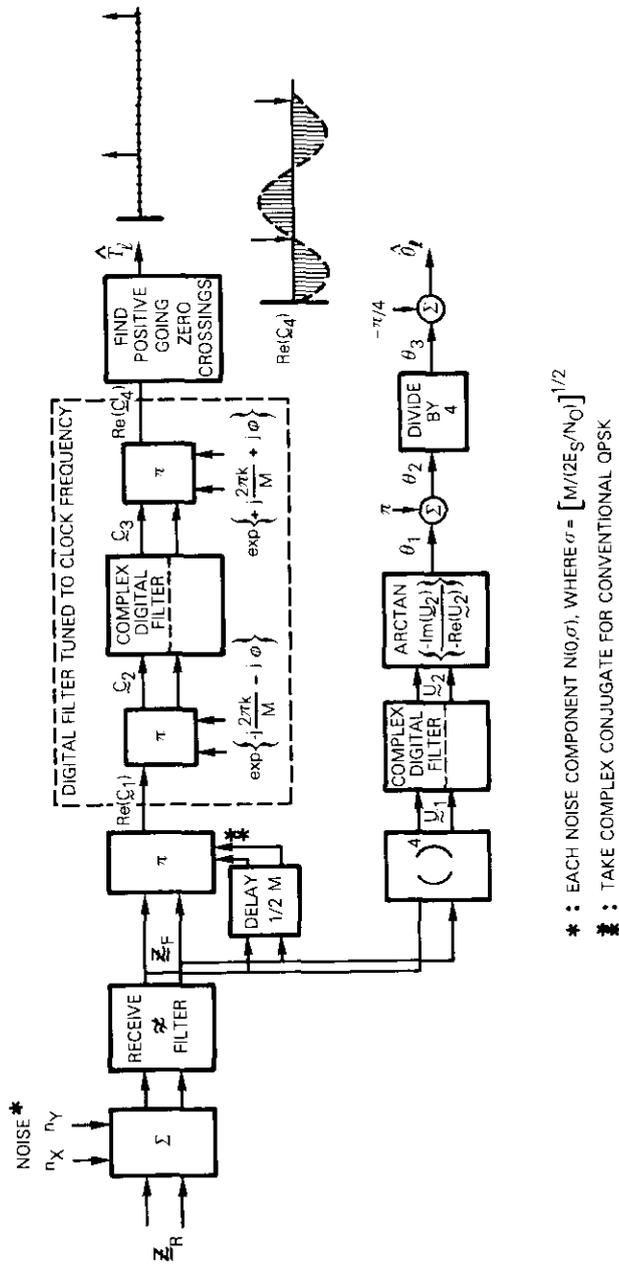


Figure 6. Data Obtained from Additional Processing of Z_R
 a. Sensitivity of $P_b(\epsilon)$ to Synchronization Errors
 b. Statistics of Phase and Timing Errors
 c. Average $P_b(\epsilon)$



* IN PRACTICE, COULD BE IMPLEMENTED BY A HARDLIMITER/DIFFERENTIATOR/HALF-WAVE RECTIFIER IN CASCADE

Figure 7. Model of Symbol Timing and Carrier Phase Synchronizers



* : EACH NOISE COMPONENT $N(0, \sigma)$, WHERE $\sigma = [M/12E_s/N_0]^{1/2}$
 ** : TAKE COMPLEX CONJUGATE FOR CONVENTIONAL QPSK

Figure 8. Simulation Model for Carrier and Clock Recovery

Q channels, respectively, assume random ± 1 values from symbol to symbol. Angle θ , which represents the arbitrary phase of the carrier at the output of the receiver filter, is unknown but fixed.

Signal Z_F is delayed by $M/2$ samples to produce the delayed signal Z'_F , where

$$Z'_k = A_k \exp(j\theta) + jB_k \exp(j\theta), k = 1, 2, \dots, \frac{M}{2}$$

$$= A_{k+1} \exp(j\theta) + jB_{k+1} \exp(j\theta), k = \frac{M}{2} + 1, \dots, M \quad (14)$$

The sampled signals are shown in Figure 9, where it is assumed for clarity that $M = 8$. For conventional QPSK, shown at the left of this figure, the output of the delay and multiply operation is

$$Re\{C_1\} \doteq Re\{Z_R \cdot Z_R^*\} \quad (15)$$

where $*$ denotes the complex conjugate. Note that this output has a constant value of $+2$ over the second half of the symbol interval, but the value over the first half can assume values of $+2, 0$, or -2 , depending on whether zero, one, or two transitions occur in the data. For offset QPSK, the complex conjugate operation is omitted in equation (15) to yield the product function shown at the right of Figure 9. Again, the amplitude of the clock component depends on the transitions in the data.

The important observation about the waveforms in Figure 9 is that the clock component maintains a constant phase even though its amplitude is a random variable depending on the occurrence of transitions (from $+$ to $-$ or from $-$ to $+$) in the data sequence on the two quadrature channels. In the actual simulation the processed waveforms are distorted and depart considerably from the ideal rectangular shape shown in Figure 9. In addition, the waveforms are perturbed by the noise added at the input to the receiver filter.

The tuned filter is simulated as a combination of down-conversion, low-pass complex filtering, and up-conversion which gives an output sampled sine wave $Re\{C_1\}$ with a phase angle that reflects a smoothed or averaged version of the phase difference between the symbol timing on input signal Z_F and the phase (ϕ) used in the up- and down-conversion. The digital filtering simply implements separate identical recursive filtering of the real and imaginary components of C_2 to produce C_3 . For example, the

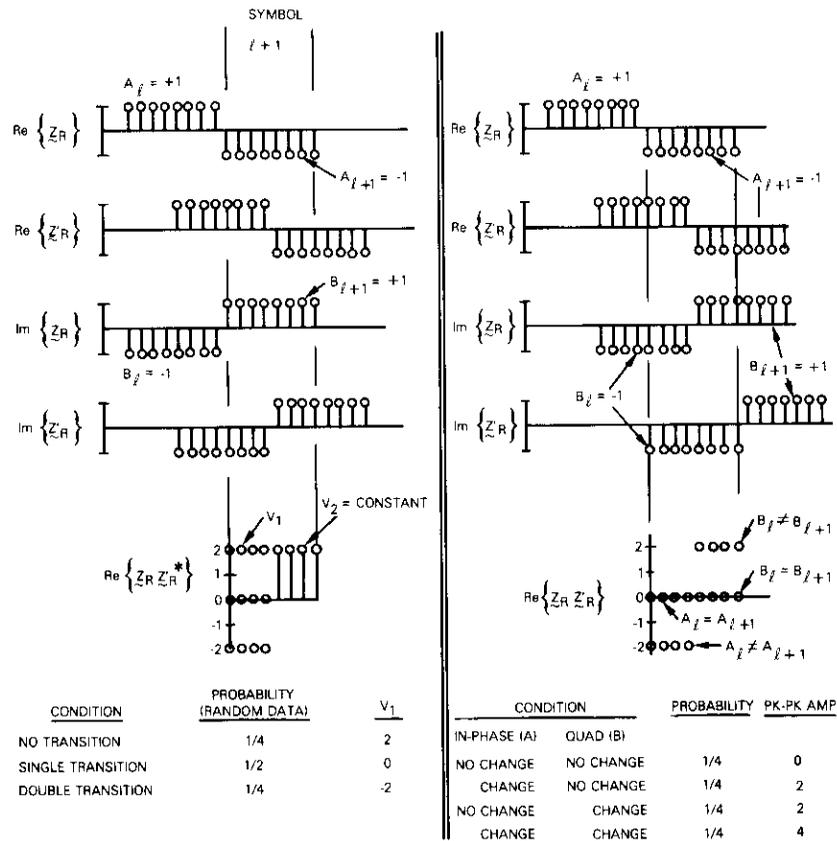


Figure 9. Delay and Multiply Output over One Symbol

real part of C_3 on the k th sample is obtained as

$$Re\{C_3(k\Delta)\} = (1 - \alpha) Re\{C_3(k\Delta - \Delta)\} + \alpha Re\{C_2(k\Delta)\} \quad (16)$$

OR

$$Re\{C_3(k\Delta)\} = \alpha \sum_{i=0}^k (1 - \alpha)^i Re\{C_2[(k - i)\Delta]\} .$$

The parameter α is determined by the bandwidth of the filter, B_c :

$$\alpha = \frac{2\pi(B_c/2)}{MR_s} \quad (17)$$

where B_c = total 3-dB bandwidth of the bandpass filter
 M = number of samples/symbol
 R_s = symbol rate.

For example, if a tuned filter with a total 3-dB bandwidth of 500 kHz is to be simulated and the symbol rate is 40×10^6 symbols/second, then α would be selected as

$$\alpha = \frac{2\pi(500 \times 10^3/2)}{16 \times 40 \times 10^6} = 0.00245$$

and $1 - \alpha = 0.9975$.

The real part of the filter output is a negative sine wave which is cyclic at the incoming symbol rate. On the average, the positive-going zero crossings in $Re\{C_4\}$ coincide with the center of each QPSK symbol. An exact sampling point, $\hat{\tau}_i$, can be determined on each symbol by sensing these zero crossings, and interpolating to determine the exact zero-crossing point between discrete samples. This point is a random variable from symbol to symbol because of the random fluctuations in the zero crossings of the demodulated (*i.e.*, baseband) waveform and the noise on the input waveforms.

The simulated carrier phase estimation process shown in Figure 8 also uses data samples Z_F . The first step is a sample-by-sample quadrupling operation,

$$U_1(k\Delta) = [Z_F(k\Delta)]^4$$

where Z_F is given by equation (13). This operation is performed in the program by raising the magnitude of each complex sample to the fourth power and multiplying the phase angle by four. The two components of U_1 are filtered separately in 1-pole filters, where the parameter α is obtained from equation (17) by replacing B_c with B_p . Filter output U_2 is simply a smoothed or averaged version of U_1 . After arbitrary shifting and division by four, the phase of U_2 yields an estimate of the carrier phase, $\hat{\theta}_i$, for each sample. One of the M phase estimates for each symbol is selected as the local phase estimate on the n th symbol.

The model in Figure 8 is used to process Z_R to yield the statistics of $\hat{\theta}_i$ and $\hat{\tau}_i$. Since noise is added at the input to the receiver filter, this process

requires multiple (typically 10 to 50) repetitions of Z_R with different Gaussian random numbers added for each repetition. Fewer repetitions are required as E_b/N_o is increased because the pattern jitter rather than thermal noise eventually dominates the fluctuations in $\hat{\theta}_i$ and $\hat{\tau}_i$.

General concept of the RECEIVER program

The RECEIVER program has been developed to evaluate the performance of various RECEIVER structures with signals that have been subjected to various satellite link impairments (*i.e.*, adjacent channel interference, cochannel interference, intermodulation distortion, and intersymbol interference). RECEIVER uses data that have been generated and stored by CHAMP. This link between the two programs allows RECEIVER to utilize the full simulation capabilities of CHAMP.

Design objectives

The design objectives which guided the development of the software are summarized as follows:

a. Modularity. The program is designed in a modular fashion to facilitate the addition of new techniques and capabilities.

b. Ease of use. The user communicates with the program by means of a mnemonic-oriented unformatted command language. Commands and parameter values are interpreted by the program; parameters are checked when possible, and the command and all pertinent data are stored in a table. The user may also preface certain commands with specific directives such as equalize and/or output. When the program senses that all commands have been entered by recognizing an END command, execution commences. All processed data are stored in core-resident units. To specify a particular data record, the user merely references the output unit number of the module that processed that data.

c. Task-oriented definable architecture. A library of basic operations, or modules, allows the user to simulate the effects of each module in the receiver. The user is allowed complete freedom in specifying the receiver architecture, including such factors as the number and type of components and the specific interconnection between them. He may also specify that a selected group of commands will form a task. As many as 10 tasks may be defined in each computer run. Within each task any combination of modules and interconnections may be chosen. In addition, repetitive looping operations can be implemented so that

the same set of commands within a particular task may be exercised over and over. This has proved to be invaluable for probability histograms in which the input noise must be averaged over many values. Another major advantage of the task concept is that it allows the user to compute values for a set of parameters in one task, and to then use those parameters in another task.

d. User-constructed library. Since many RECEIVER configurations are almost identical (perhaps differing in bandwidths or values of E_b/N_o), each user may store particular configurations in a library file. Each configuration is given a unique name and can be retrieved by RECEIVER when the user inputs that name.

Program architecture

As mentioned above, RECEIVER is implemented as a command-driven modularized program. Each command that the user inputs activates a particular module (algorithm). Numerical parameters entered with the command indicate to the module where to get its input data and where to send its processed data, and control particular parameters of the module. There are three temporary data units used to transport data from module to module.

Simulation results

The simulation procedure described in the second section of this paper has been applied, using the modem implementation described in the preceding section, to two representative models of the communications satellite channel. For the first channel model, the results were carefully compared to laboratory measurements using a 60-Mbps modem [6], [7].

Experiments using a simple channel model

The simulation model used for this first series of experiments is shown in Figure 10. The transmitter filter (F_1) is an approximation of an "inverse $\sin x/x$ " filter. Modem filters for minimum intersymbol interference during back-to-back tests (*i.e.*, tests in which the transmitted signal, Z_T , is applied directly to the RECEIVER filter) can be selected by combining this transmitter filter with a receive modem filter (F_4) with $BT_s = 1.15$.

A back-to-back test in which the channel is eliminated and the simulation model contains only filters F_1 and F_4 is performed first. The simulation results for this case do not include the effect of synchronization errors.

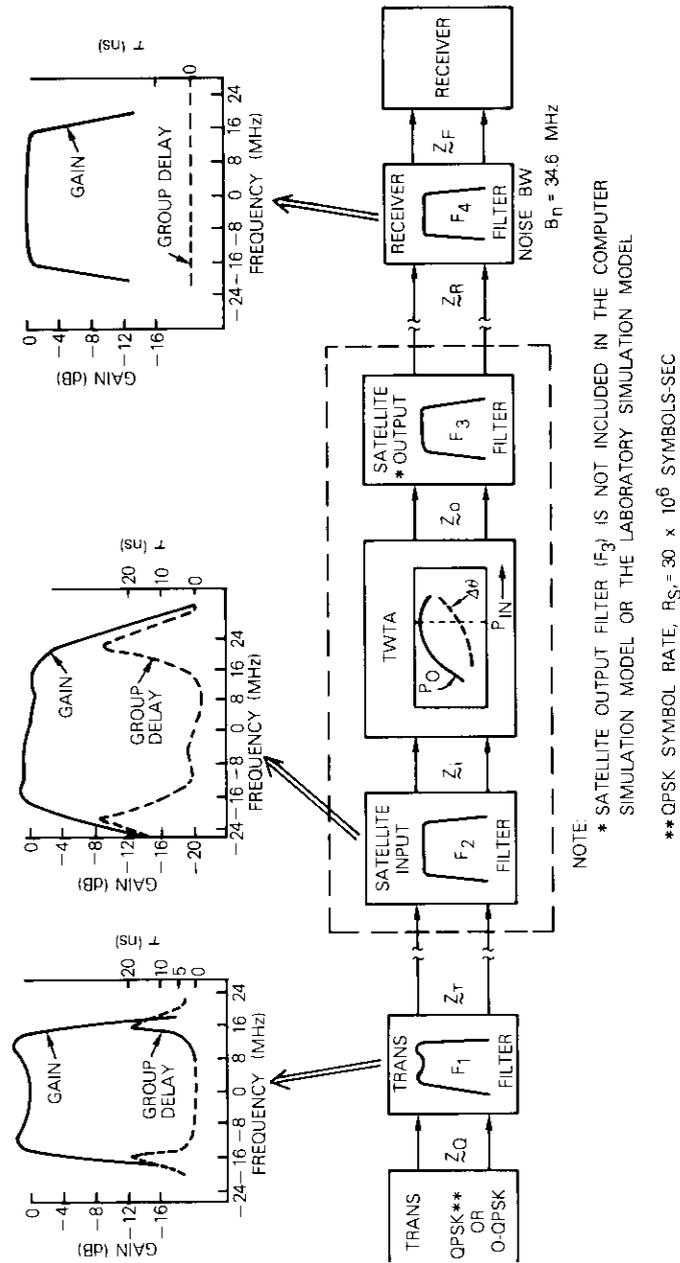


Figure 10. Channel Model

Likewise, the hardware simulation results apply to the case in which carrier and clock references are applied directly to the receive modem so that synchronization is not recovered from the noisy input signal. Thus, both of these cases apply to the situation in which the carrier and clock references are free of noise and jitter.

For the hardware simulation results, the hard-wired clock reference was adjusted for minimum error rate. In the computer simulations, static timing and phase estimates were obtained in CHAMP by using the technique described in the second section of this paper. As noted in that section, the procedure used in CHAMP does not necessarily minimize $P_b(\epsilon)$ although experience has shown that the synchronization is very close to optimum for simple channel configurations containing symmetrical filtering.

The calibration procedures used for the two measurements of this channel are compared in Figure 11. In the hardware calibration procedure shown in the lower part of this figure, the power out of the receive filter is measured in the presence of an unmodulated carrier (C_H) or noise (N_H). Because of the particular implementation of the modem used in the laboratory, the power reading of C_H does not change when modulation is applied to the transmitter. To convert the results to a common measure E_b/N_o (energy-per-bit-to-noise-density ratio), measurement C_H/N_H is modified to yield

$$(E_b/N_o)_H = \frac{C_H}{N_H} - 3 \text{ dB} + 10 \log_{10} (B_n T_s) \quad (18)$$

This conversion has been used consistently for the laboratory measurements.

The computer simulation model requires a different calibration procedure. First, the QPSK transmitter generates a unit power signal, Z_Q , which is then passed through the transmit filter F_1 . The filters in the model are scaled to have a voltage gain of one at their center frequency so that the modulated signal will experience a power loss in passing through this filter. This loss, measured in the program as P_R , in this case measures -0.52 dBW.

To assist in calibrating the simulation program, an unmodulated preamble of 32 symbols is generated at the start of each simulation run. This preamble yields the unmodulated waveforms $\{X_U, Y_U\}$ out of the RECEIVER filter; after an initial buildup interval, these waveforms become constant values. A final measurement in the simulation program is the mean and

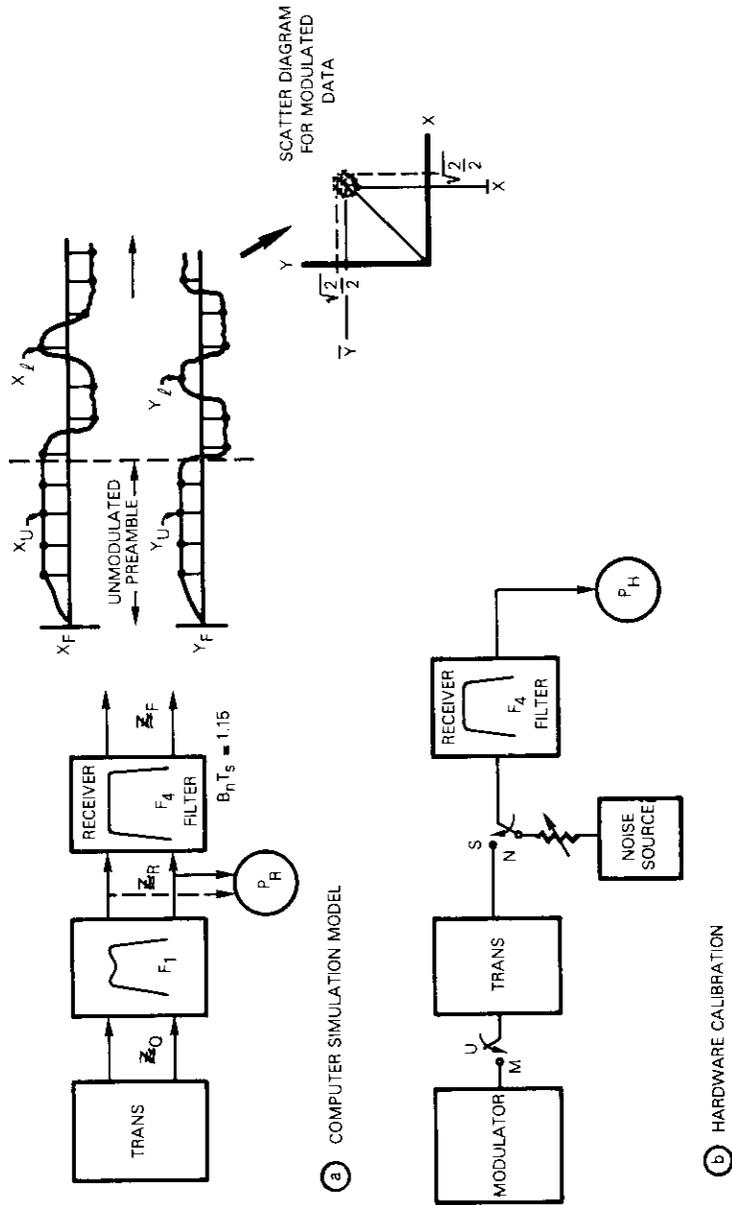


Figure 11. Calibration for Back-to-Back Test

variance of the in-phase and quadrature scatter diagrams out of the coherent demodulator, *i.e.*, \bar{X} , \bar{Y} , $\text{Var}(X)$, and $\text{Var}(Y)$.

These measurements in the simulation program allow the simulation results to be corrected to correspond to the laboratory measurements. In this case, results are expressed in terms of the carrier power that appears at the input to the receive modem filter so that the simulation results are normalized as follows:

$$(E_b/N_o)_s = \text{CNR} - 3 \text{ dB} - 10 \log_{10} (\bar{X}^2 + \bar{Y}^2) + 10 \log_{10} P_R + 10 \log_{10} (B_n T_s)$$

where CNR = variable used in the analytical calculation of error rate (corresponding to E_s/N_o for an ideal matched filter)

$\bar{X}^2 + \bar{Y}^2$ = average modulated energy per received symbol, which was used to normalize the error rate calculation*

P_R = modulated signal power at the input to F_4

$B_n T_s$ = correction factor for the analytical error rate calculation assumption of ideal matched filter detection in which the receiver noise bandwidth equals the symbol rate $1/T_s$.

With these corrections to the two sets of data, the results can be plotted as shown in Figure 12. It can be seen that the results of the hardware and computer simulations agree to within 0.2 dB.

With the back-to-back tests used to calibrate the simulation program and to provide a basis for comparing the computer simulation data with the laboratory measurements, a series of simulation runs were made for the complete channel shown in Figure 10. Although computer simulations were performed for both conventional and offset QPSK [8], laboratory measurements were available only for the conventional format. The simulation data were obtained from four runs using the CHAMP program, two each for conventional and offset QPSK at 0-dB and 14-dB input backoff to the TWTA. The data from these four runs were saved and each run was processed in the RECEIVER program for four values of E_s/N_o .

*This normalization is removed by equation (19) and a new correction is substituted.

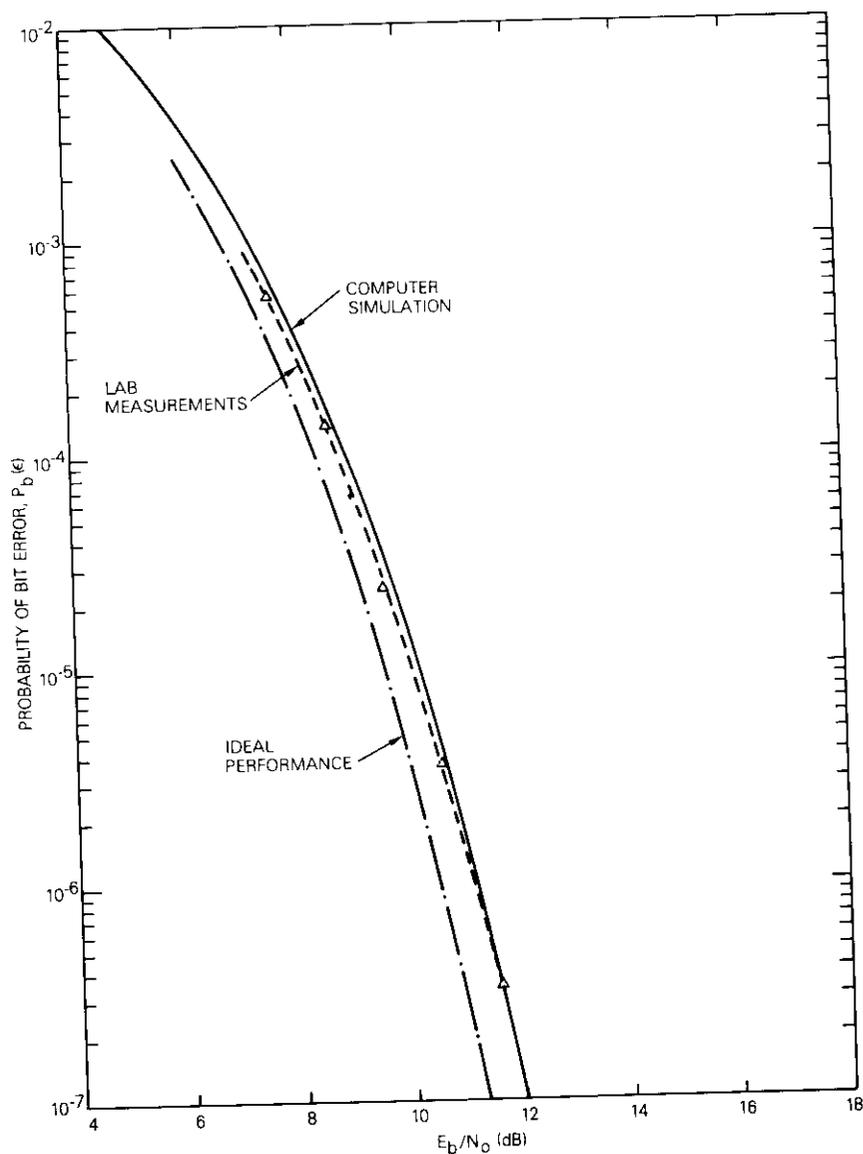


Figure 12. Comparison of Back-to-Back Test Results
(no carrier phase or timing jitter)

The data from the 16 simulation runs are shown in Table 1. The first three columns give the conditions for the particular run, and the next three columns give the relative timing points selected by the symbol synchronizer (τ_R), CHAMP (τ_c), and the timing that minimizes error probability (τ_o). The three phase values, θ_R , θ_c , and θ_o , give the same information for the phase estimators. The rms synchronization errors, σ_τ and σ_θ , give the total synchronization error experienced during the run. The last five columns in the table give the error probabilities including synchronization jitter for the optimum synchronization point τ_o , θ_o ; the synchronization point selected by the symbol and phase estimators τ_R , θ_R ; the optimum point with a 5° phase offset; the optimum point with a 5-percent (18°) time offset; and the optimum point with both a 5° phase and a 5-percent time offset.

In Table 1 it can be seen that the timing and phase recovery circuits select static or average operating points τ_R , θ_R biased from points τ_o , θ_o , respectively, that minimize error probability. The timing offsets ($\tau_R - \tau_o$) are relatively large (approximately 15°) for the linear channel and reduce to about 5° for the case in which the TWTA is operated at saturation (perhaps because the limiting action of the nonlinearity results in more symmetrically shaped pulses). It is also observed that rms timing errors for these runs are significantly larger than rms phase errors. For both synchronizers, the averaging time ($1/B_c$ and $1/B_p$) was approximately 20 symbols, which matches the values for the 60-Mbps modem that was used for the hardware simulations.

The QPSK data in Table 1 are compared to the hardware simulation results in Figure 13. The hardware simulation carrier-to-noise (C/N) ratios are converted to E_b/N_0 using equation (18). The CHAMP simulation results are converted to E_b/N_0 using

$$(E_b/N_0)_c = CNR - 3 \text{ dB} + 10 \log_{10} P_U - 10 \log_{10} (\bar{X}^2 + \bar{Y}^2) + 10 \log_{10} (B_n T_s) \quad (20)$$

where all terms have been defined previously except P_U , which is the unmodulated carrier power ($X_U^2 + Y_U^2$) that enters the RECEIVER during the preamble. This normalization is necessary because of the nonlinearity in the channel.

The procedure used to normalize the nonlinearity in the simulation program is described in Figure 14. Figure 14a shows the mapping between input and output envelope which takes place in the nonlinearity. For each sample of Z_{in} an envelope is calculated as $E_{in}(k\Delta) = [Z_{in}(k\Delta)$

TABLE 1. SUMMARY OF RECEIVER SIMULATION DATA FOR THE SIMPLE CHANNEL

Modulation	TWTA BO_{in} (dB)	E_b/N_o	Timing Point (deg)			Phase (deg)			rms Sync Error (deg)		$P_b(\epsilon)$ at τ_o, θ_o with Jitter	$P_b(\epsilon)$ at τ_R, θ_R with Jitter	$P_b(\epsilon)$ at $\tau_o, \theta_o + 5^\circ$ with Jitter	$P_b(\epsilon)$ at $\tau_o + 5\% T_s,$ θ_o	$P_b(\epsilon)$ at $\tau_o + 0.05 T_s,$ $\theta_o + 5^\circ$
			τ_R	τ_C	τ_o	θ_R	θ_C	θ_o	σ_r	σ_θ					
QPSK	14	6	18.2	8.9	4.7	0.8	0	-1.4	38.2	5.58	1.3×10^{-2}	1.4×10^{-2}	1.38×10^{-2}	1.7×10^{-2}	1.74×10^{-2}
		9	18.9	8.9	3.4	0.7	0	-1.9	23.9	3.75	8.7×10^{-4}	1.26×10^{-3}	1.12×10^{-3}	1.44×10^{-3}	1.9×10^{-3}
		12	19.3	8.9	2.6	0.7	0	-1.8	17.25	2.99	6.9×10^{-6}	5.0×10^{-5}	1.0×10^{-5}	2.3×10^{-5}	2.5×10^{-5}
	0	13.5	19.2	8.9	2.3	0.8	0	-1.8	15.2	2.75	2.5×10^{-7}	4.0×10^{-6}	5.0×10^{-7}	3.5×10^{-6}	6.3×10^{-6}
		6	11.5	9.8	6.4	1.3	0	-1.5	22.8	5.94	1.1×10^{-2}	1.1×10^{-2}	1.2×10^{-2}	1.4×10^{-2}	1.8×10^{-2}
		9	11.6	9.8	5.4	1.3	0	-2.1	15.2	3.62	9.6×10^{-4}	1.3×10^{-3}	1.3×10^{-3}	1.8×10^{-3}	2.4×10^{-3}
0-QPSK	14	12	11.5	9.8	4.9	1.3	0	-2.1	11.3	2.83	3.0×10^{-5}	4.5×10^{-5}	4.5×10^{-5}	1.0×10^{-4}	1.3×10^{-4}
		13.5	11.6	9.8	4.7	1.3	0	-2.56	10.0	2.56	3.3×10^{-6}	5.0×10^{-6}	6.3×10^{-6}	2.2×10^{-5}	4.0×10^{-5}
		6	19.6	8.6	3.9	1.6	0	-0.6	37.6	6.0	1.3×10^{-2}	1.8×10^{-2}	1.5×10^{-2}	1.8×10^{-2}	2.5×10^{-2}
	0	9	19.4	8.6	2.6	1.5	0	-0.9	24.7	4.1	9.8×10^{-4}	1.2×10^{-3}	1.2×10^{-3}	1.3×10^{-3}	1.7×10^{-3}
		12	20.4	8.6	1.8	1.3	0	-0.8	18.3	3.2	8.4×10^{-6}	4.9×10^{-5}	1.3×10^{-5}	4.5×10^{-5}	7.1×10^{-5}
		13.5	20.6	9.6	1.8	1.4	0	-0.7	15.9	2.9	5.2×10^{-7}	5.6×10^{-6}	1.3×10^{-6}	4.5×10^{-6}	1.3×10^{-5}
0	6	8.4	7.6	3.4	0.4	0	-0.7	19.1	8.8	1.3×10^{-2}	1.4×10^{-2}	1.5×10^{-2}	1.7×10^{-2}	1.8×10^{-2}	
	9	7.8	7.6	2.3	0.5	0	-1.0	14.1	5.2	1.4×10^{-3}	1.5×10^{-3}	1.7×10^{-3}	2.1×10^{-3}	2.5×10^{-3}	
	12	8.0	7.6	1.7	0.3	0	-0.9	11.7	3.7	4.8×10^{-5}	5.6×10^{-5}	7.1×10^{-5}	1.5×10^{-4}	2.0×10^{-4}	
	13.5	8.1	7.6	1.5	0.3	0	-0.7	11.0	3.3	5.3×10^{-6}	7.9×10^{-6}	1.1×10^{-5}	3.5×10^{-5}	7.9×10^{-5}	

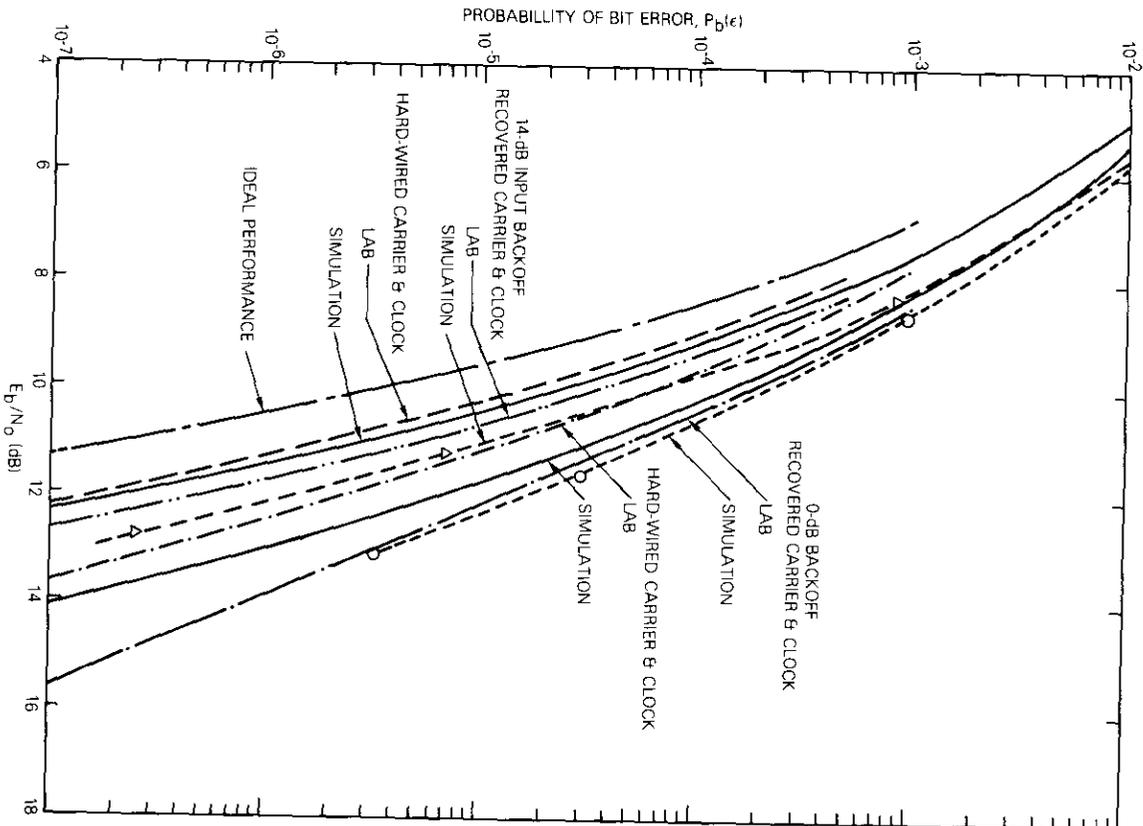


Figure 13. Comparison of Simulation and Laboratory Measurements for Conventional QPSK

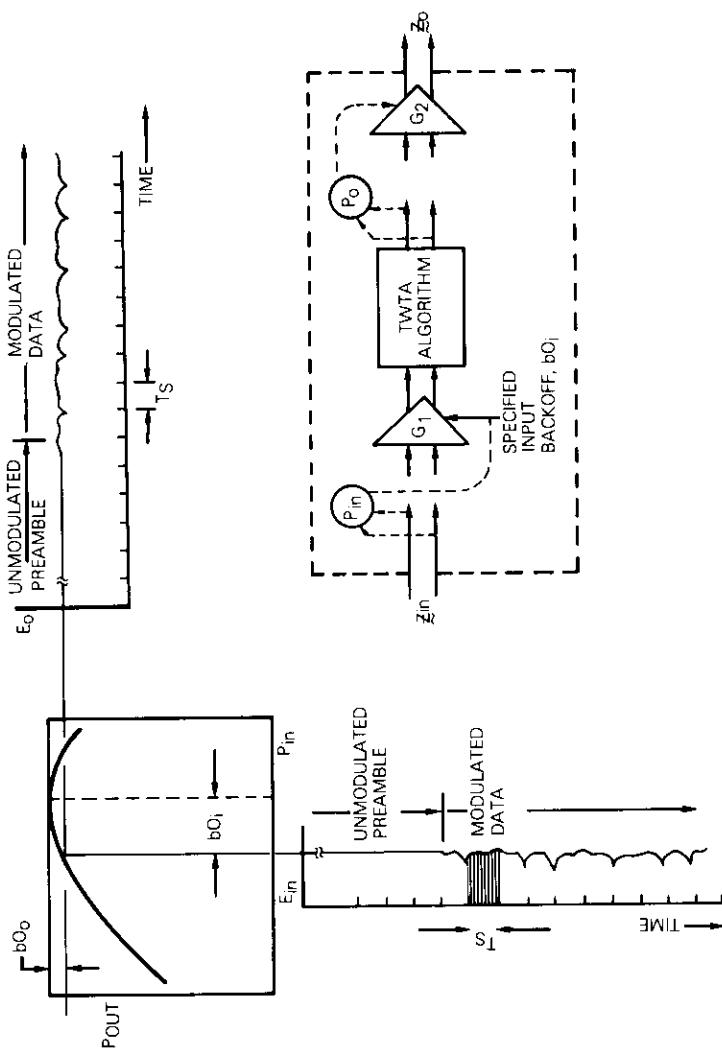


Figure 14. Calibration of TWTA Nonlinearity in the Simulation Program

$Z_{in}^*(k\Delta)]^\dagger$ and a look-up table is used to determine output envelope $E_o(k\Delta)$.[†] When the input signal has a constant envelope, for example, when the unmodulated preamble is sent through the traveling wave tube amplifier (TWTA), the output envelope is also constant. When the input envelope fluctuates as it would with a filtered QPSK signal, the output envelope would also fluctuate, although these fluctuations tend to be reduced or compressed for operation near TWTA saturation.

The input/output table is stored in normalized form so that saturation corresponds to input and output envelope values of unity. A specified input backoff is achieved by measuring the power in the input envelope ($P_{in} \doteq 1/N \sum Z_{in} Z_{in}^*$ as shown in Figure 14b) and then adjusting the input signal by gain G_1 to enter the look-up tables at the correct power level. Similarly, the output power level from the TWTA is measured ($P_o \doteq 1/N \sum E_o^2$) and this level is normalized so that the output power from the nonlinearity is 1 watt. Power measurements P_{in} and P_o utilize the total data records and therefore measure modulated signal power which is generally less than the unmodulated (preamble) signal power by approximately 0.5 dB.

The net effect of this method of calibrating the nonlinearity is that the unmodulated preamble is increased to a value that generally exceeds 1. This difference is an indirect measure of the power "lost" out of the TWTA due to envelope fluctuations on the output signal. The calibration in equation (20) accounts for this loss and relates the simulation results to an unmodulated C/N measured at the output of the receive modem filter. This same calibration point is used in the hardware simulations so that the two sets of measurements can be compared absolutely using the same definitions.

From the results in Figure 13, the losses compared to ideal performance can be summarized as shown in Table 2 at an error rate of 10^{-4} . The agreement between the computer and laboratory simulations is good; the largest difference between the two is 0.5 dB.

The performance of conventional and offset QPSK is compared in Figure 15. For these particular channel configurations, note that the offset QPSK format consistently gives slightly more loss than conventional QPSK. Additional comparisons of the two modulation formats are shown in Figure 16, which consists of eye and scatter diagrams for the noise-free channel.

[†]AM/PM conversion is applied using a second look-up table but this operation does not affect the power in the output signal.

TABLE 2. LOSS COMPARED TO IDEAL PERFORMANCE,
 $P_b(\epsilon) = 10^{-4}$

	14-dB Input Backoff		Saturated Transponder	
	Ideal Sync	Recovered Sync	Ideal Sync	Recovered Sync
Lab Measurements	0.5 dB	0.9 dB	1.3 dB	2.0 dB
Computer Simulation	0.7 dB	1.3 dB	1.8 dB	2.2 dB

Note that the eye and scatter diagrams give some qualitative justification for the slightly poorer performance of offset QPSK for this particular channel configuration. As shown in the next section, this conclusion concerning the relative performance of the two modulation techniques is not true in general.

Finally, the results for these cases are compared in terms of the variation in mean square synchronization error with E_b/N_o , as shown in Figure 17. The ordinate of this curve is the synchronizer signal-to-noise (S/N) ratio, which is defined as $10 \log_{10} (1/\sigma^2)$, where σ^2 is the mean square synchronization error measured in radians. The timing errors for these cases are an order of magnitude larger than the phase errors, which is directly attributable to the large zero-crossing jitter in the eye diagram of Figure 16. With the relatively short averaging times (approximately 20 symbols) used in the timing synchronizer, this jitter is not averaged sufficiently to yield a small net error in the timing point.

Simulation of a more complex channel configuration

A second series of computer simulation experiments has been performed for the more complex channel configuration shown in Figure 18. This configuration includes a desired signal channel, which is represented by the elements in the center of the diagram, and two adjacent transponder channels. Three conventional or offset QPSK transmitters are simulated, each modulated by independent random data sources. The symbol rate, R_s , of each transmitter is 30×10^6 symbols/second and the carriers are spaced 62.5 MHz apart. Each signal is filtered in a 4-pole Butterworth filter with a 3-dB bandwidth of 45 MHz ($BT_s = 1.5$). These filters are perfectly equalized. The filtered outputs of each transmitter are applied to the simulated high-power amplifier (HPA) whose AM/AM and AM/PM characteristics are shown in the insert in Figure 18. For the case considered

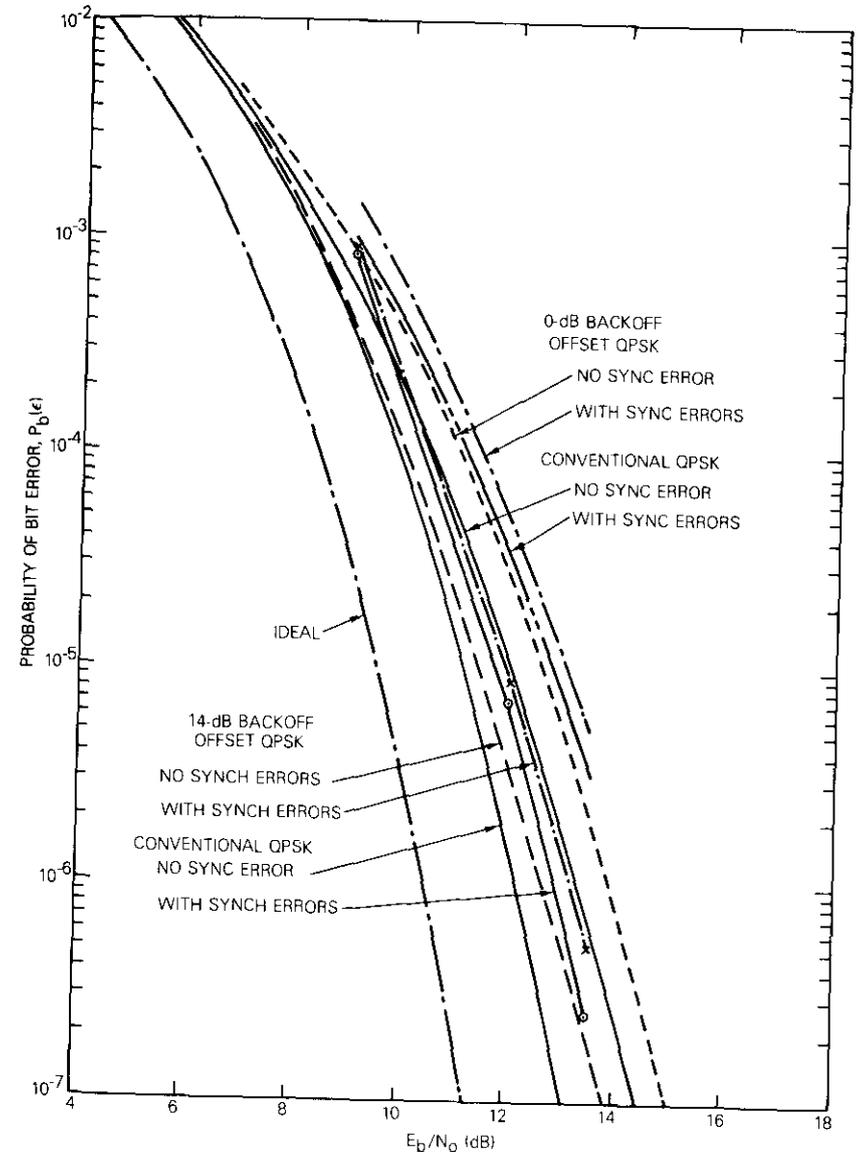


Figure 15. Comparison of Conventional and Offset QPSK

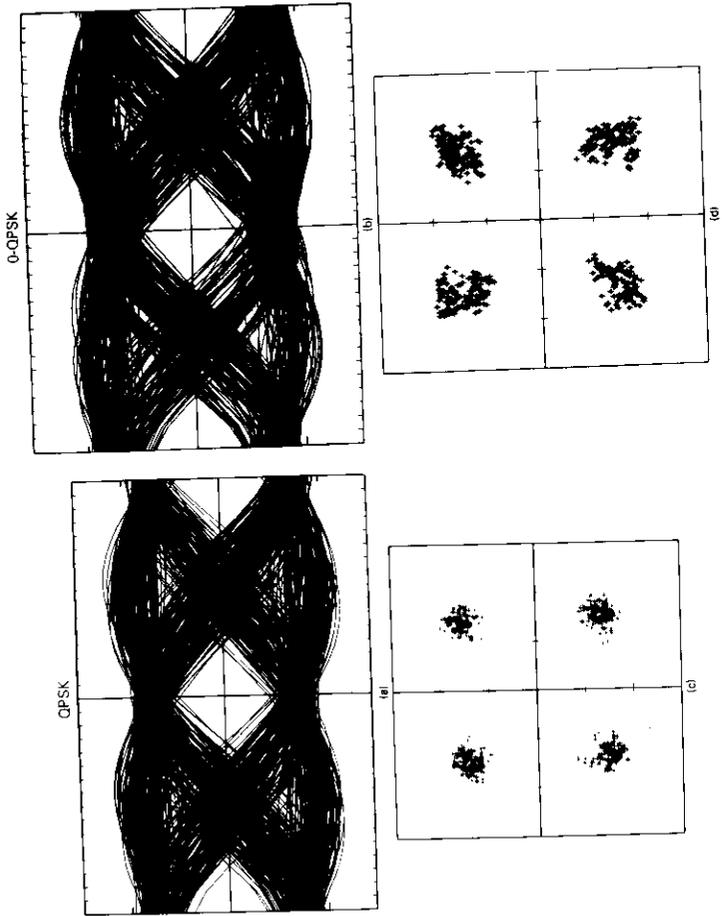


Figure 16. Eye and Scatter Diagrams for Conventional and Offset QPSK

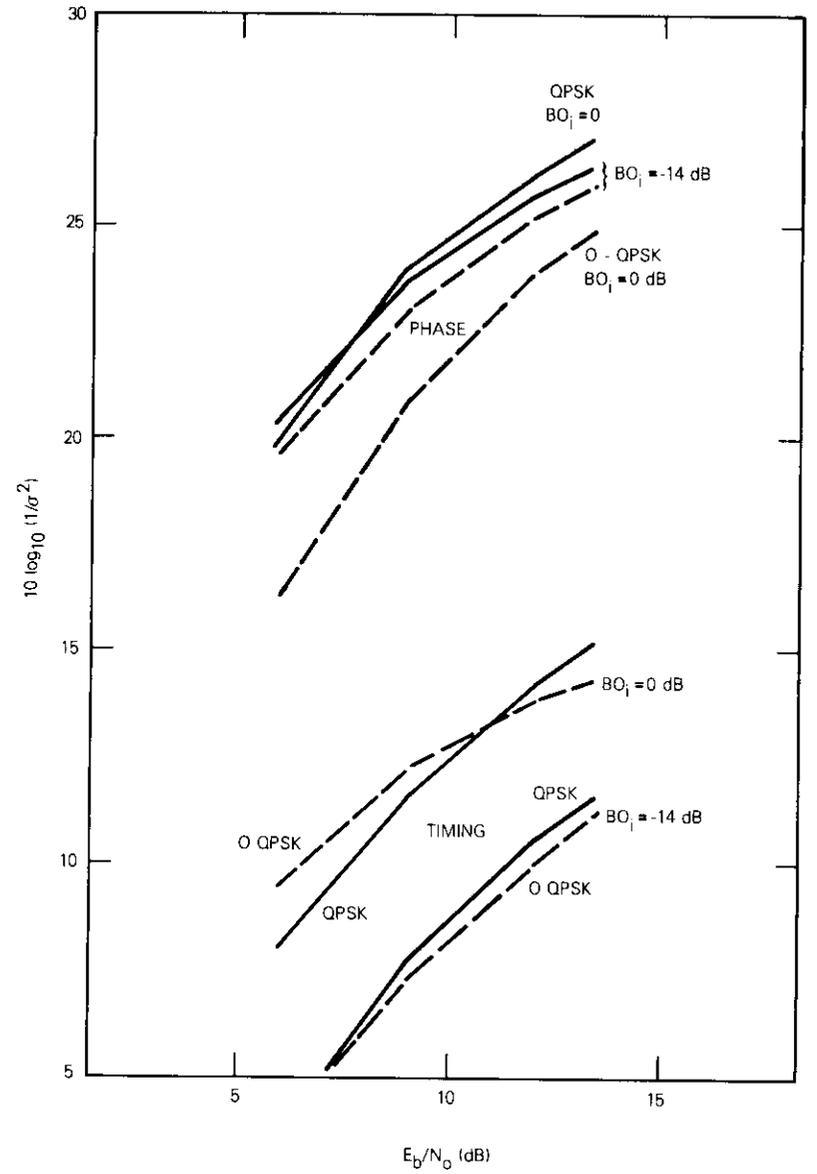


Figure 17. Phase and Timing "Signal-to-Noise Ratio"

here, all three HPAs are operated at saturation. The outputs of the HPAs are summed to produce the composite signal at the input to the satellite.

The satellite transponder model includes three identical channels; the desired channel, one channel 62.5 MHz above it, and one 62.5 MHz below it. Each channel contains a satellite input filter, a simulated TWTA nonlinearity, and a satellite output filter. The characteristics of these three elements are also shown (for one channel centered at frequency f_A) in Figure 18. The outputs of the three transponder channels are finally summed to produce the signal that enters the desired demodulator.

The model in Figure 18 is a complete model of the major sources of impairment encountered with single-carrier digital transmission (*i.e.*, TDMA transmission) in a channelized satellite transponder. It differs from the simple model used in the preceding subsection in two important respects. First, the HPA nonlinearity is included in the more complex model and, when combined with the TWTA nonlinearity and the intervening satellite input filter, the cascaded nonlinearity can lead to additional signal distortion. Secondly, in the more complex model the inclusion of the adjacent channel interfering signals, which can enter the desired signal path on either the up- or the down-link, adds additional interference to the desired signal.

The eye and scatter diagrams for the CHAMP runs for this case are shown in Figure 19. The data from the RECEIVER runs are given in Table 3. For all of these runs, the time constants used in both synchronizers were selected as 50 symbols. This increase in averaging time, combined with the decrease in zero-crossing jitter in the eye diagrams, results in a significant reduction in the root-mean-square timing error (σ_t) relative to the case considered in the preceding subsection.

Error rate for these cases is plotted in Figure 20. In contrast to the previous case, offset QPSK gives a small but consistent improvement in performance. Histograms of phase and timing errors for the six simulation runs are given in Figure 21.

Summary and conclusions

The overall objective of the work reported in this paper has been to develop a technique for including the effect of synchronization errors in time-domain simulations of coherent digital transmission systems that operate over the satellite channel. The inclusion of these effects, which account for one of the significant sources of impairment encountered in

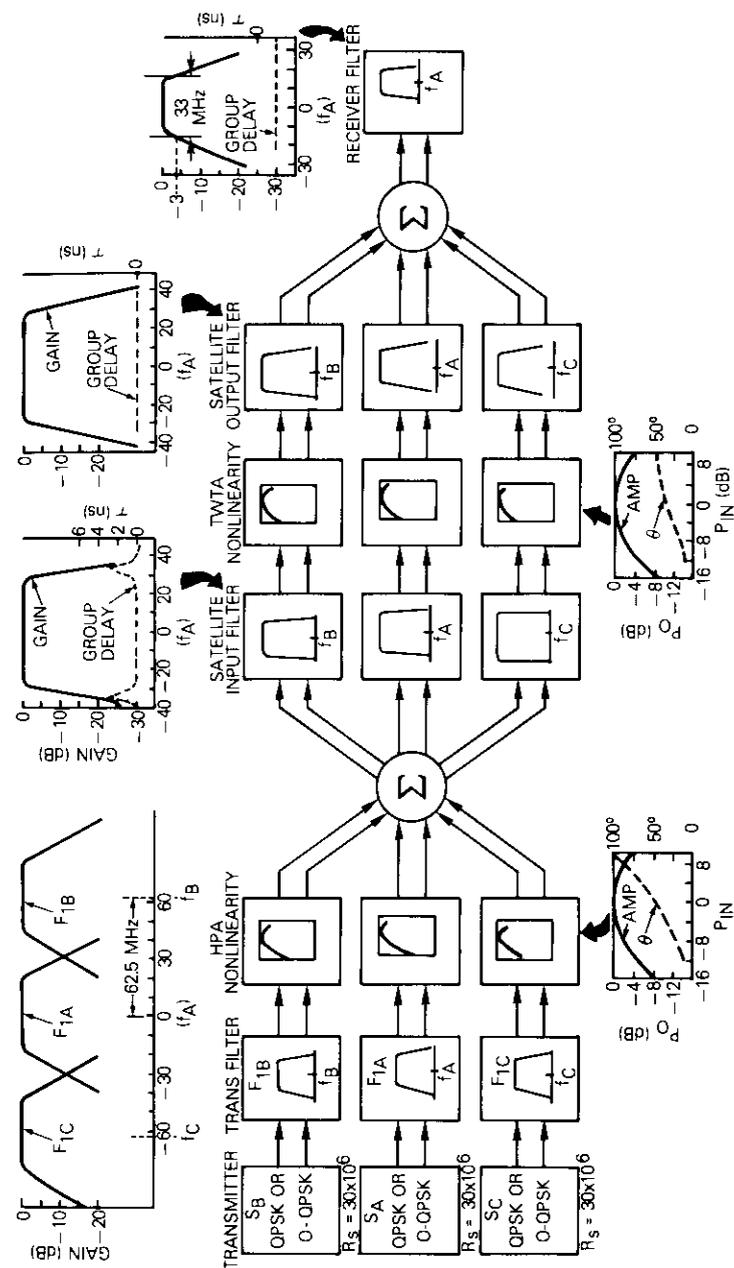


Figure 18. More Complex Channel Configuration

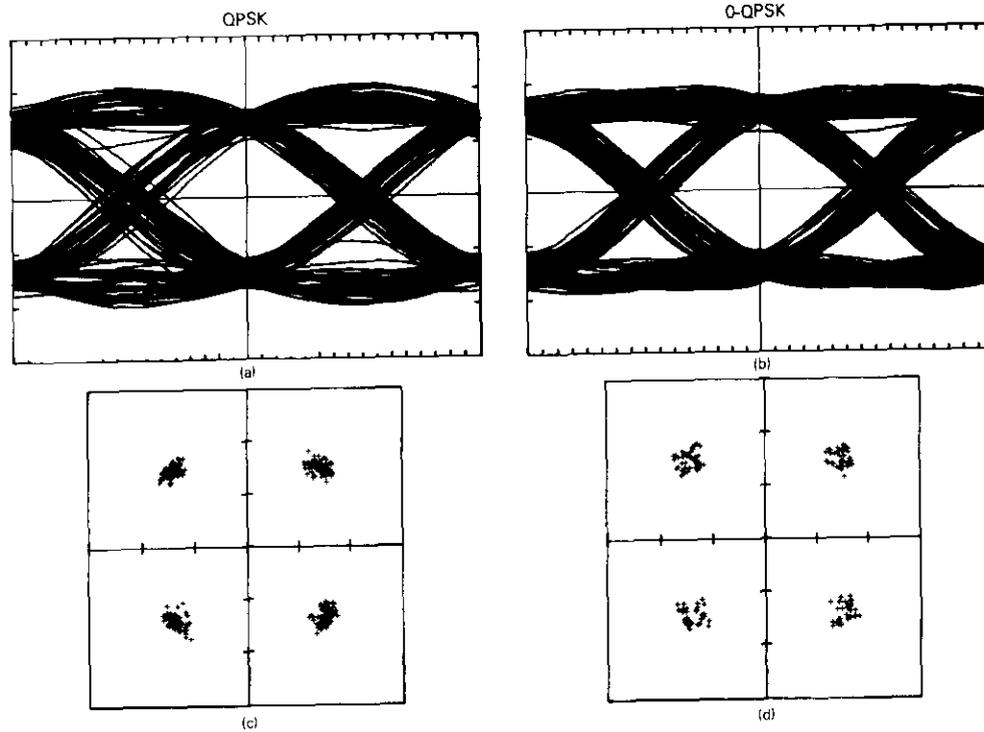


Figure 19. Eye and Scatter Diagrams for Complex Channel Model

TABLE 3. DATA FROM RECEIVER RUNS

Modulation	E_b/N_0	Timing Point (deg)			Phase (deg)			rms Sync Error (deg)		$P_b(\epsilon)$ at τ_o, θ_o with Jitter	$P_b(\epsilon)$ at $\tau_o + 5\% T_s, \theta_o$ with Jitter	$P_b(\epsilon)$ at $\tau_o + 5\% T_s, \theta_o + 5^\circ$ with Jitter
		τ_r	τ_c	τ_o	θ_R	θ_C	θ_o	σ_r	σ_θ			
QPSK	6	9.08	16.03	15.49	-1.6	-0.02	0.03	7.40	3.34	6.5×10^{-3}	7.2×10^{-3}	7.6×10^{-3}
	9	8.84	16.03	14.17	-1.58	-0.02	-0.05	5.83	2.17	3.4×10^{-4}	4.3×10^{-4}	5.3×10^{-4}
	12	9.34	16.03	11.59	-1.30	-0.02	-0.09	4.11	1.35	3.0×10^{-6}	5.9×10^{-6}	1.0×10^{-5}
O-QPSK	6	8.88	16.33	18.07	-2.11	-0.01	0.09	5.68	3.60	4.4×10^{-3}	4.6×10^{-3}	5.2×10^{-3}
	9	9.72	16.33	18.21	-2.22	-0.01	0.03	4.03	2.57	1.6×10^{-4}	1.9×10^{-4}	2.5×10^{-4}
	12	9.84	16.33	18.83	-1.65	-0.01	-0.06	3.17	1.80	6.0×10^{-7}	1.0×10^{-6}	1.6×10^{-6}

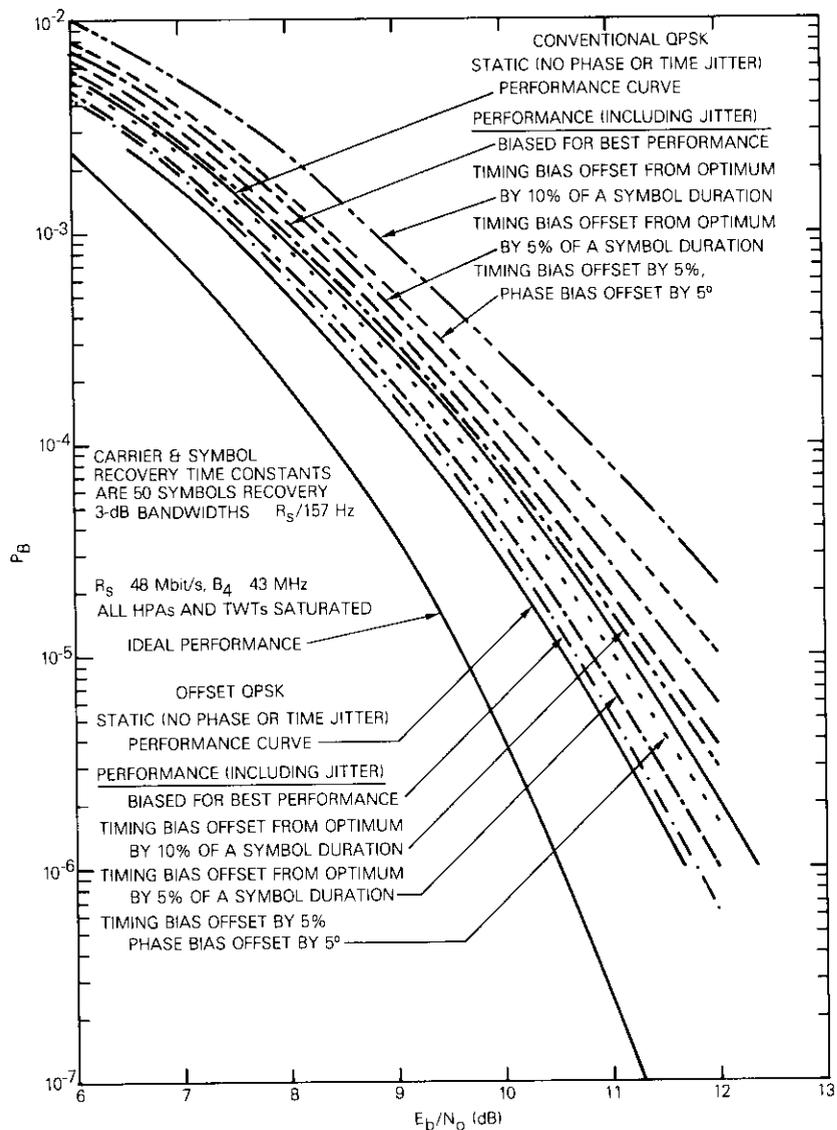


Figure 20. Error Rate vs E_b/N_0 .

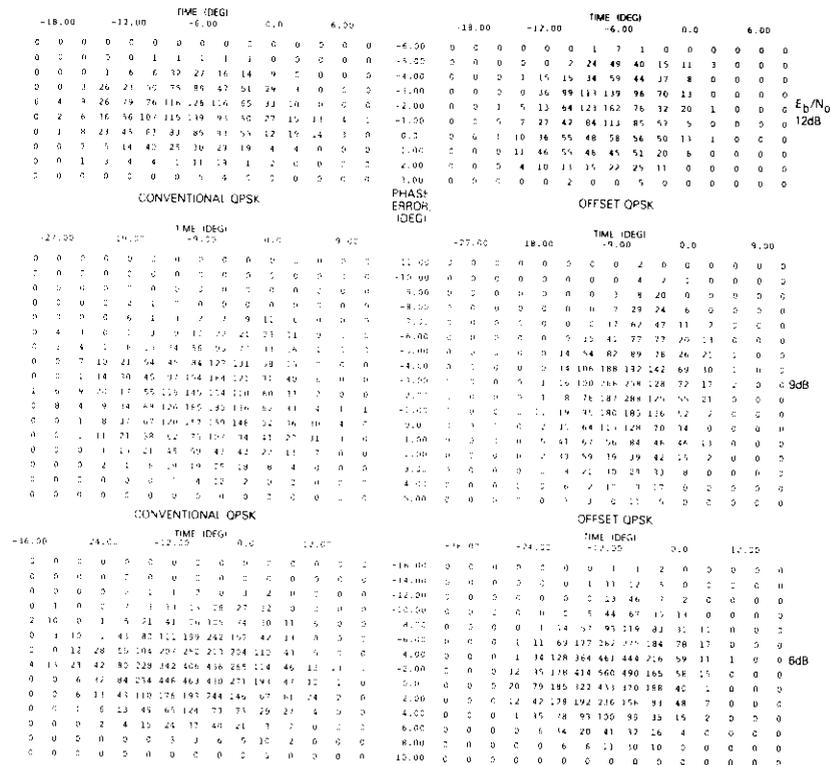


Figure 21. Histograms of Phase and Timing Errors

the communications satellite channel, increases the accuracy of the performance predictions provided by CHAMP simulation.

To accomplish this objective, a general simulation procedure has been developed to augment COMSAT's existing CHAMP. This new simulation model, RECEIVER, has a flexible structure which allows expansion to include detailed models of different carrier phase and symbol timing recovery implementations. It operates on data generated by CHAMP and processes these data further to determine the sensitivity of error probability to errors in phase and symbol timing at different values of E_b/N_0 , the joint statistics of phase and timing errors due to both thermal noise and symbol distortion caused by the channel, and the degradation in error rate due to these synchronization errors at different values of E_b/N_0 .

Since carrier phase and symbol timing recovery implementations generally involve nonlinear operations, the RECEIVER program must utilize a "pure simulation" to determine the joint statistics of phase and timing errors. During this phase of the processing, thermal noise is applied at the input to the simulated receiver in the form of Gaussian random numbers which are added to the simulated data record obtained from the CHAMP simulation run. Although this mode of operation is minimized by the approach used in RECEIVER, inclusion of the effects of synchronization errors significantly increases the overall simulation time. This increase in overall simulation (*i.e.*, CPU) time was expected. It is envisioned that the RECEIVER program would be used selectively to add the effect of synchronization errors to certain CHAMP results that are of particular interest, or for which absolute performance predictions are needed.

The program has been implemented to model one particular implementation of the carrier phase and symbol timing recovery operations, namely, a quadrupling circuit and tuned filter for carrier recovery and a half-symbol delay, multiply (for QPSK, or a squaring circuit for offset QPSK), and filter combination for clock recovery. The model was verified by a series of simulation experiments for QPSK operation over a simulated communications satellite channel for which extensive laboratory measurements were available. The goal of these particular simulation experiments was to model the hardware simulation channel as accurately as possible in the computer model, and to compare results on an absolute basis. Difficulties were encountered with definitions (C/N , E_b/N_o , and whether the reference signals were modulated or unmodulated) and with calibration, and these problems were compounded by the presence of the channel nonlinearities. However, once these inconsistencies were understood and resolved, agreement between the hardware and the computer simulation results was excellent. For the cases for which comparisons could be made, the results from the two sources agreed to within ± 0.5 dB.

Additional simulation results in the preceding section exercise the model for a more complex, and realistic, model of the satellite channel. Both conventional and offset QPSK were used for these experiments. Although these simulations certainly do not provide an exhaustive comparison of these two modulation techniques, it can be concluded from the preliminary results that the relative performance of conventional and offset QPSK may be highly dependent on the channel configuration, and in particular on the choice of transmit and receive modem filters.

Acknowledgments

The authors wish to acknowledge helpful discussions with Mr. Smith Rhodes concerning methods of modeling the symbol synchronization functions. The hardware simulation results were obtained by M. E. Jones and D. Weinreich of the Transmission Systems Laboratory, COMSAT Laboratories.

References

- [1] "Final Report on DSCS Channel Model," prepared for the Defense Communications Agency by Computer Sciences Corporation, Contract DCA-100-72-C-0002, August 31, 1972.
- [2] D. L. Hedderly and L. Lundquist, "Computer Simulation of a Digital Satellite Communications Link," International Conference on Communications, Philadelphia, Pennsylvania, 1972.
- [3] W. L. Cook, "Interactive Computer Simulation of Satellite Transmission Systems," *Proceedings of the Fifth Annual Pittsburgh Conference on Modeling and Simulation*, April 1974, pp. 867-872.
- [4] L. C. Palmer, "Characterization of the Effects of Phase and Symbol Timing Errors for QPSK Transmission," *Proc. ICC'77*, p. 340.
- [5] F. M. Gardner, *Phaselock Techniques*, New York: John Wiley & Sons, Inc., 1966.
- [6] A. J. Viterbi, *Principles of Coherent Communications*, New York: McGraw-Hill, 1966.
- [7] M. E. Jones and M. R. Wachs, "Measured 60-Mbps 4-Phase PSK Signal Impairments Through Earth Station HPA and Satellite Transponder," National Telecommunications Conference, November 29-December 1, 1976, *Proc.*, p. 21.2-1.
- [8] M. E. Jones, Private Communication.
- [9] S. A. Rhodes, "Effect of Noisy Phase Reference On Coherent Detection of Offset-QPSK Signals," *IEEE Transactions on Communications*, August 1974, p. 1046.



Larry C. Palmer received a B.S. degree from Washington and Lee University, and a B.E.E. from Rensselaer Polytechnic Institute, both in 1955; and an M.S. and Ph.D. from the University of Maryland in 1963 and 1970. In 1963 he joined IIT INTELCOM, in Falls Church, Va., where he worked on the military communications satellite system. After acquisition of INTELCOM by Computer Sciences Corporation in 1965, he was involved in several projects involving systems analysis, modeling, and simulation related to com-

munications and navigation satellite systems, and digital signal processing systems. In 1974 he became senior staff scientist in the Transmission Systems Laboratory of COMSAT Laboratories. He is responsible for systems analysis related to future communication satellite systems. Specific areas include modulation and multiple-access techniques, and channel modeling to predict transmission impairments. He is a member of Tau Beta Phi, Eta Kappa Nu, and the IEEE.

Sheldon H. Lebowitz received a B.E.E. degree from the City College of New York in 1964. He has worked at the U.S. Navy Maritime Engineering Laboratory, the Electromagnetic Compatibility Analysis Center, Norden Division of United Aircraft, and Computer Sciences Corporation. In 1974 he joined COMSAT Laboratories, where he has been involved with digital computer simulations of communication systems.



Index: multiplexer, function filter, experimental design

Singly terminated pseudo-elliptic function filter

M. H. CHEN

(Manuscript received May 16, 1977)

Abstract

The synthesis, realization, and tuning methods for a singly terminated pseudo-elliptic function filter have been developed. An experimental model of a 6-pole dual mode filter has been constructed in a cascading structure of three square waveguide cavities and tuned to realize the expected performance. This development allows a contiguous band multiplexer to use advanced pseudo-elliptic function filters with considerable space and weight savings.

Introduction

Singly terminated filters may be used advantageously as channel filters in a contiguous band multiplexer. A recently developed multiplexer [1] that uses singly terminated Chebychev filters as channel filters has demonstrated such excellent performance that the concept was selected for the INTELSAT V system. The use of contiguous band multiplexers instead of conventional even- and odd-channel multiplexers results in considerable space and weight savings in the satellite. Further weight savings

This paper is based upon work performed at COMSAT Laboratories under the sponsorship of the International Telecommunications Satellite Organization (INTELSAT). Views expressed in this paper are not necessarily those of INTELSAT.

can be achieved by replacing Chebychev channel filters with pseudo-elliptic function filters. However, the techniques for synthesis, design, and tuning of a singly terminated pseudo-elliptic function filter were not readily available. This paper describes these techniques.

Synthesis and realization for a singly terminated filter

The transfer function, $t(p)$, for a pseudo-elliptic filter of the given selectivity and in-band ripple is of the form [2], [3]

$$t(p) = \frac{P(p)}{A(p) + pB(p)} \tag{1}$$

where $P(p)$, $A(p)$, and $B(p)$ are all even polynomials of $p = j\omega$. If a singly terminated filter connects a unit voltage source and a load resistance R_2 , as shown in Figure 1, the transfer admittance Y'_{21} can generally be represented as

$$Y'_{21} = \frac{P(p)}{A(p) + pB(p)} \tag{2}$$

Since the input voltage and output current in a $L - C$ circuit must be 90° out of phase [4], the factor j is introduced in equation (2)

$$Y'_{21} = \frac{jP(p)}{A(p) + pB(p)} \tag{3}$$

Since the transfer admittance Y'_{21} is also related to the short-circuit parameters by

$$Y'_{21} = \frac{y'_{21}}{1 + y'_{22}} \tag{4}$$

y'_{21} and y'_{22} may be derived by comparing equations (3) and (4) so that

$$y'_{21} = \frac{jP(p)}{A(p)} \tag{5a}$$

$$y'_{22} = \frac{pB(p)}{A(p)} \tag{5b}$$

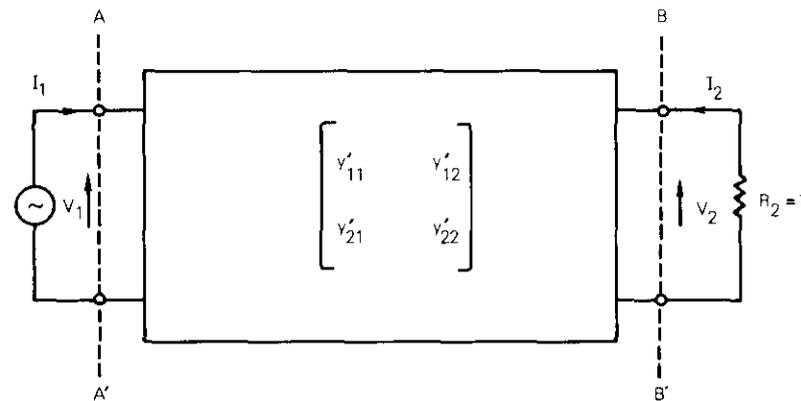


Figure 1. Network Configuration for a Singly Terminated Filter

An N -pole synchronously tuned multiple-cavity filter may be represented [5] by a lumped circuit, as shown in Figure 2, with the voltage and current related by the equation

$$\begin{bmatrix} V_1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ V_N \end{bmatrix} = Z \begin{bmatrix} I_1 \\ I_2 \\ \cdot \\ \cdot \\ \cdot \\ I_{n-1} \\ I_N \end{bmatrix} \tag{6}$$

where $Z = j(\omega 1_N + M)$. (7)

The coupling matrix M is an $N \times N$ real symmetrical square matrix

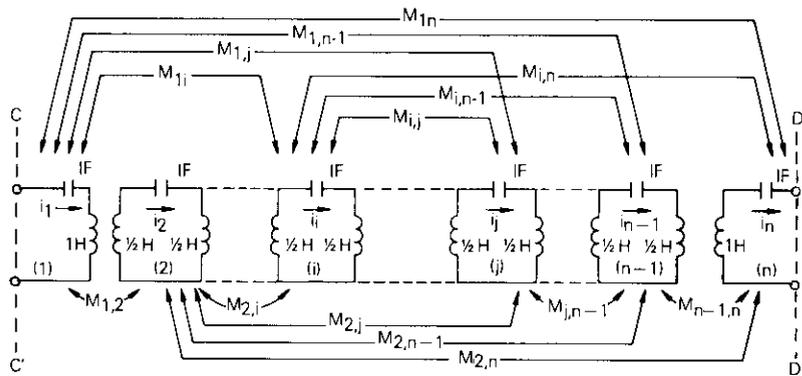


Figure 2. Lumped Circuit Representation for an *N*-Pole Synchronously Tuned Cavity Filter

which can be transformed to the Jordan normal form [6] by a non-singular matrix *T* as follows:

$$M = TJT^{-1} \tag{8}$$

where *J* is a diagonal matrix with eigenvalues, $\lambda_1, \lambda_2, \dots, \lambda_N$, and *T* is constructed by *N* orthogonal and normalized eigenvectors, t_1, t_2, \dots, t_N .

The voltage and current relationship for the first and last cavities should be related to the synthesized short-circuit admittance parameters, y'_{21} and y'_{22} as given in equation (5) to obtain the specified transfer characteristics for a filter. The networks in Figures 1 and 2 may be connected by a normalization scheme [7] as shown in Figure 3. The network parameters defined in Figure 3 are related to the impedance matrix in equation (6) as follows*:

$$y_{11} = (Z^{-1})_{11} \tag{9a}$$

$$y_{21} = (Z^{-1})_{N1} \tag{9b}$$

$$y_{22} = (Z^{-1})_{NN} \tag{9c}$$

*If a filter with an odd number of poles is considered, the synthesis method described by Saito [8] should be followed.

As discussed in Reference 7, the short-circuit admittance parameters have the partial fraction expansion

$$y_{11} = \sum_{i=1}^N \frac{t_{1i}^2}{j\omega - j\lambda_i} \tag{10a}$$

$$y_{21} = \sum_{i=1}^N \frac{t_{1i}t_{Ni}}{j\omega - j\lambda_i} \tag{10b}$$

$$y_{22} = \sum_{i=1}^N \frac{t_{Ni}^2}{j\omega - j\lambda_i} \tag{10c}$$

where $\lambda_1, \lambda_2, \dots, \lambda_N$ are the eigenvalues of *M* and

$$t_1 = \begin{bmatrix} t_{11} \\ t_{12} \\ \cdot \\ \cdot \\ t_{1N} \end{bmatrix}; \quad t_N = \begin{bmatrix} t_{N1} \\ t_{N2} \\ \cdot \\ \cdot \\ t_{NN} \end{bmatrix} \tag{11}$$

The normalization scheme represented in Figure 3 with input and output transformers would relate normalized and unnormalized short-circuit admittances as follows:

$$y'_{11} = n_1^2 y_{11} \tag{12a}$$

$$y'_{22} = n_1 n_2 y_{21} \tag{12b}$$

$$y'_{22} = n_2^2 y_{22} \tag{12c}$$

Consequently, the eigenvalues and eigenvectors of coupling matrix *M* may be solved based on the polynomials given by equation (5).

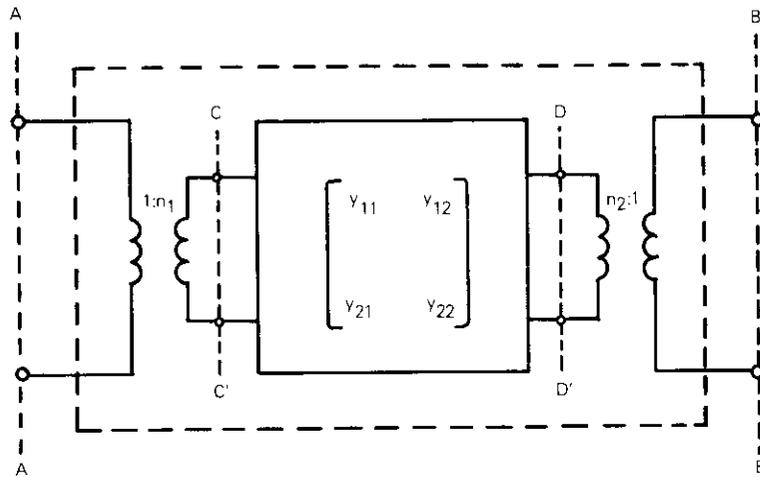


Figure 3. Normalization Scheme for Network Representations

Experimental model of a 6-pole dual mode filter

A 6-pole singly terminated pseudo-elliptic function filter was designed, fabricated, and tuned to verify the synthesis and realization methods outlined in the previous section. The step-by-step design procedures are described in this section.

First, a selectivity, k , of 0.8 is chosen, and two poles and two zeros for a fully elliptic function filter are derived [2]. For a pseudo-elliptic function filter, one pole is dropped. With an in-band ripple of 0.1 dB, the transfer function for a pseudo-elliptic function filter is solved with the polynomials

$$P(p) = 0.5615 p^2 + 1 \tag{13a}$$

$$A(p) = 2.1583 p^6 + 7.7871 p^4 + 6.6631 p^2 + 1 \tag{13b}$$

$$B(p) = 4.4120 p^4 + 8.3448 p^2 + 3.4933 \tag{13c}$$

Consequently, the short-circuit admittances are derived:

$$y'_{21} = j \frac{0.5615 p^2 + 1}{2.1583 p^6 + 7.7871 p^4 + 6.6631 p^2 + 1} \tag{14a}$$

$$y'_{22} = \frac{4.4120 p^5 + 8.3448 p^3 + 3.4933 p}{2.1583 p^6 + 7.7871 p^4 + 6.6631 p^2 + 1} \tag{14b}$$

A partial fraction expansion of equation (14) expressed in the desired form is

$$y'_{21} = n_1 n_2 \sum_{i=1}^6 \frac{t_{1i} t_{6i}}{j(\omega - \lambda_i)} \tag{15a}$$

$$y'_{22} = n_2^2 \sum_{i=1}^6 \frac{t_{6i}^2}{j(\omega - \lambda_i)} \tag{15b}$$

where λ_i , t_{1i} , and t_{6i} , $i = 1, 2, \dots, 6$, are solved and tabulated in Table 1, and n_1 and n_2 are found to be

$$n_1^2 = 0.68820 \tag{16a}$$

$$n_2^2 = 2.04417 \tag{16b}$$

The coupling matrix M may be computed according to equation (8) if both the matrix J and matrix T are found. Matrix J is a diagonal matrix whose diagonal elements are equal to the eigenvalues listed in Table 1.

TABLE 1. EIGENVALUES AND EIGENVECTORS FOR COUPLING MATRIX M

i	T_{1i}	T_{6i}	λ_i
1	-0.61228	0.35833	-0.43598
2	0.35283	0.20673	-1.00699
3	0.02497	0.57347	-1.55041
4	0.61228	0.35833	0.43598
5	-0.35283	0.20673	1.00699
6	-0.02497	0.57347	1.55041

The two eigenvectors given in Table 1 may serve as base vectors 1 and 6 in constructing matrix T . The remaining four base vectors, 2-5, may be obtained by the method of orthogonal projection [6] upon vectors 1 and 6. Consequently, with these six eigenvectors, the matrix T is constructed so that

$$T = \begin{bmatrix} -0.61228 & 0.35283 & 0.02497 & 0.61228 & -0.35283 & -0.02497 \\ 0.00000 & 0.00000 & 0.00000 & 0.40889 & 0.74669 & -0.52466 \\ 0.70478 & 0.20142 & -0.26987 & 0.34974 & -0.41163 & -0.31326 \\ 0.00000 & 0.89006 & -0.08202 & -0.40509 & 0.18500 & -0.05241 \\ 0.00000 & 0.00000 & 0.76874 & -0.20764 & -0.26752 & -0.54255 \\ 0.35833 & 0.20673 & 0.57347 & 0.35833 & 0.20673 & 0.57342 \end{bmatrix} \quad (17)$$

From equation (8), the following coupling matrix is obtained:

$$M = \begin{bmatrix} 0.00000 & -0.13585 & 0.37876 & -4.48490 & 0.03087 & 0.00000 \\ -0.13583 & 1.06111 & 0.00766 & 0.10952 & 0.20317 & -0.24716 \\ 0.37876 & 0.00766 & 0.00577 & -0.32784 & 0.66438 & -0.22166 \\ -0.48490 & 0.10952 & -0.32784 & -0.69791 & 0.12868 & -0.18373 \\ 0.03087 & 0.20317 & 0.66438 & 0.12868 & -0.36897 & -1.25401 \\ 0.00000 & -0.24716 & -0.22166 & -0.18373 & -1.25401 & 0.00000 \end{bmatrix} \quad (18)$$

This coupling matrix may be reduced to a realizable form via similarity transformations [5]. This results in the reduced coupling matrix

$$M_{\text{reduced}} = \begin{bmatrix} 0.00000 & -0.63086 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ -0.63086 & 0.00000 & -0.57573 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & -0.57573 & 0.00000 & 0.41025 & 0.00000 & -0.60390 \\ 0.00000 & 0.00000 & 0.41025 & 0.00000 & -0.99680 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & -0.99680 & 0.00000 & -1.16268 \\ 0.00000 & 0.00000 & -0.60390 & 0.00000 & -1.16268 & 0.00000 \end{bmatrix} \quad (19)$$

The coupling matrix given by equation (19) and the input and output transformer ratios given by equation (16) provide all the normalizing coupling coefficients and the input and output resistances for a pseudo-elliptic function filter with a selectivity of 0.8 and an in-band ripple of 0.1 dB. In the design of a filter with a specified bandwidth, these normalized values must be multiplied by the fraction bandwidth. For example, the coupling coefficient and the loaded Q 's at the input and output ports for a filter with a 74-MHz bandwidth and a center frequency of 3.905 GHz are obtained as follows:

$$\left. \begin{aligned} k_{12} &= -0.01195 \\ k_{23} &= -0.01091 \\ k_{34} &= 0.00777 \\ k_{45} &= -0.01889 \\ k_{56} &= -0.02203 \\ k_{36} &= -0.01144 \end{aligned} \right\} \quad (20)$$

loaded Q at input = 76.678
loaded Q at output = 25.815

With the parameters given by equation (20), a dual mode 6-pole filter has been designed with three square waveguide cavities in cascade as shown in Figure 4. The mode designation and the modal field orientation

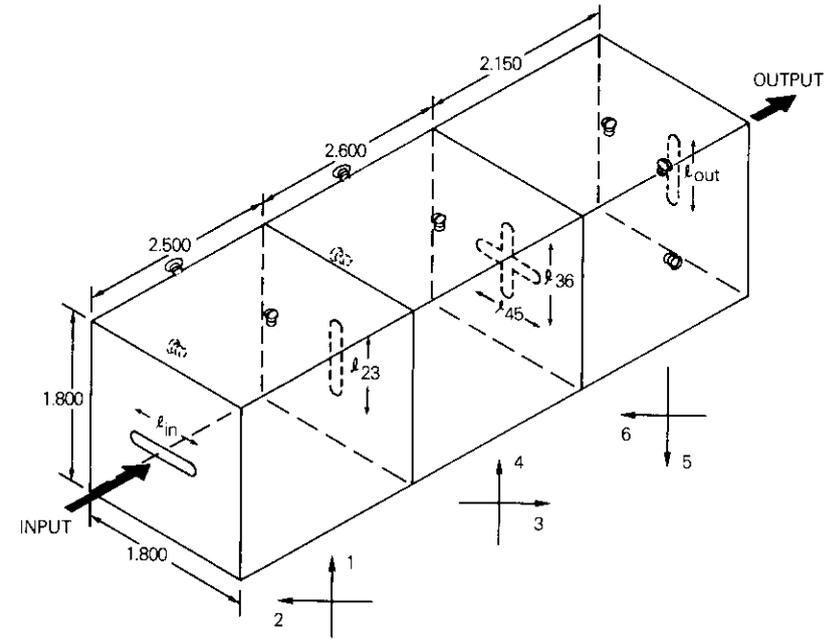


Figure 4. Filter Configuration for a 6-Pole Dual Mode Pseudo-elliptical Function Filter

are indicated by the six arrows in the figure. Couplings k_{12} , k_{34} , and k_{56} are realized by the coupling screws located at the corner of each cavity; couplings k_{23} , k_{45} , and k_{36} are realized by coupling slots. The input and output slots, as well as all the coupling slots, are designed to have the same width (0.125 in.) and lengths as follows [9]:

$$\left. \begin{aligned} l_{\text{in}} &= 1.268 \text{ in.} \\ l_{\text{out}} &= 1.388 \text{ in.} \\ l_{23} &= 0.965 \text{ in.} \\ l_{45} &= 1.091 \text{ in.} \\ l_{36} &= 0.990 \text{ in.} \end{aligned} \right\} \quad (21)$$

Tuning method for singly terminated filters

Conventional techniques of filter tuning are based upon the return loss characteristics of a filter. Minimizing the reflection in the passband with correct center frequency and bandwidth is usually the criterion for tuning. Since the input port for a singly terminated filter is not matched over the entire passband due to the existence of passband reactance, the criterion of minimum reflection cannot be used for filter tuning. A method of filter tuning based upon the filter's short-circuit impedance has been applied to tune the singly terminated filter.

The short-circuit filter tuning method is based upon a technique [10] for intercavity coupling measurement in which the tuning and coupling screws are set one by one according to the match of measured and computed input impedances for a short-circuited filter. Therefore, correct network representation is extremely important. Consider a multiple-coupled cavity network whose currents and voltages are related by an impedance matrix Z as shown by equation (6). This impedance matrix has off-diagonal elements which are the intercavity coupling coefficients as shown in equation (19) and diagonal elements which are the resonance expressions for the individual cavities as follows:

$$Z_{ii} = \frac{1}{Q_u BW} + j \frac{2}{\pi BW} \left(\frac{\lambda_{oi}}{\lambda_{g_{oi}}} \right)^2 \tan \left(\pi \frac{\lambda_{g_{oi}}}{\lambda_g} \right) \quad i = 1, 2, \dots, N \quad (22)$$

When cavity i is at resonance, wavelength λ_{oi} and guide wavelength $\lambda_{g_{oi}}$ are set according to the mid-band frequency of the filter. When cavity i is not in resonance, $\lambda_{g_{oi}}$ and λ_{oi} are set according to a detuned frequency measured in advance. This impedance matrix and the transformer ratios given by equation (16) completely define the network shown in Figure 1, with the input at AA' and output at BB' . A short-circuited filter is a 1-port network with a short circuit at BB' and the input at AA' . Thus, this short-circuited filter may be used to individually set the tuning and coupling screws as the cavity resonance condition is changed progressively from a detuned to a tuned condition. However, the reference plane for a waveguide filter in the tuning setup is usually at the input slot, which is unfortunately not plane AA' on the network model in Figure 1. Therefore, the network model would have to be modified to have a reference plane at the location of the input slot.

The first or last section of a waveguide filter is usually an impedance inverter with the parameter determined by the input or output impedance

of the filter. The network in Figure 1 may be reconstructed as shown in Figure 5, where two impedance inverters have been added in cascade at

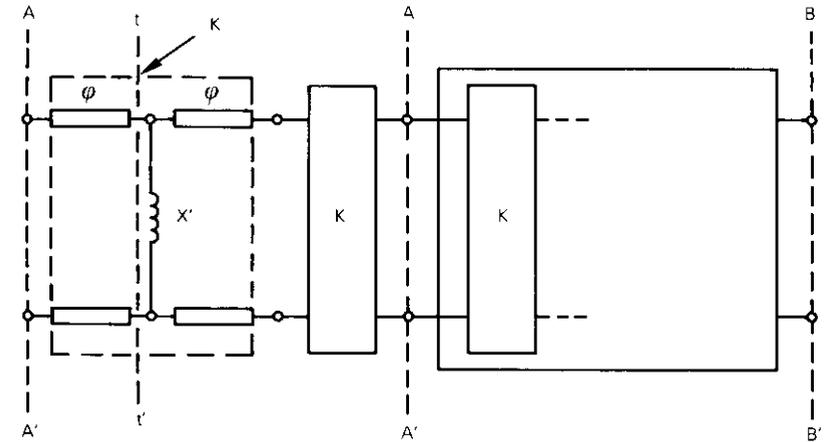


Figure 5. Modified Representation with Proper Reference Plane

the input port. Adding two impedance inverters does not change the impedance characteristics for the filter but provides access to the inside of an impedance inverter without disturbing the network representation from AA' to BB' . Shunt inductance X' is the normalized inductance for the input slot and [9]

$$\phi = -\frac{1}{2} \tan^{-1} 2X' \quad (23a)$$

$$K = \tan |\phi| \quad (23b)$$

As shown in Figure 5, a network with an input port at tt' may be realized by adding an impedance inverter K , a section of 1-ohm transmission line with a phase shift of ϕ , and a shunt impedance X' in front of the input port of the original network. Shunt impedance X' represents the input slot; the tt' plane is the location of the input slot in a practical waveguide filter. Since this network has access to the inductance that represents the input slot, the frequency dispersion of the input slot may also be conveniently added.

The experimental 6-pole pseudo-elliptic filter is tuned by dividing the filter into two short-circuited pieces and proceeding from both input and

output ports. If two cavities (four modes) are attached to the input port and one cavity (two modes) is attached to the output port, tuning screws, 1, 2, 3, and 4 and the coupling screws (1, 2) and (3, 4) may be set by the input port tuning; tuning screws, 6 and 5 and coupling screw (5, 6) may be set by the output port tuning. After the tuning screws and coupling screws have been set, the filter may be reassembled. It should have the correct responses without adjustment. The transmission loss, return loss, and reflection phase performance of the experimental model tuned by using this short-circuit tuning method are presented in Figures 6, 7, and 8. Excellent agreement between the computed and measured performances was obtained.

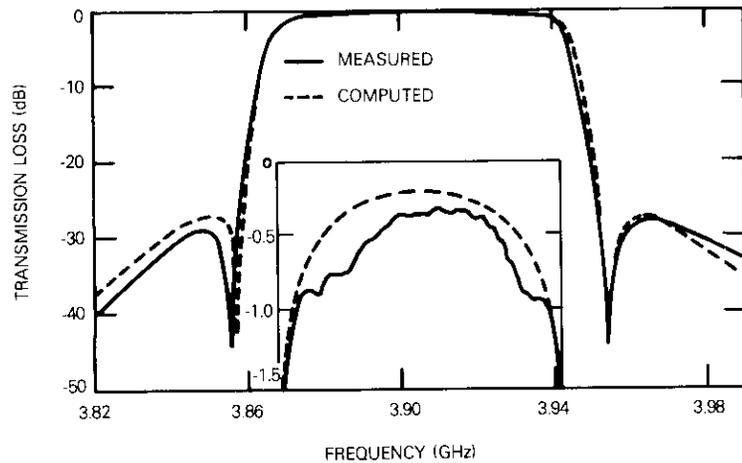


Figure 6. *Computed and Measured Transmission Loss Characteristics*

Conclusions

Methods for synthesis, realization, and tuning of a singly terminated pseudo-elliptic filter have been presented. The methods of synthesis and realization have been generalized so that they are suitable for filters with any number of poles. The design procedures for the experimental model have been described in detail to serve as a guide for filter design.

The short-circuit tuning method described in this paper is a step-by-step tuning method which is both accurate and effective. It is applicable to singly as well as doubly terminated filters.

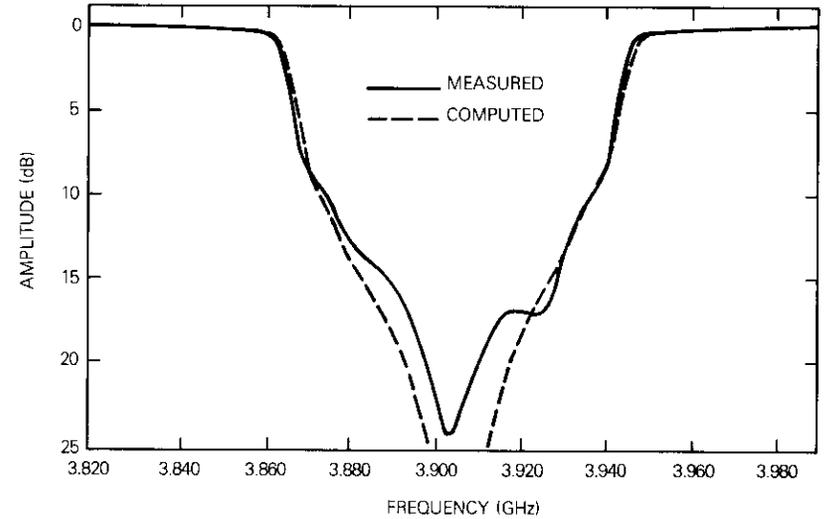


Figure 7. *Computed and Measured Return Loss Characteristics*

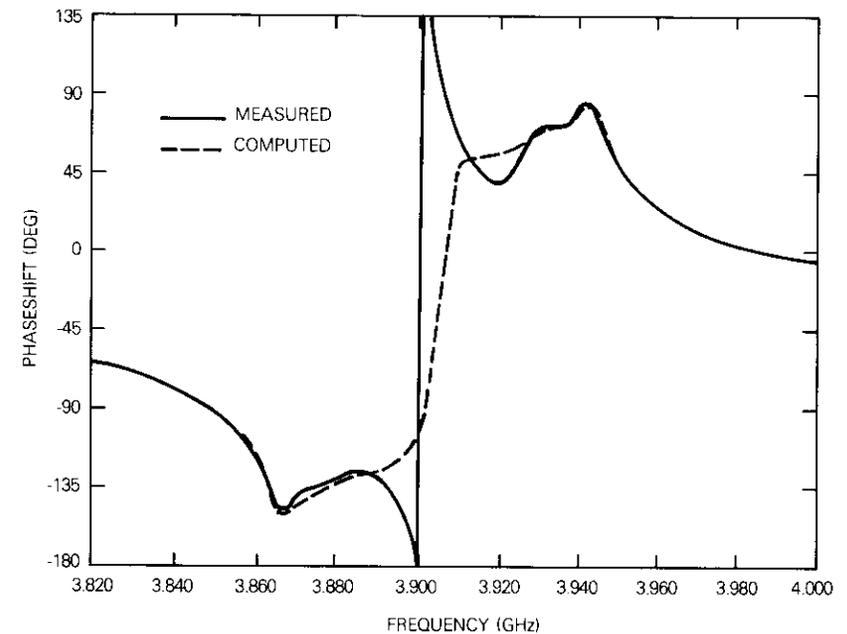


Figure 8. *Computed and Measured Reflection Phase Characteristics*

The correct tuning for the singly terminated filter is the key to successful construction of a contiguous band multiplexer. The out-of-band reactance, which has usually been disregarded in conventional tuning, plays an important role in the contiguous band multiplexer. The out-of-band reflection phase can be correctly tuned by using the short-circuit tuning method described herein.

Acknowledgments

The author wishes to thank Dr. C. Mahle for his encouragement and guidance and Mr. R. Johnson for his help in filter tuning.

References

- [1] M. Chen, F. Assal, and C. Mahle, "A Contiguous Band Multiplexer," *COMSAT Technical Review*, Vol. 6, No. 2, Fall 1976, pp. 285-306.
- [2] J. K. Swirzynski, *Design Theory and Data for Electrical Filters*, D. Van Nostrand Co., 1965, pp. 78-115.
- [3] S. Darlington, "Synthesis of Reactance 4-Poles which Produce Prescribed Insertion Loss Characteristics," *Journal of Mathematics and Physics*, Vol. 18, No. 4, September 1939, pp. 257-353.
- [4] F. F. Kuo, *Network Analysis and Synthesis*, New York: John Wiley and Sons, Inc., 1966, pp. 347-352.
- [5] A. E. Atia and A. E. Williams, "New Types of Waveguide Bandpass Filters for Satellite Transponders," *COMSAT Technical Review*, Vol. 1, No. 1, Fall 1971, pp. 21-43.
- [6] F. R. Gantmacher, *The Theory of Matrices*, Vol. I, New York: Chelsen Publishing Co.
- [7] A. E. Atia, A. E. Williams, and R. W. Newcomb, "Narrow-Band Multiple-Coupled Cavity Synthesis," *IEEE Transactions on Circuits and Systems*, CAS-21, No. 5, September 1974, pp. 649-655.
- [8] M. Saito, "Synthesis Procedure for General Coupled-Resonator Transmission Networks," *Electronics and Communications in Japan*, Vol. 53-A, No. 5, 1970, pp. 26-33.
- [9] G. L. Matthaei, L. Young, and E. M. T. Jones, *Microwave Filters, Impedance-Matching Networks, and Coupling Structures*, New York: McGraw-Hill Book Company, 1964.
- [10] A. E. Atia and A. E. Williams, "Measurement of Intercavity Couplings," *IEEE Transactions on Microwave Theory and Techniques*, MTT-23, June 1975, pp. 519-522.

Ming Hui Chen received a B.S.E.E. from Cheng Kung University, Taiwan, China, in 1960; an M.S. in electronics from Chiao Tung University, Taiwan, China, in 1962; an M.S.E.E. from Utah State University in 1964; and a Ph.D. in electrophysics from the Polytechnic Institute of Brooklyn in 1969. Before joining COMSAT, he was Chief Microwave Engineer at Microtech Co., a principal engineer at Radiation Systems, Inc., and a member of the technical staff at M.I.T. Lincoln Laboratory. In 1970 he was appointed Assistant Professor at George Washington University. He joined COMSAT Laboratories in 1974 as a member of the technical staff in the Transponders Department, Microwave Laboratory, responsible for microwave filter development. His research interests include microwave devices, antennas, and wave propagation in periodic structures.



Phase correction of IMPATT amplifiers

W. J. GETSINGER AND H. L. HUNG

(Manuscript received Nov. 3, 1976)

Abstract

It has been observed that a phase modulated signal amplified by a multistage IMPATT amplifier results in sideband compression at the output. This compression problem is related to phase distortion produced by IMPATT diode non-linearity. A simple passive device termed a spectrum conditioner effectively adjusts the spectral phases of the input signal controlling the sideband-to-carrier ratio at the output. This spectrum conditioner was used in F2 and F4 28-GHz amplifiers to achieve the specified performance.

Introduction

The Centimeter Wave Beacon 28.56-GHz IMPATT amplifier [1] is driven with a phase modulated signal of about 7.5 dBm, which has first-order sidebands 5 to 6 dB below the carrier level. The total output power of the IMPATT amplifiers as originally constructed was at the required level of about 30 dBm, but the sidebands were smaller than those at the input by 1 or 2 dB for 264.4-MHz modulation and by 4 or 5 dB for 528.9-MHz modulation. Swept frequency responses of the amplifiers were too broad to explain this in terms of bandwidth limitations.

Analysis of the nonlinear amplifier showed that the output sideband-to-carrier ratio depended strongly on the input phase relationship between the carrier and sidebands and that the IMPATT amplifiers distorted the

spectral phases. Thus, the observed sideband compression is inherent to a large-signal IMPATT amplifier. Adjustment of the spectral phases before amplification was indicated to correct the problem.

Although signal distortion by IMPATT amplifiers is well-known, little has been published on corrective measures. A technique has been described which compensates dynamically for both resistive and reactive parts of diode nonlinearity by using video amplification of a detected sample of the input signal to adjust the bias of each following IMPATT stage [2].

This paper describes a static technique which compensates the phase distortion produced by the diode nonlinearity by using a single, simple, passive device termed a spectrum conditioner. This technique was used in the F2 and F4 beacons to restore the output sideband-to-carrier ratio to its input value.

Background

An analytical model of the IMPATT amplifier revealed a different level of complex saturation for the carrier than for the sidebands. The sideband-to-carrier ratio at the output depended strongly on the relative phases of the carrier and the sidebands at the input. The term spectrum angle is introduced to relate the phases of the three input frequencies. Small index sine wave modulation can be represented by adding contra-rotating sideband phasors to a fixed carrier phasor. If the sideband phasors add in phase with the carrier phasor, amplitude modulation results. If the sideband phasors add in quadrature with the carrier phasor, phase modulation (approximately) results as illustrated in Figure 1. Spectrum angle ψ is the angle with respect to the carrier where the sideband phasors add constructively. Thus, the phase relationships among the three spectral components are characterized by a single term, ψ . If the signal is described as

$$f(t) = A_0 \cos(\omega_0 t + \theta_0) + A_{-1} \cos[(\omega_0 - \omega_1)t + \theta_{-1}] + A_1 \cos[(\omega_0 + \omega_1)t + \theta_1] \quad (1)$$

where ω_0 is the carrier angular frequency, ω_1 is the modulation angular frequency, and subscripts 0, -1, and 1 denote the carrier, lower sideband, and upper sideband, respectively. The spectrum angle, ψ , is given by

$$\psi = \frac{1}{2}(\theta_1 + \theta_{-1}) - \theta_0 \quad (2)$$

Gain at each spectral component as a function of the input spectrum

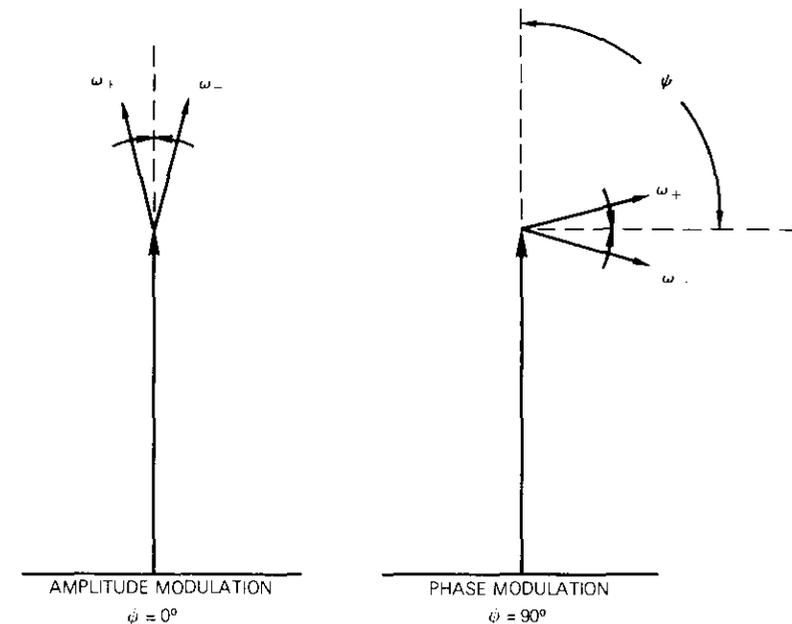


Figure 1. Spectrum Angle, ψ

angle is shown in Figure 2 for an analytical model of a large-signal IMPATT amplifier stage. At the input, each sideband is 6 dB below the carrier.

A saturated amplifier without phase distortion would suppress the sidebands at a spectrum angle of 0° . Instead, suppression occurs at 45° because of the complex nature of the IMPATT diode nonlinearity. Power is being exchanged between carrier and sidebands as the spectrum angle varies. Plots of total output power vs spectrum angle, as shown in Figure 3, indicate that power varies about 1 dB over the full 180° of the spectrum angle with little slope in the region of interest around $\pm 90^\circ$. Thus, the sideband-to-carrier ratio can be adjusted by adjusting the input spectrum angle.

One means of adjusting the spectrum angle is with a circulator terminated in a low-loss resonator which is tuned off the carrier frequency similar to a phase equalizer section. This device, termed a spectrum conditioner, is shown schematically in Figure 4. A calculated example of the spectrum angle achievable with a spectrum conditioner is given in Figure 5 for two resonator slope parameters as a function of the ratio of resonator center frequency to carrier frequency. The modulation frequency is $1/54$

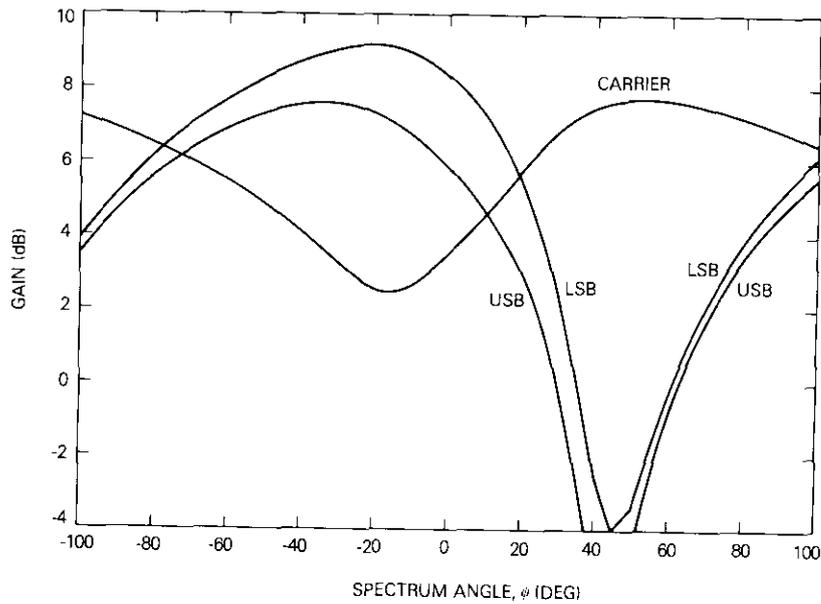


Figure 2. Gain vs Spectrum Angle, Large-Signal State

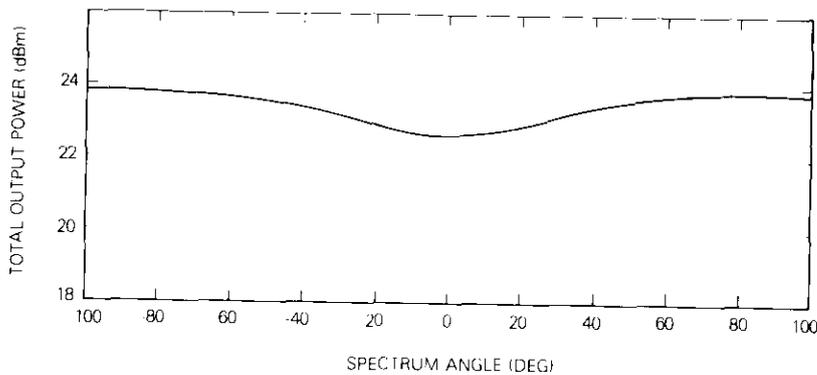


Figure 3. Total Output Power vs Spectrum Angle

of the carrier frequency. Figure 5 shows that spectrum angles from -90° to $+90^\circ$ can be generated with a single spectrum conditioner.

The curves of Figure 5 were calculated by assuming an input signal such as equation (1) to the resonator with phases θ_0 , θ_{-1} , and θ_1 all zero.

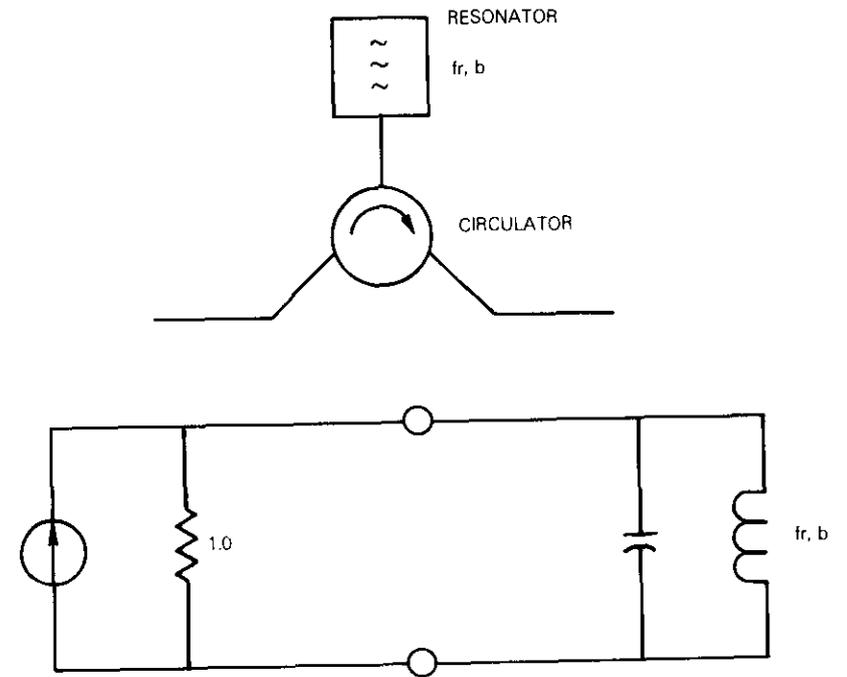


Figure 4. Spectrum Conditioner and Resonator Circuit

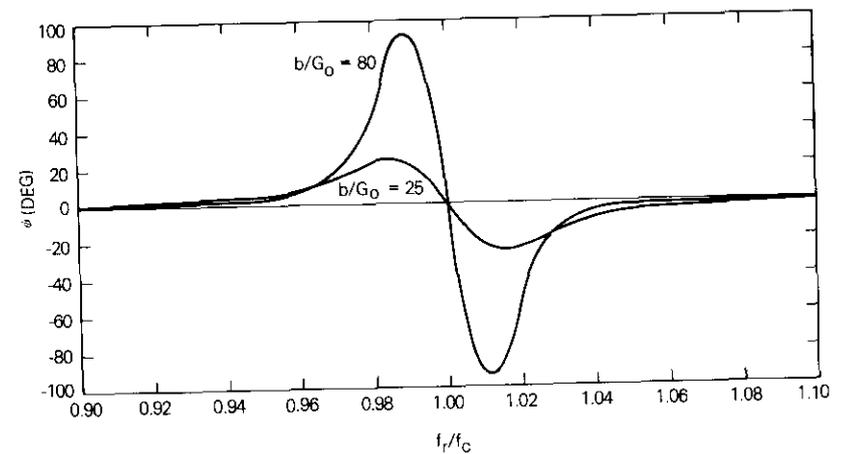


Figure 5. Spectrum Angle vs Resonator-to-Carrier-Frequency Ratio

Then the phase of the reflection coefficient for each spectral component was calculated from the expression

$$\theta = 180^\circ - 2b \arctan \left(\frac{\omega}{\omega_r} - \frac{\omega_r}{\omega} \right) \quad (3)$$

based on the simple spectrum conditioner resonator circuit representation of Figure 4.

In equation (3), b is the slope parameter of the resonator, ω_r is its resonant angular frequency, and ω is the angular frequency of the spectral component.

The solutions for all components of equation (3) are substituted into equation (2) to find ψ for a given resonator to carrier frequency ratio.

Spectrum conditioner design

The analytical models of the IMPATT amplifier stages demonstrated the nature of the sideband suppression problem and indicated a correction technique. However, since the representation of the actual amplifier was quantitatively uncertain, the correcting spectrum conditioner was designed on an empirical basis.

Correcting the entire 3-stage amplifier with a single spectrum conditioner requires the introduction of a sufficiently large spectrum angle to compensate the full 4- to 5-dB change in sideband-to-carrier ratio. Figure 2 indicated that a spectrum angle of about 30° yielded the required correction. Figure 5 showed that a normalized slope parameter of about 30 for the resonator provides such a spectrum angle at a resonator-to-carrier frequency ratio of about 0.985. Although a larger slope parameter could give the needed spectrum angle, working at the peaks of curves such as those of Figure 5 minimizes the tuning sensitivity. Also, use of the smallest possible slope parameter is preferable, since the spectrum conditioner will have the smallest dissipation loss. Therefore, a circulator-mounted waveguide resonator was designed with shims to change the frequency and replaceable irises to change the slope parameter.

With a single spectrum conditioner correcting the sideband levels of a complete 3-stage 28.5-GHz beacon IMPATT amplifier, the optimal location for the resonator was after the first medium power stage. The carrier-to-sideband levels were less sensitive to cavity size tuning than with the resonator placed before the first stage or after the second. Figure 6 shows

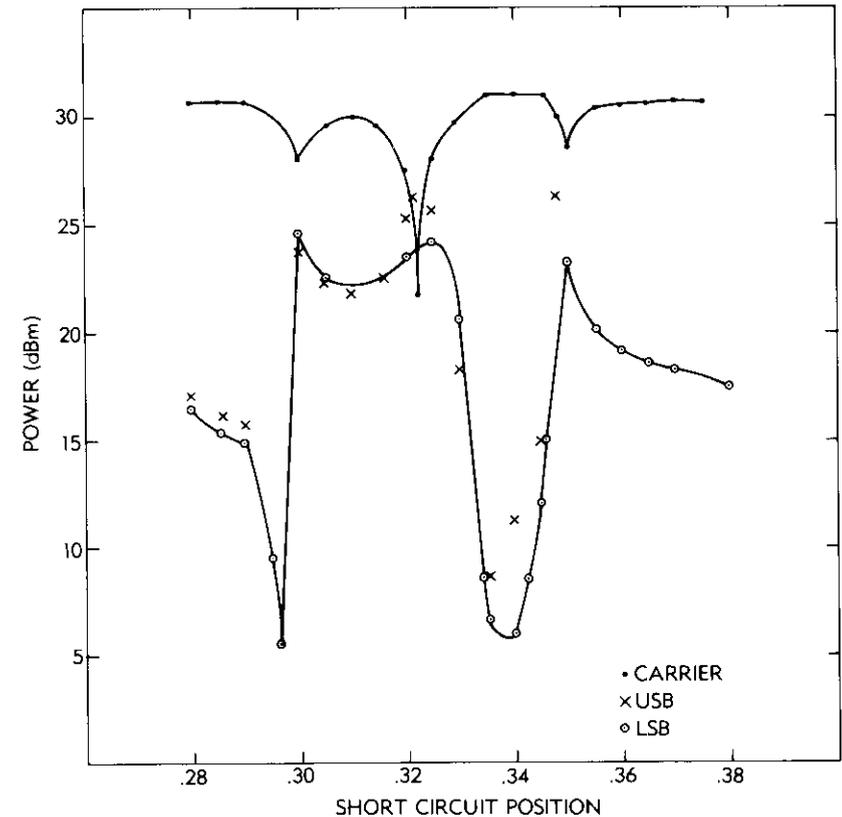


Figure 6. Spectral Levels vs Resonator Tuning, 3-Stage Amplifier

the output behavior of a representative 3-stage amplifier with the spectrum conditioner at the optimal location. In the region around the short-circuit position of 0.31, both the carrier and the sidebands were well-behaved with respect to the cavity tuning, and their levels exceeded the specifications. The effect of the spectrum conditioner on amplifier performance at nominal input power over the required temperature range (0°C – 50°C) is illustrated in Figure 7. The sidebands improved by 4 dB at the expense of only 1 dB for the carrier with the spectrum conditioner.

To investigate the effectiveness of the spectrum conditioner, experiments were performed with reduced input power. If a single resonator with slight modification of the first amplifier module is used, the carrier and the sidebands of the 3-stage amplifier can reach specified output levels over

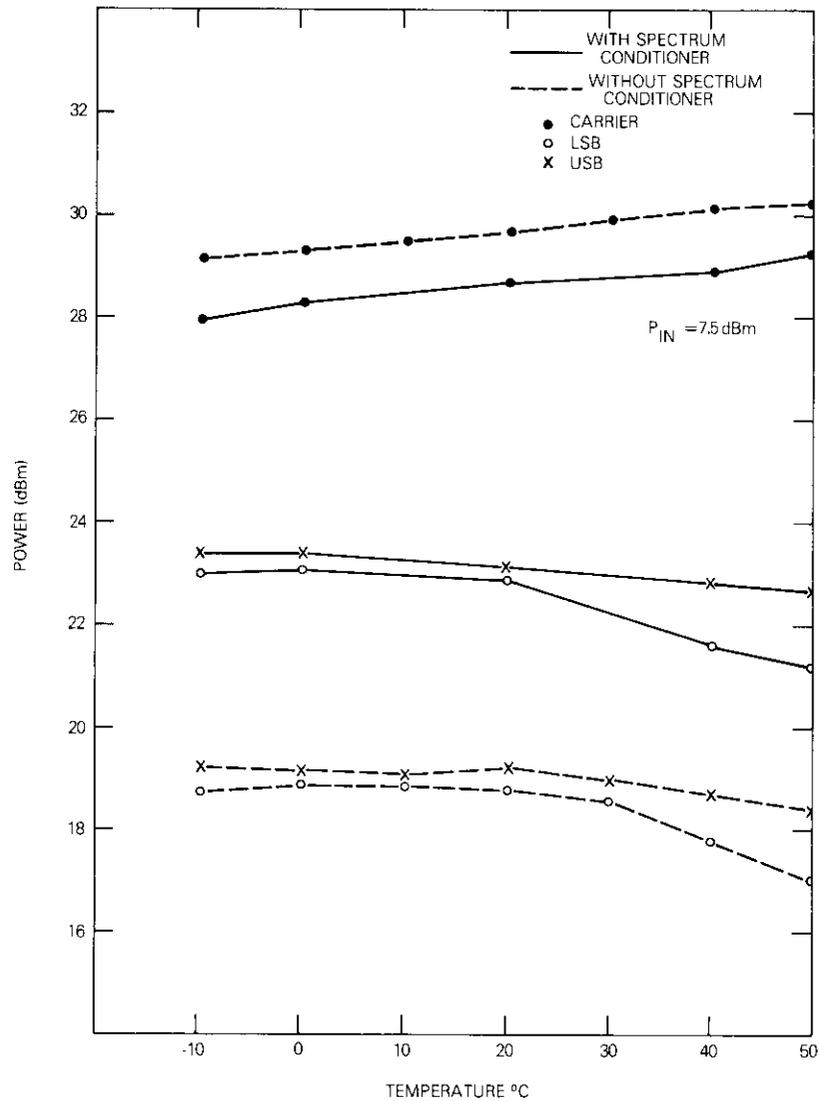


Figure 7. Spectral Levels vs Temperature with and without Spectrum Conditioner

the required temperature range even with a 3.5-dB reduction in input power. Figure 8 shows the results of the output of the amplifier with an

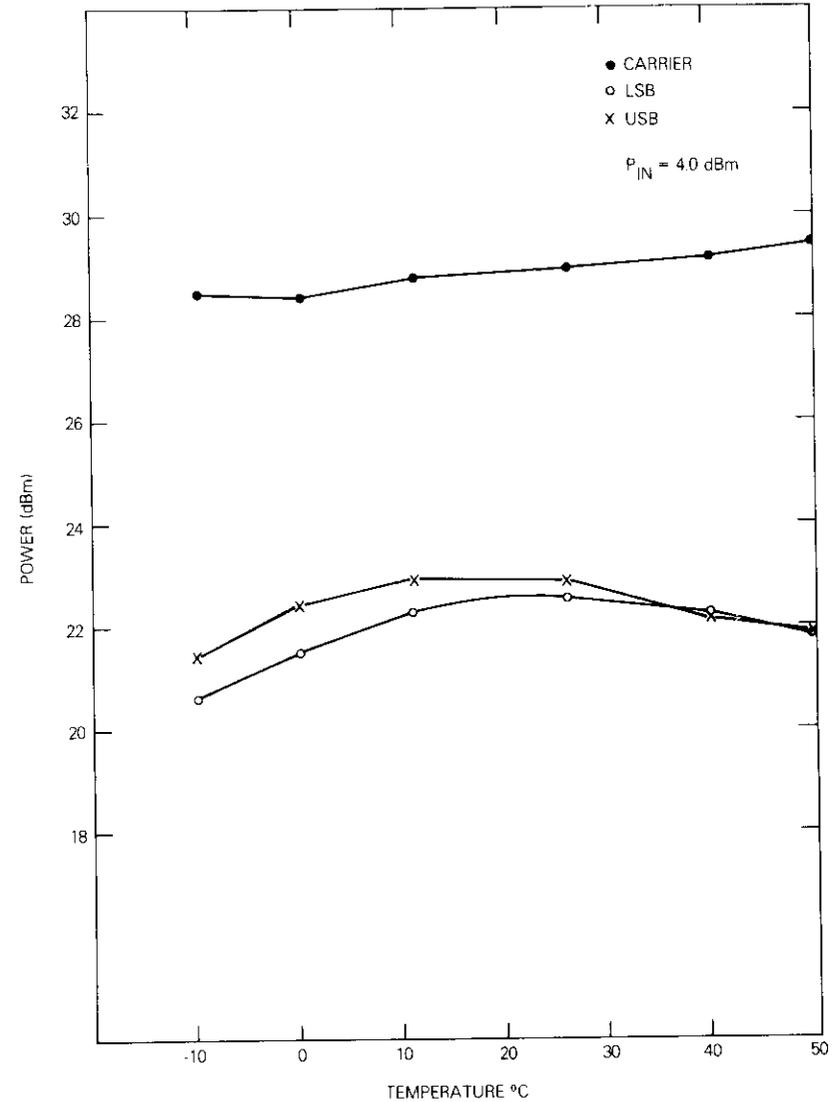


Figure 8. Spectral Levels vs Temperature (reduced input level)

input signal of 4.0 dBm rather than the nominal 7.5 dBm.

Figure 9 is a sketch of the spectrum conditioner resonator in beacons F2 and F4. Resonant frequencies of the two cavities were calculated from

their physical dimensions. The frequency of F4 was measured to within 0.4 percent of the calculated value. Calculated values of the ratio of resonator center frequency to carrier frequency and slope parameter with the physical dimensions of the resonators used for the two flight units are listed in Table 1.

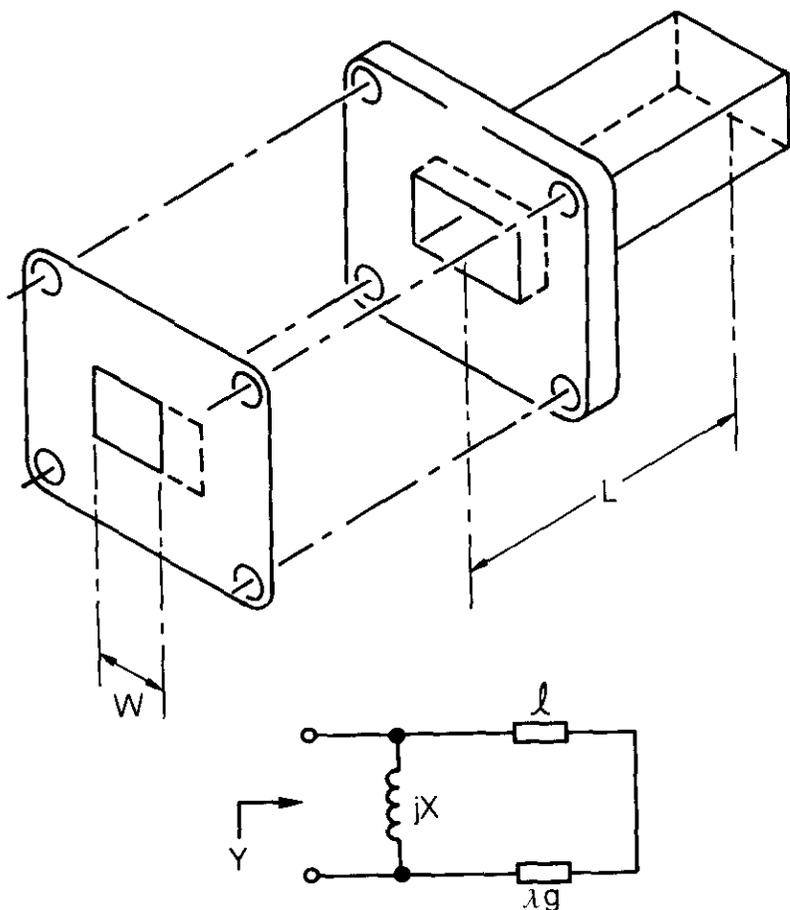


Figure 9. Spectrum Conditioner Resonator and Circuit

TABLE 1. PHYSICAL AND CALCULATED ELECTRICAL PARAMETERS OF F2 AND F4 SPECTRUM CONDITIONERS

Parameter	Unit	
	F-2	F-4
Cavity Length (L)	1.45 cm	1.47 cm
Iris Width (W)	0.47 cm	0.36 cm
Guide Wavelength (λ_g)	1.60 cm	1.52 cm
Resonant Frequency (f_r)	28.22 GHz	28.92 GHz
f_r/f_{center}	0.988	1.013
Normalized Slope Parameter (b/Y_0)	22	170
Estimated Spectrum Angle (ψ)	22°	-155°

The parameters of the F4 spectrum conditioner were determined almost entirely by experiment, and although performance was satisfactory, the slope parameter was larger than necessary, as shown by subsequent analysis. The F2 conditioner was designed following the principles outlined in this paper and achieved minimum slope parameter and correct frequency tuning. The F4 conditioner has different frequency tuning and slope parameter, but Figures 2 and 5 indicate that it provides the same sideband correction as the F2 conditioner.

Conclusion

A simple spectrum conditioner design based on the results of an analytical study effectively corrects the phase distortion caused by the nonlinearity of a multistage IMPATT amplifier. The carrier-to-sideband ratio was controlled and specified output levels of the F2 and F4 28-GHz beacons were achieved.

References

- [1] M. J. Barrett, "19- and 28-GHz IMPATT Amplifiers," *COMSAT Technical Review*, Vol. 7, No. 1, Spring 1977, pp. 129-152.
- [2] H. Komizo et al., "Improvement of Nonlinear Distortion in an IMPATT Stable Amplifier," *IEEE Transactions on Microwave Theory and Techniques*, MTT-21, November 1973, pp. 721-728.



William J. Getsinger was born in Waterbury, Conn., on January 25, 1924. He received a B.S.E.E. from the University of Connecticut in 1949 and an M.S.E.E. and the degree of Engineer in electrical engineering from Stanford University in 1959 and 1961. Since 1950 he has worked on microwave components at Technicraft Laboratories, the Westinghouse Electric Company, Stanford Research Institute, and M.I.T. Lincoln Laboratories. In 1969 he joined COMSAT Laboratories, where he is presently Manager of the Microwave

Circuits Department. Mr. Getsinger was Project Manager of the COMSTAR Centimeter Wave Beacon Project at COMSAT Laboratories.

Hing-Loi Hung received a B.S. in electrical engineering from M.I.T., and an M.S. and a Ph.D. from Cornell University. His research at Cornell concerned microwave solid-state device fabrications and circuit designs. In 1974, he joined COMSAT Laboratories, where he is presently a member of the technical staff of the Microwave Circuits Department, Microwave Laboratory.



Index: COMSTAR, beacon, frequency doubler, varactor diode

High-performance frequency doublers for the COMSTAR beacon

R. MOTT

(Manuscript received May 27, 1977)

Abstract

This paper describes the design approach for the high-performance, space-qualified, 19- and 28.5-GHz varactor frequency doublers used in the experimental, satellite-borne, centimeter wave beacon transmitters. The procedure for final optimization of the frequency conversion efficiency and a discussion of the measured transfer characteristics are presented. Power conversion efficiencies as high as 60 percent were attained for flight units, with both models exhibiting stable, oscillation-free behavior over a temperature range of -45°C to $+40^{\circ}\text{C}$ in hard vacuum.

Introduction

High conversion efficiency and excellent signal purity are provided by the 9.52- to 19.04-GHz and 14.28- to 28.56-GHz varactor frequency doublers of the centimeter wave beacons on the COMSTAR domestic communications satellites [1]. These rugged, compact frequency multipliers, designed and fabricated at COMSAT Laboratories, exhibit stable, oscillation-free behavior under all vibration and thermal vacuum conditions encountered during the launch phases and in orbit.

The design approach, based on computer-aided circuit analysis, yielded near-optimum performance. Final optimization was obtained from small-

signal measurements on the actual circuit using bias sweeping techniques [2]-[4]. The optimized design, realized in aluminum split-block form, is lightweight and mechanically rugged and contains no sensitive tuning elements.

Doubler circuit design

Circuit description

Figure 1 is a cross section of the doubler circuit. A capacitive probe couples the input signal from the input waveguide to the input coaxial impedance transformer and low-pass filter. The characteristic impedance of the filter is equal to the real part of the diode impedance at the input frequency. At the input frequency the diode is tuned by the inductance of the output waveguide below its cutoff frequency. At the output frequency it is tuned by both the output waveguide back-short inductance and the stopband reactance of the low-pass filter. The output waveguide impedance transformer is designed to present the optimum load impedance to the diode. The bias low-pass filter provides decoupling from an adjustable external bias resistor and stabilization against potential parametric oscillations.

Mounted diode equivalent circuits

Successful doubler design depends on proper tuning and matching of the varactor at the input and output frequencies. Figure 2 is the equivalent circuit of the mounted diode at the input frequency. The elements are as follows:

- Z_{ji} = varactor junction operating impedance at the input frequency
- L_D = diode package lead inductance
- C_1 = diode package shunt capacitance
- L_{1i} = inductance external to the diode package
- L_2, L_3 = inductances of the output waveguide below its cutoff frequency
- R_{si} = series resistance of diode and circuitry at the input frequency
- $R_{IN(EXT)}$ = real part of the packaged and mounted diode input impedance.

The internal package elements are either measured according to the IEEE standards [5] or obtained from the manufacturer. The external inductance, L_{1i} , is obtained using the method suggested by Getsinger [6]. The output waveguide, operating below its cutoff frequency at the input frequency, produces inductances L_2 and L_3 . These inductances are calculated from output waveguide reactances X_2 and X_3 , shown in Figure 3,

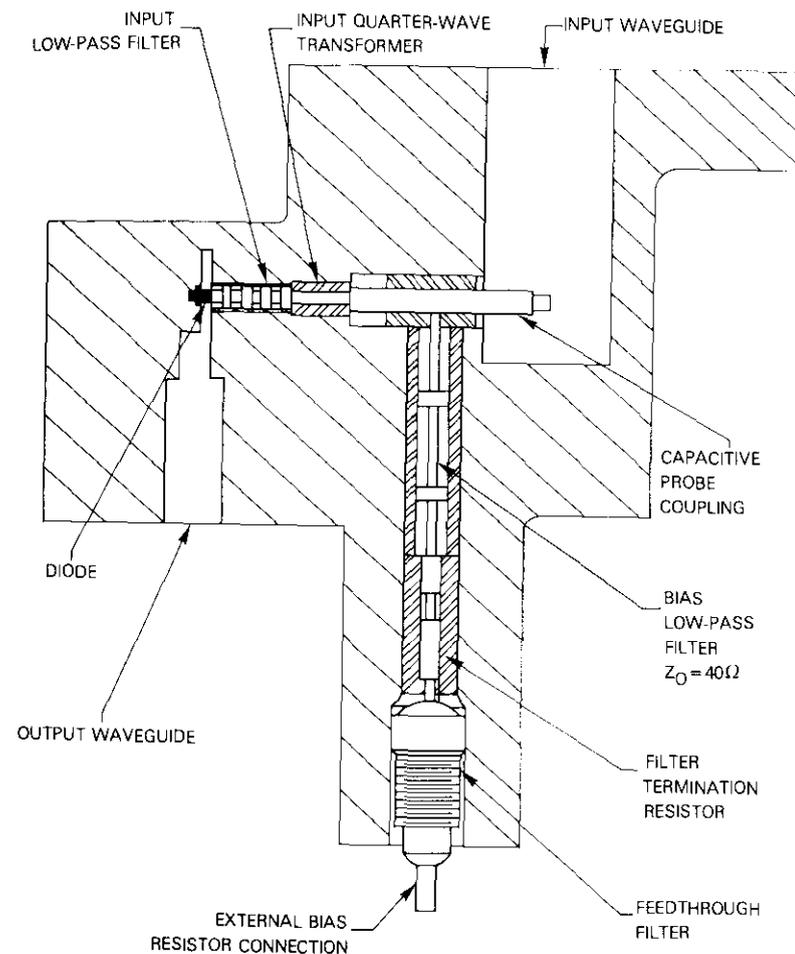


Figure 1. Doubler Circuit Layout

using equations (2) and (3). The characteristic impedance of the output waveguide for frequencies below cutoff is imaginary:

$$Z_o = j \left(\frac{2b}{a} \right) \sqrt{\frac{\mu}{\epsilon}} \left[\frac{1}{\sqrt{(f_c/f)^2 - 1}} \right], f < f_c \quad (1)$$

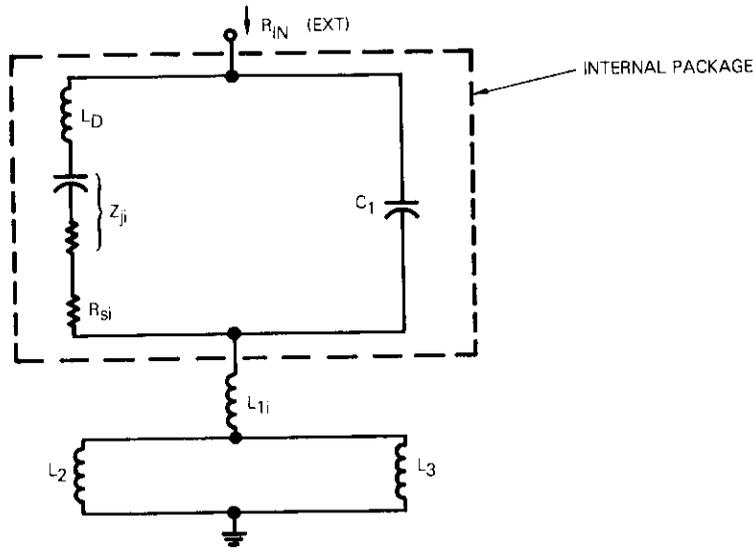


Figure 2. Equivalent Circuit of the Mounted Diode at the Input Frequency

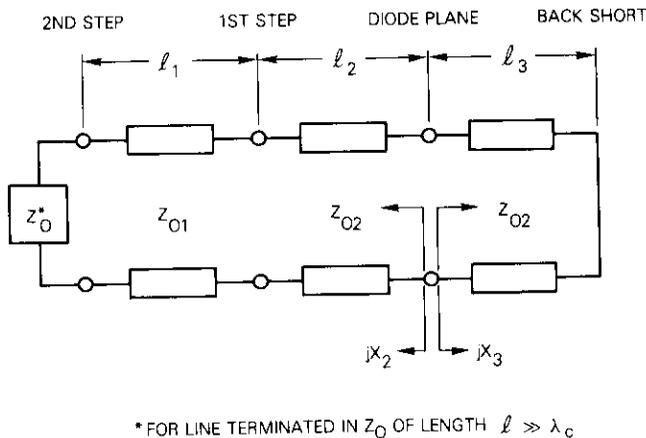


Figure 3. Equivalent Circuit of the Output Waveguide at the Input Frequency

where b and a are the waveguide height and width, respectively. Reactance X_2 is obtained by successive application of the impedance transformation equation for a transmission line below the cutoff frequency:

$$jX_i = Z_{0i} \left(\frac{Z_{Li} + Z_{0i} \tanh \gamma l_i}{Z_{0i} + Z_{Li} \tanh \gamma l_i} \right), i = 1, 2 \quad (2)$$

where X_i = input reactance
 Z_{Li} = load impedance of the i th line (imaginary)
 Z_{0i} = characteristic impedance of the i th line
 γ = propagation constant (real for waveguide below cutoff)
 l_i = length of the i th line

Reactance X_3 is given by

$$jX_3 = Z_{02} \tanh \gamma l_1 \quad (3)$$

The input reactance is tuned using computer-aided circuit analysis. The value of waveguide impedance Z_{02} is determined by the varactor diode package height. The input real part match is achieved by designing the low-pass filter with a characteristic impedance equal to the packaged and mounted diode input impedance real part, $R_{IN(EXT)}$.

The input probe transition is matched to 50-ohm coaxial line. The input low-pass filter impedance is matched to the 50-ohm line with a quarter-wave transformer.

The mounted diode equivalent circuit at the output frequency is shown in Figure 4. Internal package parameters L_D and C_1 are used at both the input and output frequencies, while R_{SO} is the series loss resistance at the output frequency only. External inductance L_{10} at the output frequency is determined with the same technique used to obtain the input circuit inductance, L_{1i} . The characteristic impedance of the coaxial line connecting the low-pass filter to the diode circuit is equal to the low-pass filter characteristic impedance, R_0 . The length, l , of this line, together with the out-of-band reactance of the low-pass filter, is adjusted to obtain the reactance required to resonate the diode at the output frequency. The optimum diode load resistance, R_L , is achieved by adjusting Z_{01} to transform the full height output waveguide characteristic impedance to R_L . All adjustments are performed using computer-aided design techniques before the circuit is fabricated.

The diode is terminated out of band in a real impedance to prevent parametric oscillations. The effectiveness of this real impedance out-of-band scheme depends on the proper choice of cutoff frequencies for the input and bias low-pass filters. The bias filter and the input waveguide form a diplexer network with a crossover frequency equal to the cutoff

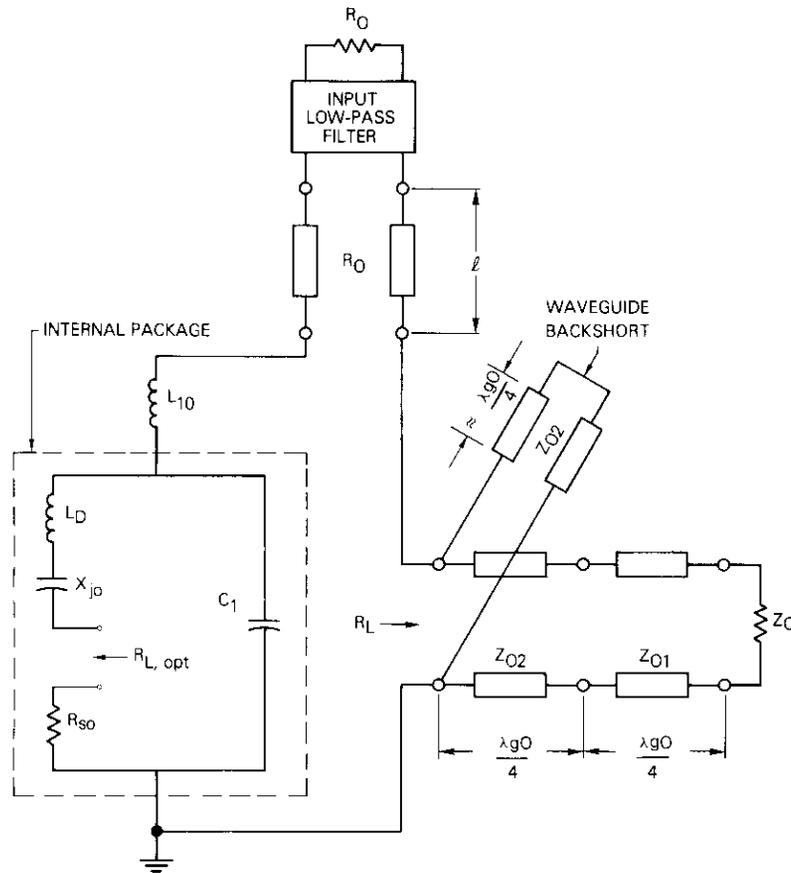


Figure 4. Equivalent Circuit of the Mounted Diode at the Output Frequency

frequency of the input waveguide. Similarly, the input low-pass filter and output waveguide form a diplexer with a crossover frequency equal to the cutoff frequency of the output waveguide. This technique terminates the varactor in a real impedance at all frequencies out of band, effectively inhibiting parametric instabilities.

Important design parameters

Two doubler circuits were designed around gallium arsenide varactors with two different package configurations. The important design param-

eters, listed in Table 1, are used to fabricate the initial hardware. Optimization of circuit performance requires slight circuit adjustment because of the approximations used to calculate parasitic reactances and the assumptions adopted for parasitic loss resistances.

TABLE 1. DESIGN PARAMETERS FOR TWO PACKAGE CONFIGURATIONS

Parameter	9.5- to 19-GHz Doubler	14.28- to 28.56-GHz Doubler
Diode		
C_j min	0.16 pF	0.11 pF
V_B	20 V	15 V
f_c (-6) V	400 GHz	500 GHz
Package Style	Micro-Pill	Min-Dot
L_D	0.16 nH	0.12 nH
C_1	0.2 pF	0.12 pF
Input Circuit		
Diode Junction [7]-[9] Impedance	10 Ω - j33 Ω *	10 Ω - j33 Ω *
L_{1i}	0.07 nH	0.04 nH
L_2	0.856 nH	0.445 nH
L_3	0.440 nH	0.225 nH
Signal Low-Pass Filter, R_o	6 Ω	10 Ω
Bias Low-Pass Filter, R_o	40 Ω	40 Ω
$R_{s,i}$ (assumed)	2.2 Ω	3.5 Ω
Output Circuit		
Diode Junction [7]-[9] Impedance	16.5 Ω - j16.5 Ω *	16.5 Ω - j16.5 Ω *
External Inductance L_{10}		
Input Low-Pass Filter Reactance	-j2.9 Ω	-j3.7 Ω
$R_{L, opt}$	13 Ω	15.4 Ω
R_{so} (assumed)	2.0 Ω	3.0 Ω

*Drive level, $M = 2$.

Doubler circuit performance

Bias sweeping measurement technique (cold testing)

Preliminary measurements of the doubler return loss showed that the circuits were closely but not optimally matched. Circuit performance could be improved by better matching (see Table 2). Each doubler was

"cold tested" to determine the input and output impedances presented to the diode and to define the coupling circuit elements which relate these impedances to their respective measurement planes.

TABLE 2. DOUBLER PERFORMANCE TABLE FOR COPPER ENGINEERING MODELS

Parameter	19-GHz Doubler Performance ($P_{IN} = +17.0$ dBm)			28-GHz Doubler Performance ($P_{IN} = +14.5$ dBm)		
	Measured		Pre- dicted*	Measured		Pre- dicted*
	Initial	Final		Initial	Final	
Conversion Loss (dB)	3.0	2.2	2.1	4.1	3.2	3.3
Return Loss (dB)	9.5	27.1		17.8	22.0	
1-dB Output Bandwidth (MHz)		500		1,200	1,700	
Efficiency, η (%)	50	60	62	39	48	47

*Diode theoretical loss + 0.5-dB circuit loss estimate [see equation (11)].

The varactor circuit cold test theory and procedure are discussed in detail in Reference 10. The following is a practical application of the theory. The bias sweeping measurement technique permits the input and output circuits to be decoupled so that matching and tuning can be accomplished separately. Measurements of return loss on an operating multiplier indicate only that a matching problem exists; cold testing reveals the location and nature of the problem, e.g., input tuning, input real part match, output tuning, or output real part match.

An equivalent circuit of the coupling between the input or output waveguide flange and the diode junction is shown in Figure 5. The elements are as follows:

- Z_{INPUT} = impedance presented to the varactor junction
- Z_j = varactor junction operating impedance (small signal)
- R_c = series resistance representing circuit losses
- R_s = diode junction series loss resistance

- L_s = series inductance which resonates the average varactor capacitance
- $S(V)$ = elastance as a function of junction voltage (V)
- R_{SH} = shunt loss resistance.

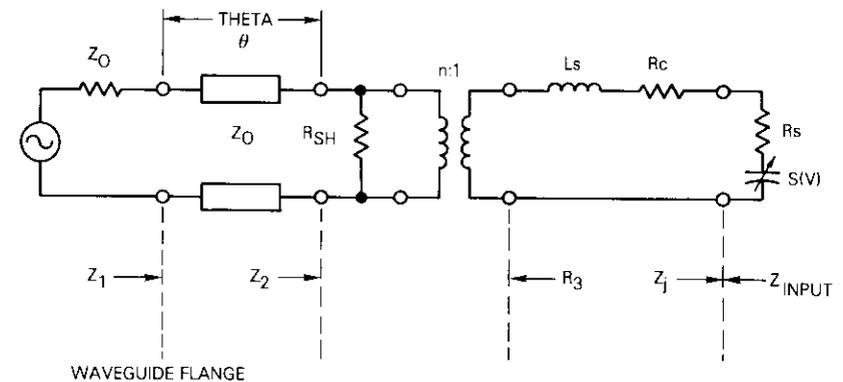


Figure 5. Small-Signal Equivalent Circuit Between the Varactor Diode Junction and the Input or Output Terminals

The circuit is correctly designed when the diode is resonant and the real part impedance is matched at both the input and output ports. The condition for resonance is given by

$$L_s = \frac{S_{01}}{\omega_o^2} \quad (5a)$$

at the input port, and

$$L_s = \frac{S_{02}}{(2\omega_o)^2} \quad (5b)$$

at the output port, where S_{01} and S_{02} are the average diode elastances at the input and output frequencies, respectively.

A real part match is achieved at the input when*

$$R_3 = R_c + R_s + R_{in} \quad (6a)$$

*The effect of R_{SH} is generally found to be negligible; hence, it has been omitted here.

and at the output when

$$R_3 = R_L - R_c - R_s \quad .$$

Under full power conditions, R_{in} is the real part of the pumped, lossless varactor junction impedance at the input frequency; R_L is the optimum load resistance for the lossless varactor. These values are obtained from the analyses given in References 7-9.

Note the difference between equations (6a) and (6b). When the diode and circuit losses are nontrivial, accurate values should be used to obtain the proper value of R_3 for matching.

The elements of the coupling circuit are derived from slotted-line, small-signal impedance measurements of the doubler as a function of junction bias voltage. The impedance points lie on a circle on a Smith chart. The angle through which the circle is rotated until it is coincident with a constant resistance circle is the electrical distance θ corresponding to shifting the reference plane of measurement from the waveguide flange to the Z_2 measurement plane of Figure 5. Figure 6 is the Smith chart plot, where

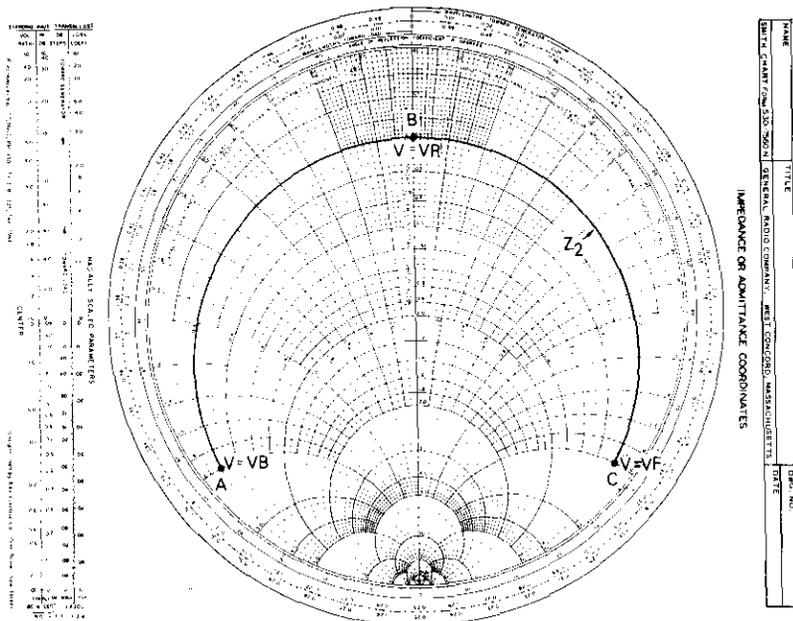


Figure 6. Typical Impedance Contour in the Z^2 Plane

points A, B, and C refer to Z_2 and are identified with respect to the notation in Figure 5 as follows. For point A, $V_o = V_B$, where V_B is the reverse breakdown voltage, and

$$Z_A = n^2 \left[R_c + R_s - j \frac{S_{MAX}}{\omega_o} + j(\omega_o L_s) \right] \quad (7)$$

where n is the transformer turns ratio as shown in Figure 5 and $S_{MAX} = S(V_B)$. For point B, $V_o = V_R$ (tuned condition),

$$\frac{S(V_R)}{\omega_o} = \omega_o L_s \quad (8)$$

and

$$Z_B = n^2(R_c + R_s)$$

For point C, $V_o = V_F$, where V_F is an arbitrary forward bias voltage required to produce the reactance shown in Figure 6, and

$$Z_C = n^2 \left(R_c + R_s - j \frac{S_{MIN}}{\omega_o} + j\omega_o L_s \right) \quad (9)$$

where $S_{MIN} = S(V_F)$.

The value of $S(V_o)$ is determined by measuring the elastance voltage characteristics of the varactor on a capacitance bridge at 1 MHz. Series inductance L_s can be determined from the resonance condition, equation (8). From Figure 6, it follows that

$$X_2(V_B) - X_2(V_F) = - \frac{n^2}{\omega_o} \left[S(V_B) - S(V_F) \right] \quad (10)$$

The values of $X_2(V_B)$ and $X_2(V_F)$ are taken from the Smith chart. The values of $S(V_B)$ and $S(V_F)$ are known from the measurements of $S(V)$ at 1 MHz; hence, n^2 is calculated from equation (10). The circuit efficiency can be calculated by determining $R_s + R_c$ from equation (7) or (9). Input and output circuit efficiencies are given by

$$\eta_{e,in} = \frac{1}{1 + (R_c + R_s)/R_{in}} \quad (11a)$$

and

$$\eta_{c,out} = \frac{L}{1 + (R_c + R_s)/R_L} \quad (11b)$$

respectively. Some high-frequency low-capacitance varactors contain a significant voltage-variable resistive term; the small-signal impedance is more accurately represented by

$$Z_i = R_s' - R_s''(V_o) - j \frac{S(V_o)}{\omega_o} \quad (12)$$

The voltage-variable resistive term $R_s''(V_o)$ is small and can be neglected when deimbedding the varactor on the 9.5- to 19-GHz doubler. However, for the 14- to 28-GHz doubler, this term must be included due to the lower capacitance and hence smaller junction area of the diode for the higher frequency multiplier. The resistive term is larger and the voltage-variable part more prominent [11].

Special techniques developed by Atia [12], which consider the voltage-variable resistance, were employed for deimbedding both multipliers.

Optimum performance impedance matching technique

The matching technique changes the impedance presented to the diode junction by the initial circuit design into that required for optimum performance by using the measured elements of the coupling circuit of Figure 5 to calculate impedance Z_{input} . Both the measured and desired Z_{input} are transformed through the coupling network into the waveguide of characteristic impedance Z_0 where the cold test measurements were performed. The two points Z_{MEAS} and Z_{DES} , plotted on a Smith chart in Figure 7, are referred to the waveguide flange for convenience. The corresponding admittance points, Y_{MEAS} and Y_{DES} , are rotated through an angle ϕ , preferably toward the diode,* to a new reference plane where $R_d[Y'_{MEAS}] = R_d[Y'_{DES}]$. The imaginary part difference, jB , indicates the susceptance of the iris required at the new reference plane for matching.

The input and output flanges are machined to the reference plane for the matching irises to facilitate installation. In addition, sets of irises are made in small increments of susceptance around the nominal value to allow unit-to-unit matching variations in a production run.

*It is desirable to effect the matching close to the diode for optimum bandwidth.

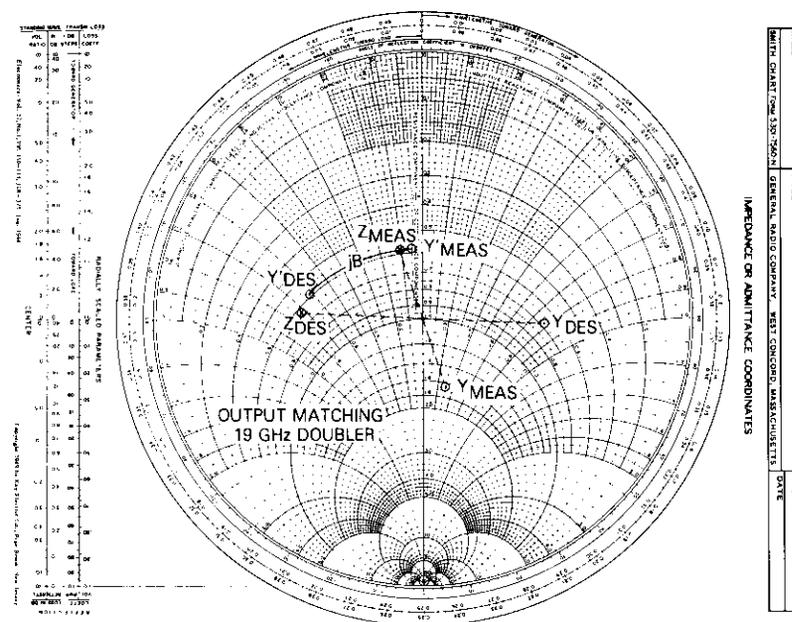


Figure 7. Iris Matching Technique

Doubler engineering models

The final swept frequency output power of the two doubler engineering models is shown in Figures 8a and 8b at various power levels. Gold-flashed copper parts are used throughout both models. The 28-GHz unit has considerably more bandwidth and a smoother response than the 19-GHz circuit due to a better diode match in the 28-GHz circuit and hence less perturbation from the matching irises. This demonstrates the importance of accuracy in the initial circuit design.

The matching effort increased the efficiency from 50 to 60 percent for the 19-GHz doubler and from 39 to 48 percent for the 28-GHz unit. Both models exhibited stable, well-behaved frequency responses over the temperature range of -60°C to $+40^{\circ}\text{C}$, and input power ranges to 740 mW for the 19-GHz model and 630 mW for the 28-GHz model.

Doubler flight models

The flight model doublers have iridited (iridite is a chromate conversion coating) aluminum bodies and gold-flashed copper center conductor parts.

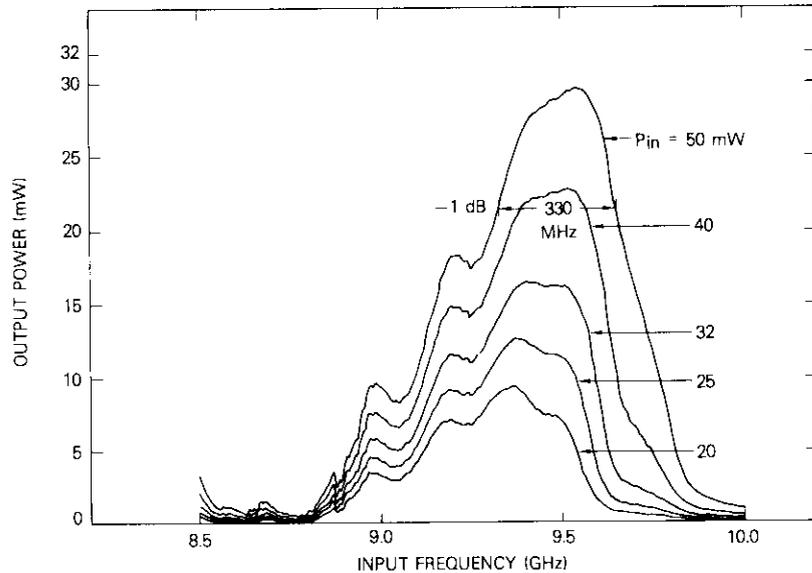


Figure 8a. Swept Frequency Output Power for the 19-GHz Copper Engineering Model Doubler

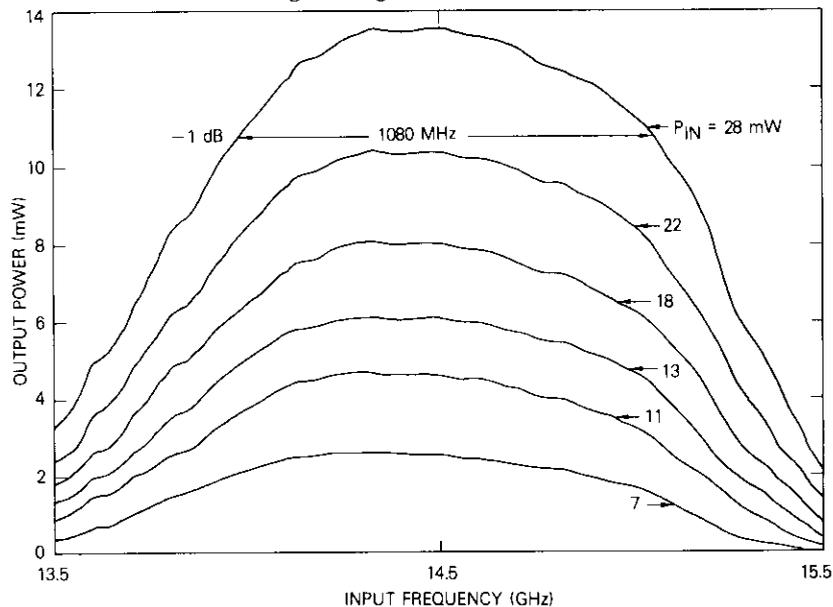


Figure 8b. Swept Frequency Output Power for the 28-GHz Copper Engineering Model Doubler

Five units of each model were assembled and tuned within a narrow selection of matching irises. Figure 9 is a photograph of the production flight model doublers.

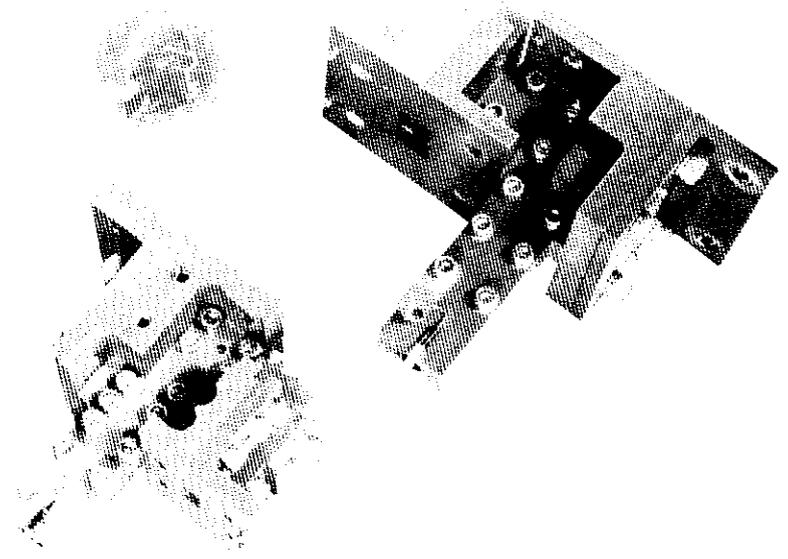


Figure 9. 9.5- to 19-GHz and 14- to 28-GHz Doubler Aluminum Production Flight Models

Performance measurements

Typical flight model transfer characteristics and return loss measurements are shown in Figures 10a through 10f. The improvement in circuit performance achieved through iris tuning for the flight models is shown in Table 3.

Changing the doubler construction materials from copper to aluminum created a potential increase in conversion loss for several reasons. The iridized aluminum contact surfaces required high contact pressure for low contact loss. Most importantly, the increased resistivity and contact loss changed the impedance matching requirements. As iris matching planes

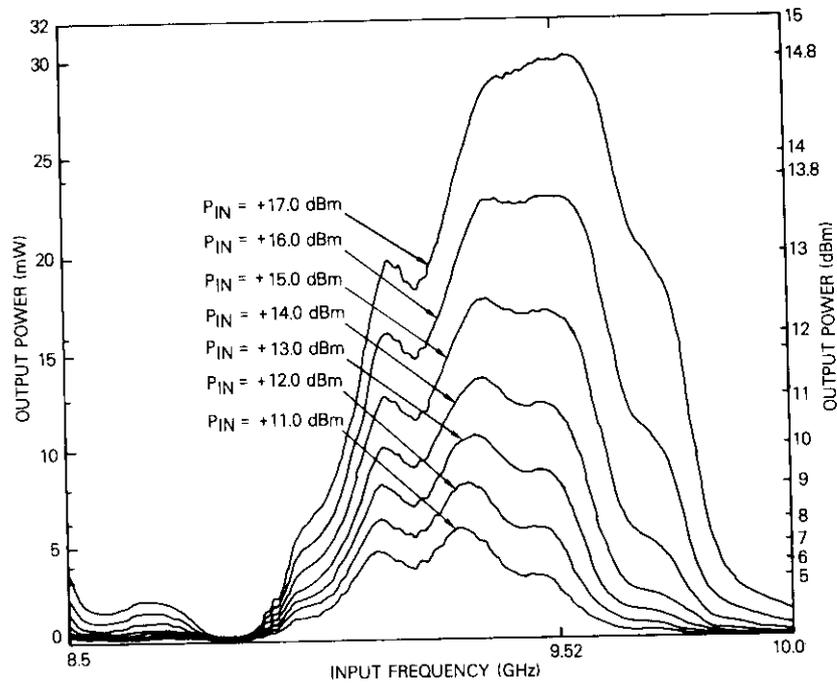


Figure 10a. Swept Frequency Output Power for the Beacon 19-GHz Aluminum Flight Model Doubler (SN01)

were established only for the copper engineering models, slight impedance mismatches in the aluminum flight models were possible. Finally, greater resistivity increased waveguide loss. These factors reduced the bandwidth and efficiency of the 28-GHz doubler aluminum flight model.

The 19-GHz doubler flight model was essentially unaffected by the materials change. Only a slight matching degradation was noted, probably due to the reduced contact loss of iridized surfaces surrounding the diode at 19 GHz.

Aluminum construction reduced the doubler weight by 70 percent relative to copper construction at a maximum loss of 19 percent (see Tables 2 and 3) in microwave power conversion efficiency. The weights of the 28-GHz and 19-GHz doubler flight models are 74 and 94 g, respectively.

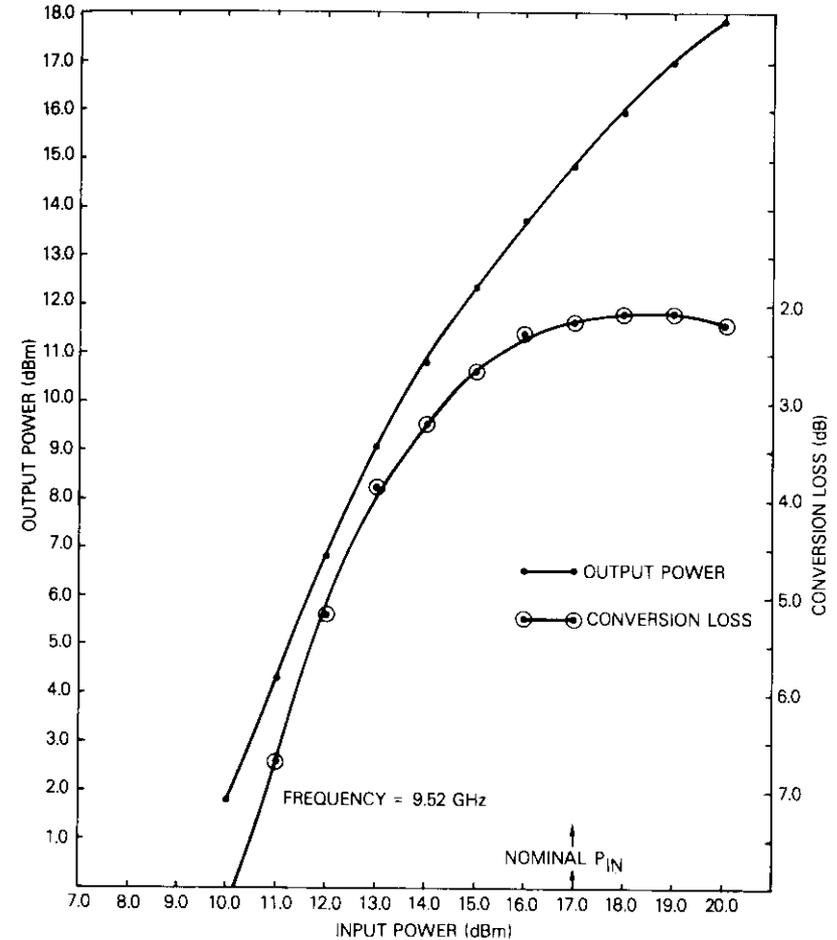


Figure 10b. Power Transfer Characteristic for the Beacon 19-GHz Aluminum Flight Model Doubler (SN01)

Assembly techniques

Certain assembly techniques ensure successful operation of the doubler flight models. Precision soldering of the bias-line low-pass filter to the main center conductor using a split-block alignment jig establishes the correct line length of the filter's first section. The press-fit of the input

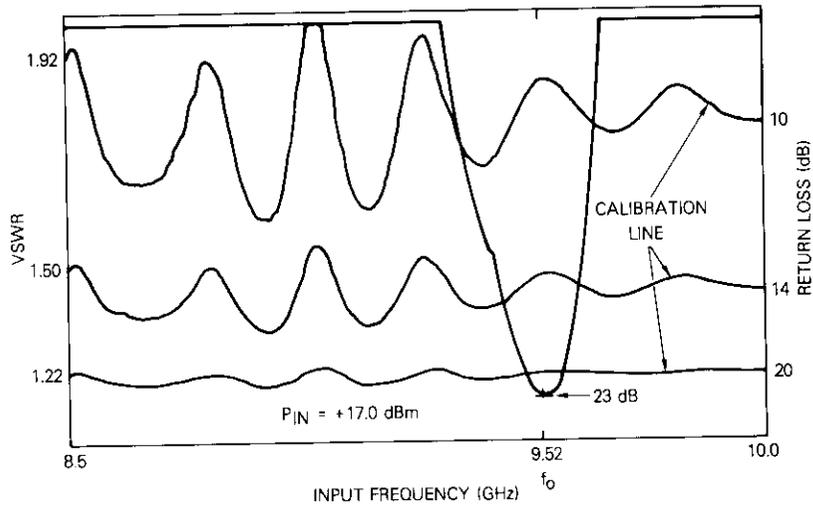


Figure 10c. Swept Frequency Return Loss for the Beacon 19-GHz Aluminum Flight Model Doubler (SN01)

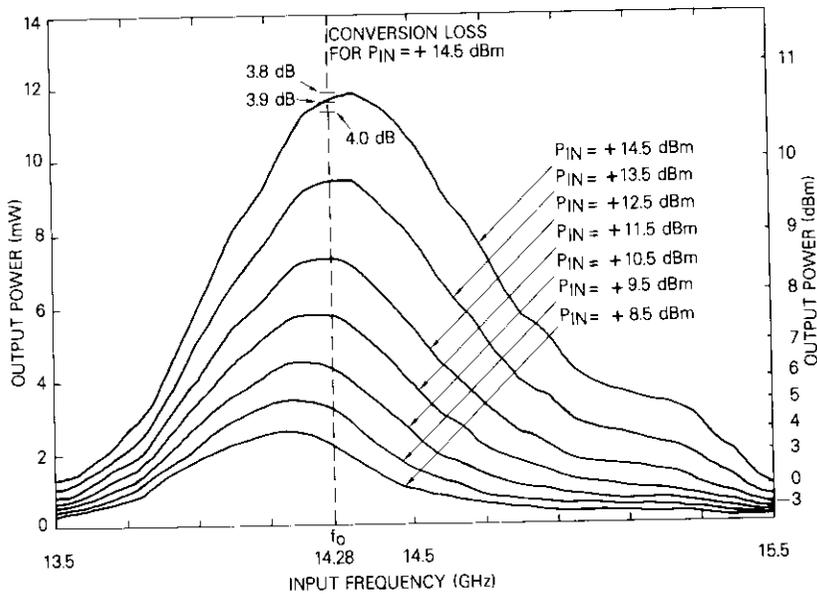


Figure 10d. Swept Frequency Power Output for the Beacon 28-GHz Aluminum Flight Model Doubler (SN03)

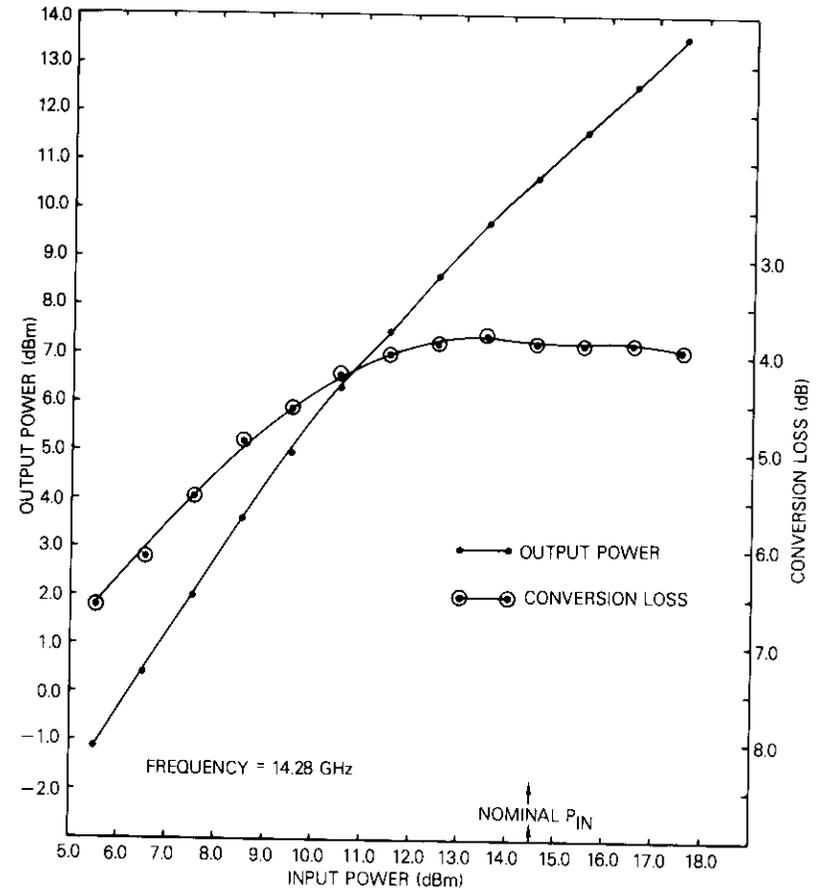


Figure 10e. Power Transfer Characteristic for the Beacon 28-GHz Aluminum Flight Model Doubler (SN03)

low-pass filter's center conductor section and dielectric into the surrounding coaxial ground shield is achieved using a preformed metal guide and tool. Most of the doubler's dimensions in this region are toleranced to one-half of one mil (± 0.0005 in.) and dimensioned for a line-to-line fit. The press-fit is essential to the proper transfer characteristics of the input low-pass filter. All teflon dielectric sleeves are cross-pinned to the doubler body. The nylon pins are epoxied into recesses in the split-block body to prevent movement of the center conductors. The filter termination re-

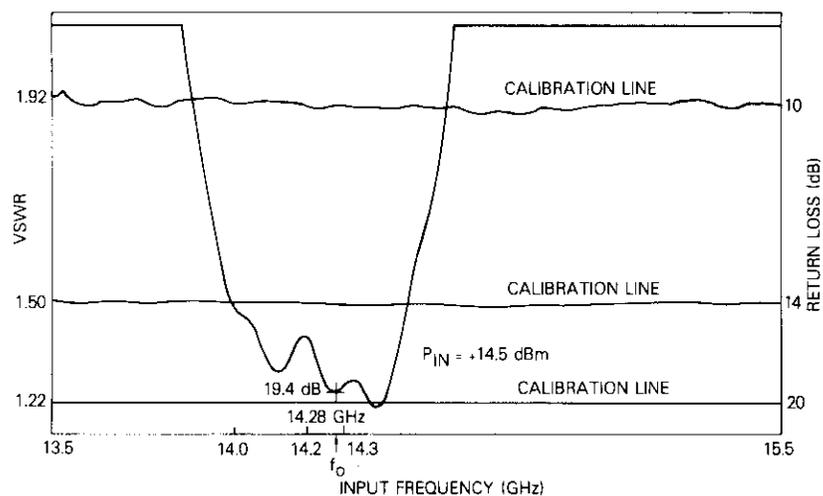


Figure 10f. Swept Frequency Return Loss for the Beacon 28-GHz Aluminum Flight Model Doubler (SN03)

TABLE 3. DOUBLER PERFORMANCE TABLE FOR ALUMINUM FLIGHT MODELS

Parameter	19-GHz Doubler Performance ($P_{IN} = +17.0$ dBm)		28-GHz Doubler Performance ($P_{IN} = +14.5$ dBm)	
	Measured		Measured	
	Initial	Final	Initial	Final
Conversion Loss Average (dB)	2.7	2.2	6.0	4.3
Return Loss Average (dB)	—	22	—	21
1-dB Output Bandwidth Average (MHz)	845	640	1,800	1,000
Efficiency Average, η (%)	54	60	25	37

sistor is low-temperature soldered to the feedthrough filter using pin-vises concentrically mounted in a jig.

Finally, the diode is silver-epoxied to the input low-pass filter using a recess in the output waveguide lid as an alignment tool. The spring-loaded

fuzz-button contact in the lid recess is then impregnated with silver-epoxy and installed permanently onto the diode. This diode attachment technique provides mechanically flexible mounting, as well as a low-loss microwave contact in this high-current-density region.

Doubler environmental tests

All units operated successfully in a 7-G vibration environment and a thermal vacuum environment over a temperature range of -45°C to $+40^{\circ}\text{C}$ in a hard vacuum. The adaptability of the designs to production lots of space-qualified hardware is excellent.

Conclusion

Two space-qualified models of centimeter wave varactor doublers having state-of-art electrical performance were built in a small production lot. The basic circuits were designed using computer-aided analysis to produce working hardware with good but less than optimum performance. Cold test measurements were used to calculate simple matching irises that improve the circuit match and hence achieve optimum performance. The iris matching technique was used to optimize a production run of five doublers of each type which were subsequently space qualified.

Acknowledgments

The author wishes to express his gratitude to R. Sicotte for his extensive design contributions to both doubler models. Special thanks are due to Messrs. J. Molz, J. Jerome, and G. Hawisher for their assistance with the assembly and measurements and to Mr. William Geisinger for his technical guidance throughout the project.

References

- [1] L. Pollack, "Centimeter Wave Beacons for the COMSTAR Satellites," *COMSAT Technical Review*, Vol. 7, No. 1, Spring 1977, pp. 103-108.
- [2] L. Kurokawa, "On the Use of Passive Circuit Measurements for the Adjustments of Variable Capacitance Amplifiers," *Bell System Technical Journal*, January 1962, pp. 361-381.
- [3] C. B. Swan, "Design and Evaluation of a Microwave Varactor Tripler," *ISSCC Digest of Technical Papers*, 1965, pp. 106-107.

- [4] D. H. Steinbrecher, M. E. Goff, and A. H. Soloman, "Iterative Synthesis of Varactor-Multiplier Microwave Networks and a Doubler with 0.17 Watt Output at 47 GHz," *G-MTT International Symposium Digest*, 1967, pp. 157-159.
- [5] "Standard Definitions, Symbols and Methods of Test for Semiconductor Tunnel (Esaki) Diodes and Backward Diodes," IEEE Standard No. 253, December 1963.
- [6] W. J. Getsinger, "The Packaged and Mounted Diode as a Microwave Circuit," *IEEE Transactions on Microwave Theory and Techniques*, MTT-14, No. 2, February 1966, pp. 58-69.
- [7] C. B. Burckhardt, "Analysis of Varactor Frequency Multipliers for Arbitrary Capacitance Variation and Drive Level," *Bell System Technical Journal*, Vol. 44, 1965, p. 675.
- [8] D. H. Steinbrecher, "Harmonic Multiplication with Punch-Through Varactors," *ISSCC Digest of Technical Papers*, 1966, p. 68.
- [9] D. Der Young Tang, "General Analysis of Idlerless Frequency Multipliers," Ph.D. Thesis, Northeastern University, Boston, Mass., February 1969.
- [10] H. A. Watson, *Microwave Semiconductor Devices and Their Circuit Applications*, New York: McGraw-Hill, 1969, pp. 245-248.
- [11] R. C. Mott, and R. L. Sicotte, "Design Space Qualified Millimeter-Wave Doublers," *Microwaves*, July 1976, pp. 46-48.
- [12] A. E. Atia, Private Communication.



Richard C. Mott received B.S.E.E. and M.S.E.E. degrees from George Washington University in 1967 and 1973, respectively. After joining COMSAT Laboratories in 1966, he participated in the evaluation of components for the INTELSAT IV communications transponder and the design of microwave stepped-digit elliptic-function filters. He developed microwave integrated circuitry for the DOMSAT simulator and several different satellite-borne waveguide circuits for the centimeter-wave Beacon transmitters. Recently, he has

developed a 14 GHz microstrip tunnel diode amplifier for an INTELSAT V study. He is presently a technical staff member of the Microwave Systems Department of the Microwave Laboratory, where he is responsible for the development of microwave satellite components and systems. In addition, he is project manager for the development of the 12 GHz "R" switch.

Index: IMPATT diode, COMSTAR, centimeter wave, beacon, reliability

Reliable IMPATT diodes for the COMSTAR centimeter wave beacon

P. KOSKOS, P. FLEMING, AND J. REYNOLDS

(Manuscript received July 22, 1977)

Abstract

High-efficiency, high-reliability IMPATT diodes were required for the transmitters to be used in the COMSTAR satellite 19- and 28-GHz beacons. Silicon, single-drift diodes were used. Close cooperation with selected sources, coupled with rigorous specification, screening, and burn-in requirements, resulted in IMPATT diodes with a demonstrated projected life in the range of 10^7 to 10^{12} hours, well over the estimated mission life requirement of 100,000 hours. Beacons launched with COMSTAR DI and D II in May and July 1976 have operated until October 1977 with no IMPATT diode failures.

Introduction

IMPATT diodes, which are the active elements in the output amplifier stages of the 19- and 28-GHz transmitters [1], [2] of the COMSTAR satellite beacons, are the subject of this paper. IMPATT diodes were not a mature, high-reliability product at the end of 1973 when the Beacon program was initiated. While several high-reliability applications were in development [3]-[5], the desired reliability had not yet been achieved and the necessary test and screening criteria had not been established. Special effort was needed to obtain the required performance as well as to establish device design, screening and test procedures, and manufacturing controls to ensure reliability.

The system tradeoff considerations resulting in the original performance requirements summarized in Table 1 have been previously described [1], [2].

TABLE 1. BEACON IMPATT REQUIREMENTS

Frequency (GHz)	19.04	28.56
System Power, Min. (dBm)	+27	+28
IMPATT Stages	2	6
Power Out per Diode (dBm)	+27	+24
Efficiency* (%)	6-10	6-10
Maximum Junction Temperature (°C)	200	200
2-Year Mission Reliability per Diode (%)	99.5	99.5
Median Time to Failure per Diode Assuming $\sigma = 1$ (hr)	400,000	400,000

*Efficiency is defined as RF power out ÷ DC power in, operating in an oscillator mode.

The frequency and power output requirements imply a high power density in the IMPATT mesa, which generates a high semiconductor junction temperature tending to shorten life. The minimum efficiency requirement of 6 percent was set by the satellite power budget limitations; however, higher efficiencies would moderate the semiconductor junction operating temperature, hence improving life and reliability. The successful implementation of a high-reliability IMPATT diode procurement program for the beacon transmitters is described in the following sections of this paper.

Device physics of IMPATT diodes

Operational principles

The negative resistance in an IMPATT (*impact avalanche transit time*) device is a result of a 180° phase difference between the RF voltage and current in a reverse biased p-n junction. The physical mechanisms involved are avalanche breakdown and carrier drift under an applied electric field. The phase difference is produced by the delay inherent in the build-up of the avalanche current and in traversing the drift zone. Figure 1 shows the doping profile and approximate field distribution for a single-drift abrupt junction. The dimensions and approximate field strength are typical of a high-frequency device.

Beacon IMPATT diodes

The IMPATT diodes for the Beacon program were single-drift silicon devices procured from two sources in n^+np^+ and p^+pn^+ configurations.

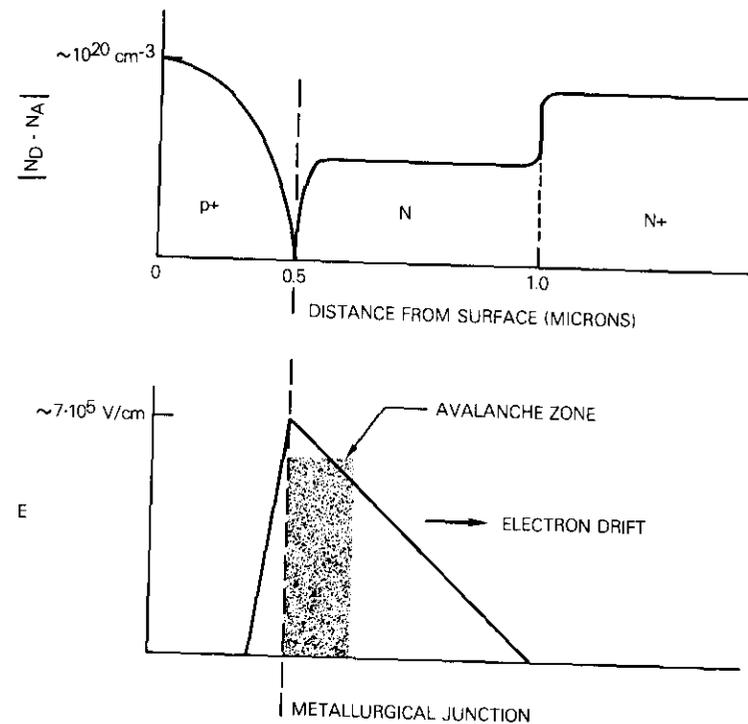
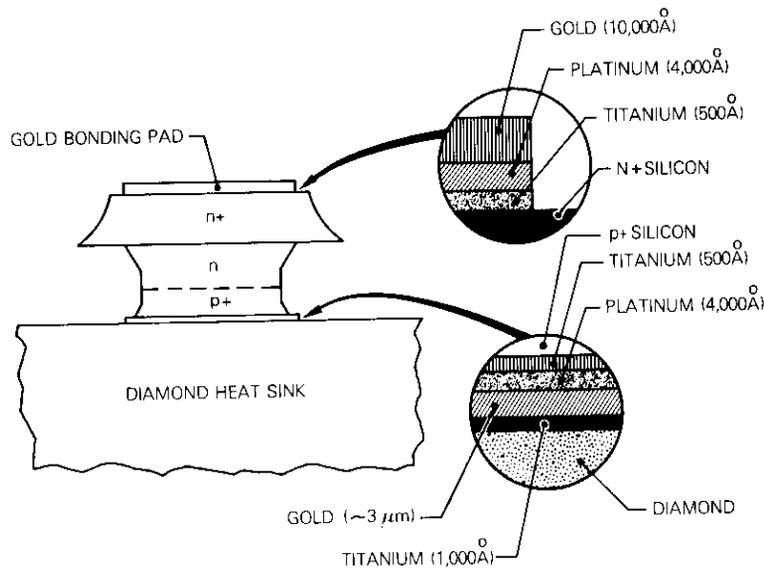


Figure 1. Doping and Field Profile vs Distance for a Single-Drift IMPATT Diode

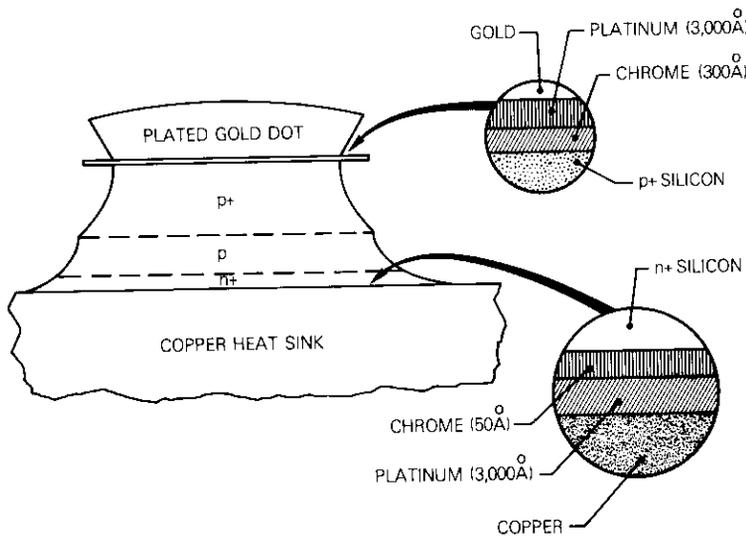
These configurations are complementary, but the p^+pn^+ device is characterized by higher efficiency [6]. Construction details are indicated in the device cross sections in Figure 2. In each case, platinum is the barrier metal which prevents diffusion of other metals into the silicon at the elevated device operating temperature. The complementary junction device has a plated copper heat sink while the other device is thermocompression bonded to a metalized diamond heat sink. The merits of these heat-sinking techniques will be discussed later.

Failure modes

The dominant IMPATT failure mode results in a metallic spike which short-circuits the junction either in the bulk or on the surface of the diode. Such failures, including "tuning-induced short-out" [7], which have been



a. Regular Device Section



b. Complementary Device Section

Figure 2. IMPATT Diode Cross Sections

attributed to the thermal runaway mechanism [8], [9], cause "shortout" of IMPATT diodes. The efficiency of single-drift IMPATT diodes is typically less than 10 percent; *i.e.*, more than 90 percent of the DC input power is dissipated as heat in the high field region (typically less than 1 micron thick) of the diode, causing the junction temperature to rise. As the junction temperature rises, the leakage current with reverse bias increases. The combined effects of leakage current, avalanche multiplication, and thermal resistance result in the formation of a current-controlled differential negative resistance in the DC voltage current characteristic. In bulk materials, such a characteristic leads to the formation of destructive high current filaments. The metallic spikes uncovered in subsequent failure analyses are replicas of these high current filaments. Figure 3 shows representative scanning electron microscope (SEM) images of IMPATT diodes subjected to failure analysis; these views are almost normal to the top of the diode, allowing a view of the mesa edge. The top contact has been removed to aid in the failure analysis. Figure 3a shows a spike largely containing gold (Au) in bulk silicon; Figure 3b shows a finger-like spike, largely containing Au, at the perimeter of the silicon mesa. Both failed diodes tested as DC shorts prior to failure analysis.

Heat sinking

Due to the temperature dependence of the runaway mechanism the thermal resistance should be as low as possible [9]. This results in lower junction operating temperatures and a higher probability of survival. A simple model for a silicon abrupt junction IMPATT is shown in Figure 4a for a single-sided heat sink [10]. The resultant thermal impedance is given by

$$\theta = \frac{L_S}{\pi r^2 K_S} + \frac{L_M}{\pi r^2 K_M} + \frac{1}{\pi r K_H} \quad (1)$$

- where θ = thermal impedance
- L_S = silicon thickness (junction to metal contact)
- L_M = metal thickness
- r = diode radius
- K_S = thermal conductivity of silicon
- K_M = thermal conductivity of metal
- K_H = thermal conductivity of heat sink.

The last term in equation (1), which represents the spreading resistance in the heat sink, usually dominates. For minimum spreading resistance,



Figure 3b. SEM Image of Failed Diode AA-35 with Spike Occurring at Edge of Silicon Mesa (SEM taken at COMSAT Laboratories)

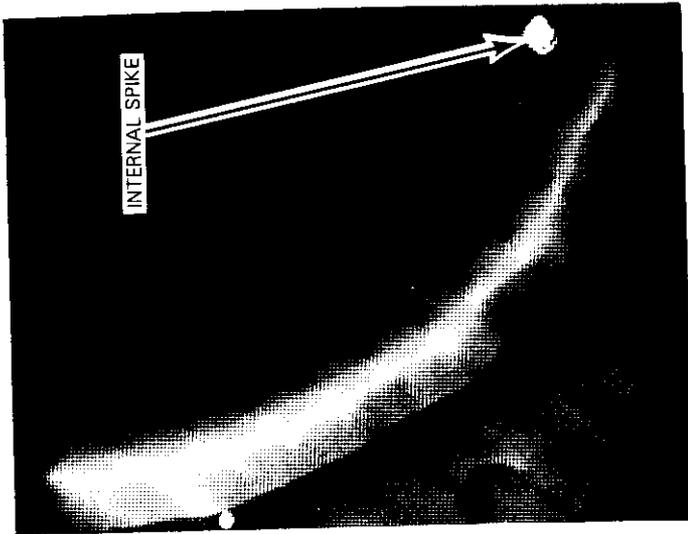


Figure 3a. SEM Image of Failed Diode AA-24 with Metallic Spike in Bulk Silicon (SEM taken at COMSAT Laboratories)

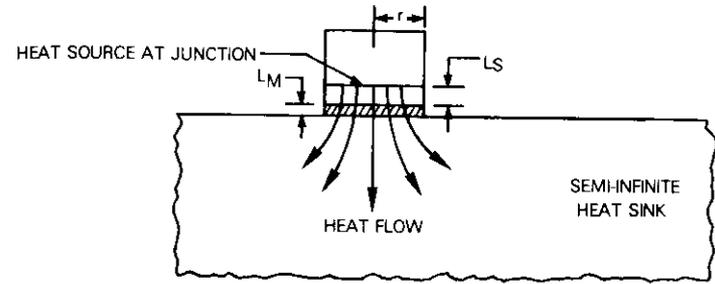


Figure 4a. Thermal Impedance Model

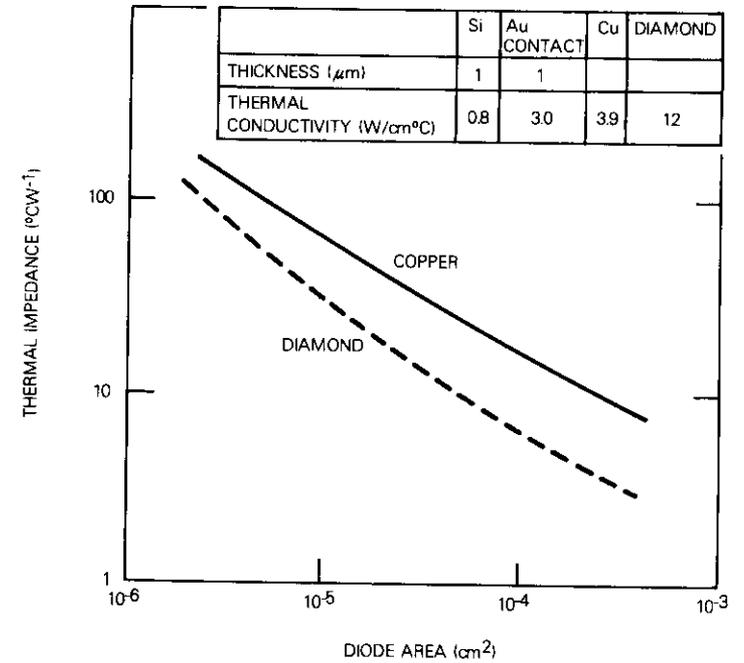


Figure 4b. Thermal Impedance vs Diode Area (reprinted from *Avalanche Diode Microwave Oscillators*, Clarendon Press: Oxford, 1973; courtesy of George Gibbons)

the ratio of heat sink radius to diode radius should be 5:1 or larger [11]. Figure 4b compares the thermal resistance of copper and of type II diamond heat sinks. The lower thermal impedance of the diamond heat sink provides a higher probability of thermal stability [8]. A given diode may have local variations in thermal impedance which could enhance formation of the destinate high current filaments.

This analysis was applied to the two types of IMPATTs procured for this program. The first type, which used a plated copper heat sink, marginally met the thermal resistance requirement and exhibited the highest early applications failure rate. The second type, which used a diamond heat sink, was very low in thermal resistance and had no unexplained applications failures. Although factors other than thermal resistance contribute to the early unexplained failures, the correlation between thermal resistance and failure rate must be considered.

Implementation

Definition of product

At the beginning of the Beacon program, IMPATT diodes from a number of sources were evaluated for electrical performance in amplifier circuits and for potential reliability. Control drawings were then prepared, covering the major details of fabrication processing, screening, and testing. Characterization tests in the control drawings included breakdown voltage, capacitance, and efficiency to ensure performance in the beacon amplifier and to detect anomalies in the diode fabrication. A typical characterization table is shown in Table 2.

Records were required for reject rates at several manufacturing points prior to final test. Lot reject criteria were established so that an entire diode lot could be rejected when high reject rates indicated an uncontrolled manufacturing process.

Product control

Periodic on-site source surveillance was coupled with frequent source progress reports on reject rates and other processing problems. The value of this process became evident when COMSAT Laboratories' tests of development samples revealed a defective metallization procedure sufficiently early to permit correction without a major program delay. Specifically, the barrier metal chosen initially did not adequately prevent gold-silicon or copper-silicon interdiffusion and was subsequently changed to platinum.

TABLE 2. ELECTRICAL CHARACTERISTICS
(nominal ambient temperature = 25°C)

Characteristic	Conditions	19-GHz Flight	
		Min	Max
Breakdown Voltage, VBR (V)	IR = 1 mA	35	45
Leakage Current, ILK (nA)	VR = 30 V	—	10
Junction Capacitance, ^a C _J (pF)	F = 1 MHz BIAS = VBR - 1V	0.3	0.7
Junction to Case Temperature, ^b DTJ (°C)	F = 19.0 ± 0.4 GHz PO = 500 mW	—	160
Efficiency, ^b η (%)	F = 19.0 ± 0.4 GHz PO = 500 mW	6.0	—
Output Power, ^b PO (mW)	F = 19.0 ± 0.4 GHz	500	—
Operating Voltage, ^b VOP (V)	F = 19.0 ± 0.4 GHz PO = 500 mW	45	55
Operating Current, ^b IOP (mA)	Same as VOP	130	220
Forward Voltage, VF (V)	IF = 300 mA	0.9 ^c	1.1 ^c

^a Capacitance will be measured at 10 equally spaced voltages between 0V and VBR, once for each diode (after burn-in). For repeated QA and REL tests, only the BIAS point specified above needs to be measured.

^b Measured in a specified oscillator circuit.

^c A curve tracer I-V presentation will show a smooth curve with no irregularities.

All products received at the Labs incoming inspection were periodically retested to detect any significant changes with shelf life during the Beacon construction program.

Screening and user application results

Less than half of the IMPATT chips entering the diode fabrication cycle survived screening and testing. Table 3 is a breakdown of the failure rates

TABLE 3. SCREENING AND TESTING YIELDS

Test Point	Source A				Source B			
	Total (Units)	Pass (Units)	Point Yield	Cum. Yield	Total (Units)	Pass (Units)	Point Yield	Cum. Yield
Assembly	300	193	0.64	0.64	347	203	0.59	0.59
First Test	193	181	0.86	0.60	203	173	0.85	0.50
Screening	181	137	0.76	0.46	173	108	0.62	0.31
Burn In	137	122	0.89	0.41	108	83	0.77	0.24

at key points for the two sources. Although the yields are different for the two sources, the rough correlation between them suggests broad similarities in the distribution of design and process flaws detectable by the screening process.

In subsequent inspection and end application the testing, results diverged markedly. For source A, few unexpected failures (<5 percent) were discovered. Failure analysis attributed these to external causes such as accidental application of excessive power. For source B, there was a further loss of over 50 percent during incoming inspection and system acceptance testing. The failure analyses, based on test history and scanning electron microscope studies, indicated that the failures were associated with leakage resistance and that the original screening criteria were not sufficiently rigorous. All of the flight stock was then rescreened to new leakage specifications which included a lower leakage current limit, and a leakage current stability with time requirements. The surviving product has functioned without failure to October 1977.

It is possible that the low overall screening yield for source B and the high failure rate in inspection and end use were related by design/fabrication deficiencies which could have been eliminated at the start by more rigorous screening. However, because of the time limitations of the Beacon program, production was completed before failure analyses identified the cause of the problem and before corrective action could be implemented at the fabrication level.

Reliability demonstration

The control drawing specified reliability demonstration tests on which to base estimates of product reliability and median time to failure within three months of fabrication completion. Stepped stress and accelerated (temperature) life tests were designated, and the results are discussed in the following section.

Reliability Results

Background of reliability analysis techniques

The failure statistics for IMPATT diodes can be approximated by a log-normal distribution which has been discussed in detail by Peck and Zierdt [12], and by F. H. Reynolds [13]. The cumulative failure probability, Q , is given by

$$Q = \Phi[z(t)] \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{\mu^2}{2}\right) d\mu \quad (2)$$

where
$$z(t) = \frac{1}{\sigma} \ln \frac{t}{t_m}$$

and Φ = the cumulative normal probability function
 z = the standard normal deviate
 t_m = the median time to failure
 σ = the dispersion.

The parameters t_m and σ characterize the distribution and are usually determined experimentally.

Under conditions of constant stress (constant temperature), t_m is constant and z is proportional to the logarithm of time. For IMPATT diodes, t_m has been shown to be an exponential function of the temperature, i.e.,

$$t_m = t_o \exp\left(\frac{qE}{kT}\right) \quad (3)$$

where t_o = a constant
 E = the activation energy
 q = the electronic charge
 k = Boltzmann's constant
 T = absolute temperature.

A life test can be designed in several ways. Equations (2) and (3) show that the cumulative failure probability, Q , is a function of time and temperature, as shown graphically in Figure 5. Time is plotted along the X-axis on a logarithmic scale, and temperature is plotted along the Y-axis scaled to the reciprocal of absolute temperature. Each curve, which is computer plotted from equations (2) and (3) assuming an activation energy of 1.7 eV, represents the failure probability as a function of time at a fixed temperature. The parallel straight lines are contours of equal failure probability. The projection of the 50-percent contour on the time-temperature plane is the Arrhenius relationship defined by equation (3).

One straightforward approach suggested by Figure 5 is the "constant stress" test or isothermal anneal. In this case devices are tested at a fixed junction temperature, and the cumulative distribution function, Q , is measured as a function of time. The data are then applied to equation (2) to determine t_m and σ . The analysis can be simplified graphically by plot-

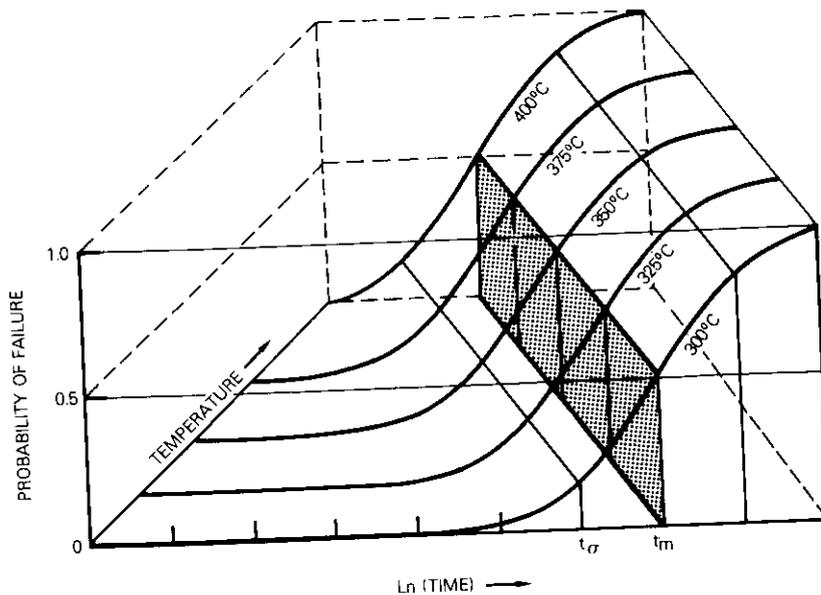


Figure 5. Time, Temperature, and Failure Probability Interrelationship

ting $\ln(t)$ vs Q on probability paper. For a log-normal distribution, a straight line is obtained.

The relationship between t_m and T can be determined by repeating the isothermal anneal at two or more temperatures, and equation (3) can be used to calculate activation energy and t_0 . Equation (3) can also be used to calculate a predicted t_m for any desired operating temperature.

An alternate approach is the "stepped stress" test or isochronal anneal during which the sample is stressed for fixed time intervals at a series of progressively higher temperatures. Under these conditions, the cumulative failure probability, Q , is a normally distributed function of the reciprocal of the absolute temperature. Giving t in equation (2) some arbitrary value ($t \equiv t^*$) and substituting equation (3) into (2) yields

$$z\left(t^*, \frac{1}{T}\right) = \frac{qE}{k\sigma} \left(\frac{1}{T^*} - \frac{1}{T}\right) \quad (4)$$

where T^* is the absolute temperature at which $t_m = t^*$ and all other parameters are previously defined.

The parameters T^* and $qE/k\sigma$ can be determined by applying the ex-

perimentally observed failure distribution to equations (2) and (4). Again, the procedure can be simplified by plotting Q vs $1/T$ on probability paper, where it should appear as a straight line. When stepped stress data are analyzed, an adjustment must be made for the time the sample has been tested at the lower stress levels, as discussed by Peck and Zierdt [12]. For properly chosen stress levels, the adjustment is small.

Both of the foregoing methods measure median time to failure at various elevated temperatures, which can be used with equation (3) to predict t_m for normal operating temperatures. As shown in the next section, normal temperature t_m 's are greater than 10^6 hours so that real-time verification measurements are impractical.

Life test design and analysis

Tests were conducted to establish a prediction criterion for the mission life reliability of the flight quality IMPATT diodes. Since information was needed well before the launch, a 3-month accelerated life test was employed. Life test data on similar devices were already available [5], [14]. These data established that, for a properly screened population characterized by the alloy spike failure mode, the failure statistics could be modeled by the log-normal distribution. They also showed that the t_m vs T relationship for IMPATT diodes followed the Arrhenius relationship, equation (3), and provided a good estimate of the activation energy.

With this background a life test program was designed to produce meaningful results within a limited time using a small number of diodes. This program consisted of a stepped stress test followed by an accelerated life test at a single stress level. The stepped stress test, which required only a few weeks to complete, established the appropriate stress level for the constant stress test. A description of the procedure illustrated by an example follows.

Stepped stress test design and results

Ten diodes were selected from each population of flight diodes and subjected to 40 hours of operation at progressively higher temperatures, starting at 325°C and increasing in 25°C steps until all of the diodes had failed. The thermal stress was applied by increasing the voltage and current to produce DC dissipation in the IMPATT junction in a non-oscillating condition.

Figure 6 shows stepped stress test results for the 19-GHz IMPATT diode lot. The junction temperature, in degrees Celsius, is plotted on probability paper as a function of the cumulative probability of failure. Both the

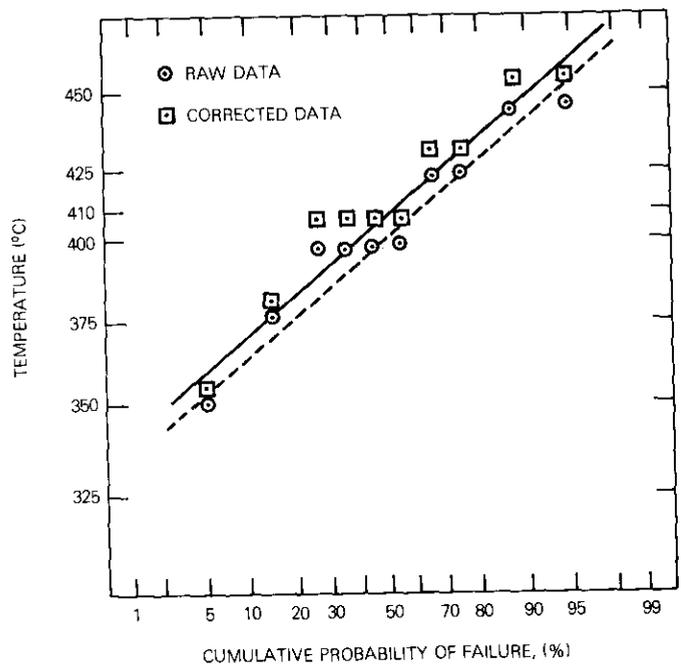


Figure 6. 19-GHz IMPATT Stepped Stress Test Results

original data and the data corrected using the procedure described by Peck and Zierdt [12] are plotted. The circles represent the junction temperature at each step, and the squares represent corresponding temperatures corrected for the effect of the preceding steps [12]. The failure distribution in Figure 6 is log-normal in that the data can be fitted by a straight line. (This result also implies that there is one dominant failure mechanism and one activation energy over the temperature-time range of the test, and that the screening program has successfully eliminated random early failures.) The 40-hour 50-percent failure point for the corrected data occurs at 410°C.

Accelerated life test design and results

The temperature for an accelerated life test may now be specified. The choice is dictated by the following considerations. The temperature should be high enough to ensure failure of a substantial fraction of the diodes (approximately 75 percent) during the 2,160-hour (3-month) planned test

period. This is necessary to more accurately establish the dispersion, σ . However, the planned median life should be substantially longer than the stepped stress lifetime (40 hours) so that the operating temperature may be more accurately extrapolated.

To estimate the appropriate temperature, the following procedure was adopted. Figure 7 presents stepped stress and accelerated life test data for three lots of beacon IMPATT diodes, as well as two earlier sets of data [5],

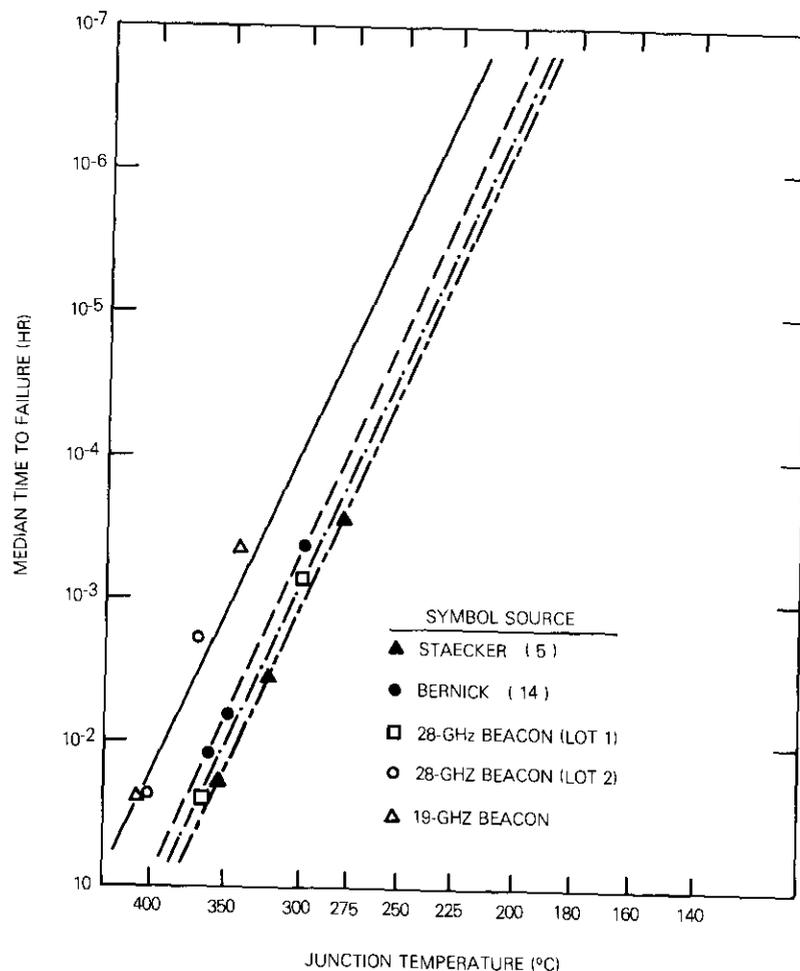


Figure 7. IMPATT Diode Life Test Results

[14]. Again, time in hours is plotted logarithmically; junction temperature is scaled to the reciprocal of absolute temperature labeled in °C for convenience. The open symbols represent beacon IMPATT data, and the solid symbols represent the data referenced earlier. The stepped stress data for the three beacon lots are the three lowermost open symbols. An Arrhenius curve was drawn through each of these symbols using an assumed activation energy of 1.7 eV, based on the data represented by the solid symbols. From these constructions, a temperature corresponding to an accelerated lifetime of 1,200 hours was estimated graphically and by calculation, and an accelerated life test conducted at that temperature.

The analysis of one of the constant stress tests (beacon 19-GHz lot) is shown in Figure 8. Twelve samples were stressed at a junction temperature of 340°C. The cumulative probability of failure vs the logarithm of time is plotted on probability paper. All failures were short circuits; through numerous failure analyses, they were traced to metal spikes through the semiconductor junction. The data fit a log-normal distribution. The observed median time to failure is 2,200 hours, and the dispersion is 1.6. This result is plotted in Figure 7 along with the corresponding test results for the two 28-GHz beacon IMPATT lots. The three constant-stress test results correspond closely to the predicted Arrhenius curves. The activation energies shown in Table 4 for the beacon lots are calculated

TABLE 4. RELIABILITY TEST SUMMARY

Lot	Accelerated Life Temp (°C)	Dispersion (σ)	Activation Energy (eV)	Application Junction Temp (°C)	Predicted t_m (hr)	Required t_m (hr)
19 GHz	340	1.6	2.1	165	10^8	2.7×10^6
28 GHz I	300	1.0	1.8	185	10^7	4×10^5
28 GHz II	370	1.0	2.0	135	10^{12}	4×10^5

from the two test points shown in Figure 7. For beacon reliability predictions, the original estimates using the activation energy of ~ 1.7 eV reported by Staecker [5] and Bernick [14] were used. This activation energy was judged more reliable because it was based on larger sample sizes and more data points. It also provided a more conservative prediction of mission reliability.

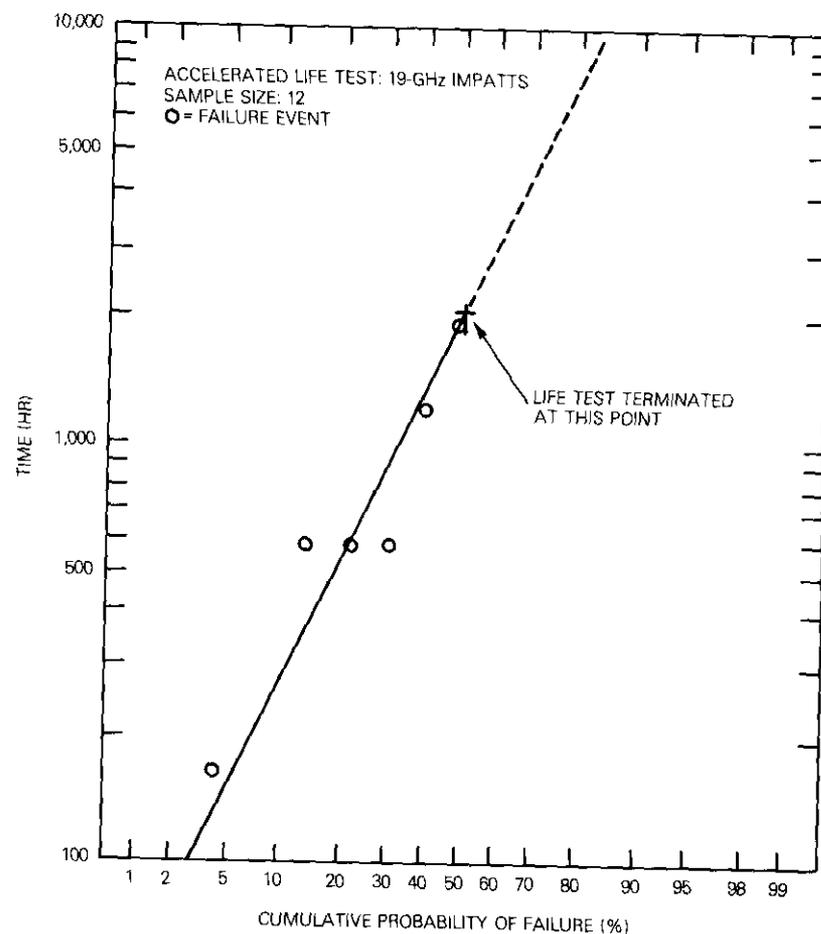


Figure 8. Accelerated Life Test Analysis

Reliability estimate

The above data must be reduced to a reliability estimate for the diodes to be compared with the mission requirements. As defined by the system reliability model, the IMPATT diode component reliability must exceed 0.995 for the 2-year duration of the mission.

Given these requirements, the corresponding minimum median time to

failure can be calculated for a log-normal distribution. The cumulative probability of failure (per diode) must not exceed 0.005, and the corresponding standard normal deviate must not exceed -2.57 . In addition, statistical error must be included in the estimate of t_m . The Student's t distribution for the sample size used in the beacon life tests indicates a 5-percent chance that the true value will lie more than 0.58 standard normal deviate below the estimate. Adding the two deviates makes it possible to predict with a 95-percent confidence that, if $z < -3.15$ [see equation (2)], the device will accomplish the required mission. When $t = 2$ years (17,520 hours), the median time to failure must exceed 17,520 $\exp(3.15\sigma)$ hours. Thus, when $\sigma = 1.6$ (Figure 8), t_m must exceed 2.7×10^6 hours.

The value of the median time to failure at the operating junction temperature is estimated by extrapolating the Arrhenius curves. The value of σ is assumed to be constant, which is supported by Staecker's results [5] showing that σ is essentially constant over a broad temperature range. The results are summarized in Table 4, which shows the values for σ , activation energy, operating junction temperature, and the estimated t_m and the required t_m based on the beacon's reliability requirements.

Conclusions

Reliable IMPATT diodes were achieved by an integrated effort involving the following:

- a. mature fabrication technology;
- b. an understanding of device physics and failure mechanisms by both manufacturer and user and close cooperation between them;
- c. control specifications covering major fabrication steps from material procurement through processing, screening, and testing;
- d. non-destructive screening and testing procedures to identify latent defects;
- e. proper circuit application.

Careful design of the accelerated life tests and thorough analysis of the results of the reliability tests for consistency demonstrate the product reliability for prediction purposes. For the Beacon program, IMPATT diode life predictions were in excess of 10^7 hours, which was well over the Beacon mission requirements. The predictions were obtained using short duration accelerated life tests which can provide timely and useful reliability esti-

mates for semiconductor performance over the 7-year life currently planned for communications satellites.

Acknowledgments

The support of Messrs. L. Pollack, R. Briskman, R. Strauss, and W. Getsinger in completing this program is acknowledged. Numerous other individuals at COMSAT Laboratories contributed to the program implementation. Thanks are due to Messrs. H. Carlson, A. Ewing, I. Feigenbaum, and J. Francis. T. Kirkendall and J. Hannsen provided the SEM images and valuable analyses of diodes failures.

References

- [1] W. J. Getsinger, "Centimeter Wave Beacon Transmitter Design," *COMSAT Technical Review*, Vol. 7, No. 1, Spring 1977, pp. 109-128.
- [2] M. J. Barrett, "19- and 28-GHz IMPATT Amplifiers," *COMSAT Technical Review*, Vol. 7, No. 1, Spring 1977, pp. 129-152.
- [3] R. L. Kuvas and J. A. Rupp, "Design, Characterization and Reliability of 6-GHz Silicon IMPATT Diodes," International Electron Devices Meeting, *Proc.*, 1973.
- [4] H. Nagao and S. Katayama, "Silicon IMPATT Diode Device Incorporating a Diamond Heat Sink," *NEC Research and Development*, N35, October 1974, pp. 67-76.
- [5] P. Staecker, W. T. Lindley, R. A. Murphy, and J. P. Donnelly, "Reliability of Silicon and Gallium Arsenide K_A -Band IMPATT Diodes," *Reliability Physics Proceedings*, 1974.
- [6] C. M. Lee et al., "A Comparison Between $n^+ - p - p^+$ and $p^+ - n - n^+$ Silicon IMPATT Diodes," *IEEE Transactions on Electron Devices*, ED-21, No. 2, February 1974, pp. 137-141.
- [7] C. A. Brackett, "The Elimination of Tuning-Induced Burnout and Bias-Current Oscillations in IMPATT Oscillators," *Bell System Technical Journal*, Vol. 52, March 1973, pp. 271-306.
- [8] H. M. Olson, "Thermal Runaway of IMPATT Diode," *IEEE Transactions on Electron Devices*, ED-22, April 1975, pp. 165-168.
- [9] H. M. Olson, "A Mechanism for Catastrophic Failure of Avalanche Diodes," *IEEE Transactions on Electron Devices*, ED-22, No. 10, October 1975, pp. 842-849.
- [10] G. Gibbons, *Avalanche Diode Microwave Oscillators*, Oxford, England: Clarendon Press, 1973, pp. 78-80.
- [11] G. Gibbons, Private Communication.

- [12] D. S. Peck and C. H. Zierdt, Jr., "The Reliability of Semiconductor Devices in the Bell System," *Proc. IEEE*, Vol. 62, No. 1, February 1974, pp. 185-211.
- [13] F. H. Reynolds, "Thermally Accelerated Aging of Semiconductor Components," *Proc. IEEE*, Vol. 62, No. 2, February 1974, pp. 212-222.
- [14] R. L. Bernick, Private Communication.



Paul Koskos received a B.S.E.E. from Columbia University and an M.S.E.E. from New York University; he has continued his postgraduate studies at George Washington University. Prior to joining COMSAT, he taught briefly at Brooklyn Polytechnic University. He was responsible for product engineering, design, and development of vacuum tubes and semiconductor devices at Western Electric and RCA. At RCA, his primary areas of responsibility were in traveling wave tubes and parametric amplifiers.

After joining COMSAT, he worked on the design and reliability of low-noise receivers in the Microwave Laboratory. He became Assistant Director, Reliability and Quality Assurance, in 1973. He is a senior member of IEEE and holds five patents on electronic devices.

Paul Fleming is Manager of the Device Physics Department at COMSAT Laboratories. He received a B.S.E.E. degree from the City College of New York in 1957 and an M.S.E.E. degree from Columbia University in 1960. He previously served as a VHF radio specialist with the Marines in South Korea. He spent ten years with the IBM Corporation, where he received an Invention Achievement Award. His work has included microwave memories, microwave modulation of light, digital microwave, 12-GHz propagation, bulk effect devices, and more recently, state-of-the-art devices in GaAs. He is a member of Eta Kappa Nu.



John Reynolds received a B.S. degree in physics from the Worcester Polytechnic Institute in 1962 and an M.S. degree in physics from Purdue University in 1964. Prior to joining COMSAT, he was a staff member at the Sprague Electric Research Center in North Adams, Mass., where he was engaged in research on MOS transistors and other solid-state devices.

Mr. Reynolds joined COMSAT in 1968 and has been engaged in the study and analysis of solar cells and more recently solid-state microwave devices. He is a member of IEEE and Sigma Xi.

Index: COMSTAR, beacon signal measurement, attenuation, data

Preliminary results of COMSTAR 19/29-GHz beacon measurements at Clarksburg, Maryland

J. M. HARRIS AND G. HYDE

(Manuscript received August 11, 1977)

Abstract

This paper reviews the characteristics of the COMSTAR beacon signals and briefly discusses the beacon receive facility at COMSAT Labs, Clarksburg, Maryland. Attenuation data covering individual events and 6-month statistics are presented, including cumulative fade distributions at 19.04 and 28.56 GHz, cumulative point rain rate distributions, diurnal distribution of fades greater than 5 dB at 19.04 and 28.56 GHz, and analysis of the relationship of 19.04-GHz and 28.56-GHz fade data. In addition, measurement results of orthogonal linearly polarized 19.04-GHz signals are described, including depolarization with and without significant attenuation, and differential attenuation of the horizontally and vertically polarized signals. Finally, phase measurements between the 28.56-GHz carrier and the upper and lower sidebands (± 264.4 MHz) are briefly discussed.

Introduction

The COMSTAR beacons [1] comprise the first stable and continuous geostationary orbit signal sources available for long-term measurements of propagation parameters of the slant path from earth to geostationary satellite in the satellite communications frequency bands near 19 and 29 GHz. Although one year is probably the shortest period from which

statistical data for the design of satellite communications systems can be determined, data from shorter periods are useful in analyzing the propagation characteristics at a given site.

This paper reports results of statistical analysis of attenuation data at 19.04 and 28.56 GHz measured using signals from the COMSTAR D1 beacons during the period of approximately six months from early July 1976 to early January 1977. Depolarization events with and without appreciable attenuation are presented for the 19.04-GHz signals. Analysis results for several events are given to establish the differential attenuation for the 19.04-GHz vertically and horizontally copolarized signals. In addition, diurnal distributions developed from the data for selected fade levels are presented graphically, along with instantaneous and statistical attenuation correlations. Finally, events which characterize the frequency dependent properties of the medium are discussed. The measurements program is expected to extend over several years, and will include measurements on the COMSTAR DII beacons as well.

Measurement facilities

Because all signals transmitted from the beacon derive by multiplication and amplification from a common 132.2-MHz source, both amplitude and "relative phase" measurements [2] are possible. These transmissions consist of vertically and horizontally polarized signals. The vertically and horizontally polarized transmit signals at 19.04 GHz (19 TV and 19 TH, respectively) derive from a single chain switched at a 1-kHz rate between the vertical and horizontal ports of a CONUS coverage horn. The vertically polarized signals at 28.56 GHz (carrier), 28.29 GHz (lower sideband) and 28.82 GHz (upper sideband) denoted as 29 C, 29 LSB, and 29 USB, respectively, are derived by phase modulating the carrier with the doubled 132.2-GHz source and fed into another CONUS coverage horn. Nominal and measured e.i.r.p. levels are given in Table 1.

The facilities at COMSAT Labs used to measure the COMSTAR Beacon signals consist of a 5-meter antenna system and COMSTAR Beacon receiver, a 12-GHz radiometric terminal operating independently through a 2-meter antenna, and a tipping bucket rain gauge. All data are recorded on chart paper for future data reduction and analysis.

An existing 5-meter Cassegrain antenna was modified by replacing the feed and orthomode transducer (OMT) to provide at least 40 dB of cross-polarization discrimination (XPD) inside the 2-dB contour of the 19.04-GHz antenna beam, at least 36 dB of cross polarization within the 3-dB

TABLE 1. COMSTAR BEACON E.I.R.P.S

Signal	Nominal e.i.r.p. (dBW)	Measured e.i.r.p. (dBW)	
		D1*	D2*
19 TV	23	24.9	24.8
19 TH	23	24.1	24.6
29 C	24	29.0	30.7
29 LSB	17	23.4	21.5
29 USB	17	23.0	21.8

*Post-launch measurements by C. Mahle and K. Betaharon at Clarksburg, Maryland; D1 on 7/6/76; DII on 9/8/76.

contour, and isolation greater than 40 dB between the 29- and 19-GHz ports. The antenna system electrical parameters are given in Table 2.

TABLE 2. ANTENNA SYSTEM ELECTRICAL PARAMETERS*

Parameter	19-GHz RV	19-GHz RH	28.56 GHz
Gain (dBi)	55.1	55.8	59.8
Isolation (dB)	>36	>36	>40
	>40	>40	>40

*Measurements performed by R. Gruner.

The 0.7-dB difference between the antenna gain for 19.04-GHz vertical polarization receive (19 RV) and 19.04-GHz horizontal polarization receive (19 RH) is due to the losses in the waveguide and diplexer beyond the 19 RH port. The antenna servo-system precision has been shown to have a backlash error of less than 0.004°.

The COMSTAR beacon receiver consists of an antenna-mounted RF section shown in Figure 1 and an IF processing section shown in Figure 2. The antenna-mounted section lies immediately behind the main reflector of the 5-meter antenna. The IF unit is some 10 meters away from the RF section in the control and data acquisition room. Since the 19 TVRV signal is the strongest signal present under most atmospheric conditions and particularly during severe fades, the phase-locked loop (PLL) receiver is designed to lock on the 19 TVRV signal. The receiver uses a voltage-controlled crystal oscillator (VCXO) controlled base frequency of 264.375 MHz multiplied as in the beacon to produce a local oscillator (LO) signal

TV), 19 THRH, 19 THRV, 39 C, 29 USB, and 29 LSB signals. They are also used to provide the inputs for phase measurement circuits. The 19-GHz phase (TVRV-THRH) results directly from phase comparison of these two signals at the same frequency. The 29 LSB and USB second IF signals are mixed with the 29 second IF to yield 69.4-kHz signals that are compared in phase to a reference 69.4-kHz signal (derived from the 5-MHz reference) to yield 29 (C-LSB) phase $\Delta\phi_{LC}$ and 29 (C-USB) phase $\Delta\phi_{UC}$. Finally, the 19 TVRV-29 C phase is obtained by mixing the IFs to produce a 2.5-MHz signal which is compared in phase to the reference 5-MHz frequency divided by two.

When the signals fade below a preset threshold, the 264.375-MHz VCXO control is transferred to a servo card reader, which is set to read the beacon diurnal frequency changes from a manually prepared card. The 1-kHz PLL has a much larger dynamic range and remains locked to the limit of useful 19-GHz amplitude data. Amplitude measurements may thus be maintained during deep fades, although phase information is lost while the signal is less than or a few decibels above the noise. When the 19 TVRV signal recovers to the acquisition level of the PLL, control is returned to the PLL and phase information is restored.

In addition to the 11 beacon signals, signals from a 12-GHz radiometric receive terminal and a tipping bucket rain gauge are recorded. The signals recorded from these 13 channels are listed in Table 3. Most of the recordings are continuous.

Table 4 lists the receive system data pertinent to a link budget calculation. Under moderately clear cloud-free skies, an atmospheric absorption loss of 0.7 dB can be anticipated. The link budget shown in Table 5 is based on these data, a satellite location of 125.5°W longitude and a receive site at Clarksburg, Maryland (77.3°W longitude, 39.2°N). In practice, the fade depth measured to actual loss of lock using DI is about 35 dB on the 19-GHz TV channel working directly into the 19-GHz mixers with no parametric amplifier. With DU, the e.i.r.p. is 0.8 dB higher, and the atmospheric absorption loss is also slightly less due to the higher elevation angle. Thus, an increase of about 1 dB in dynamic range to loss of lock can be anticipated. The threshold for restoring control of the LO frequency to the PLL from the curve-following servo is set for a signal at about 30 dB below the clear sky level, i.e., about 5 dB above loss of lock.

TABLE 3. RECORDED DATA CHANNELS

Data Channel Designation	Signal Description
19 TVRV	19.04-GHz signal, vertical polarization transmitted, vertical polarization received
19 TVRH	19.04-GHz signal, vertical polarization transmitted, horizontal polarization received
19 THRH	19.04-GHz signal, horizontal polarization transmitted, horizontal polarization received
19 THRV	19.04-GHz signal, horizontal polarization transmitted, vertical polarization received
29 Carrier	28.56-GHz signal amplitude transmitted and received
29 USB	28.29-GHz signal amplitude transmitted and received
29 LSB	28.82-GHz signal amplitude transmitted and received
$\Delta\phi_{UC}$	Phase between 28.56- and 28.82-GHz signals
$\Delta\phi_{LC}$	Phase between 28.56- and 28.29-GHz signals
19/19 Phase	Phase between 19.04-GHz TVRV and THRH signals
29/19 Phase	Phase between 28.56-GHz and 19.04 TVRV signals
12 Sky Temperature	12-GHz sky temperature variations (from radiometer)
Rain Gauge	Rainfall (measured by tipping bucket rain gauge)

TABLE 4. RECEIVE SYSTEM PARAMETERS

Parameter	Direct into Mixer	Paramp Ahead of Mixer
Waveguide Losses (dB)	1	1
Paramp Gain (dB)	—	≥ 15
Paramp Noise Figure (dB)	—	≤ 4
Mixer Conversion Loss (dB)	5	5
Mixer Noise Figure (dB)	6.2	6.2
Detection Bandwidth (Hz)	200	200
T_{system} , referred to Antenna Flange (K)	3,805	1,445

TABLE 5. LINK BUDGET FOR DI

	With Paramp	Mixer Only
Satellite e.i.r.p. for 19-GHz TV (dBW)	+24	
Path Loss (dB)	-210.0	
Atmospheric Absorption (clear, dry day) (dB)	-0.7	
Antenna Gain (dB)	+55.1	
Received Signal Level (dBW)	-131.6	
T_s (dB)	+31.6	+35.8
Boltzmann's Constant (dB)	-228.6	-228.6
Bandwidth (dB)	+23	+23
KTB (dBW)	-174	-169.8
Signal C/N (dB)	42.4	38.2
Loss in Modulation Sidebands (dB)	3.0	3.0
Dynamic Range of Amplitude Measurements (dB)	39.4	35.2

Attenuation data

For the 19- and 29-GHz vertically polarized signals, the frequency of occurrence of fading, accumulated to the 30-dB level, is given in Figure 3. During over 4,200 hours of data, the data base dynamic range of 30 dB was exceeded for nearly 100 minutes, *i.e.*, about 0.04 percent of the time, at 19.04 GHz and approximately 200 minutes, *i.e.*, about 0.08 percent of the time, at 28.56 GHz. (The fade level is referenced to a nominal clear-sky condition established from the 29-GHz data.) The cumulative rain rate statistics, measured on-site, for the same period of time are given in Figure 4. Measurable rain at the receiving site occurred for only slightly more than 1.5 percent of the time; on the other hand, measurable attenuation occurred for more than 10 percent of the time. Thus, while the cumulative fade level is not total atmospheric attenuation, it includes clouds, fog, and other non-rain effects.

Fades were by no means evenly distributed throughout the measurement period. Most of the deep fades, *i.e.*, ≥ 15 dB at 19 GHz, occurred during the late summer months, while lesser fades were more evenly distributed throughout the measurements. At these frequencies and elevation angles, light rain, which occurred throughout the measurement, induced lesser fades whereas only the higher rain rates of summer caused deep fades.

Fades were also unevenly distributed throughout the day. The diurnal

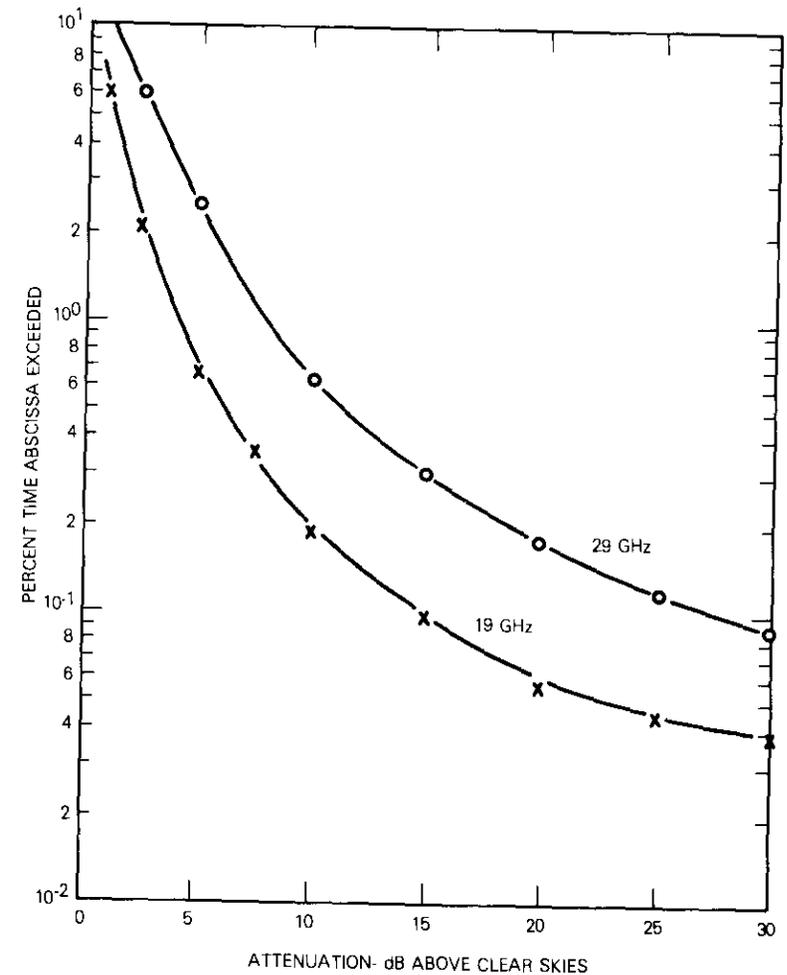


Figure 3. Cumulative Fading Statistics of 19.04- and 28.56-GHz Vertically Polarized Beacons from Mid-July 1976 to Mid-January 1977

variations in the solar heating of the atmosphere produce this uneven distribution of cloud cover and precipitation, particularly thunderstorms; hence, the probability density of fades is similarly distributed. Figures 5a and 5b give the diurnal distribution for fades exceeding 5 dB for the 19 TVRV and the 29-GHz carrier, respectively. These histograms show the relative time for which the fade level is exceeded during each hour of the

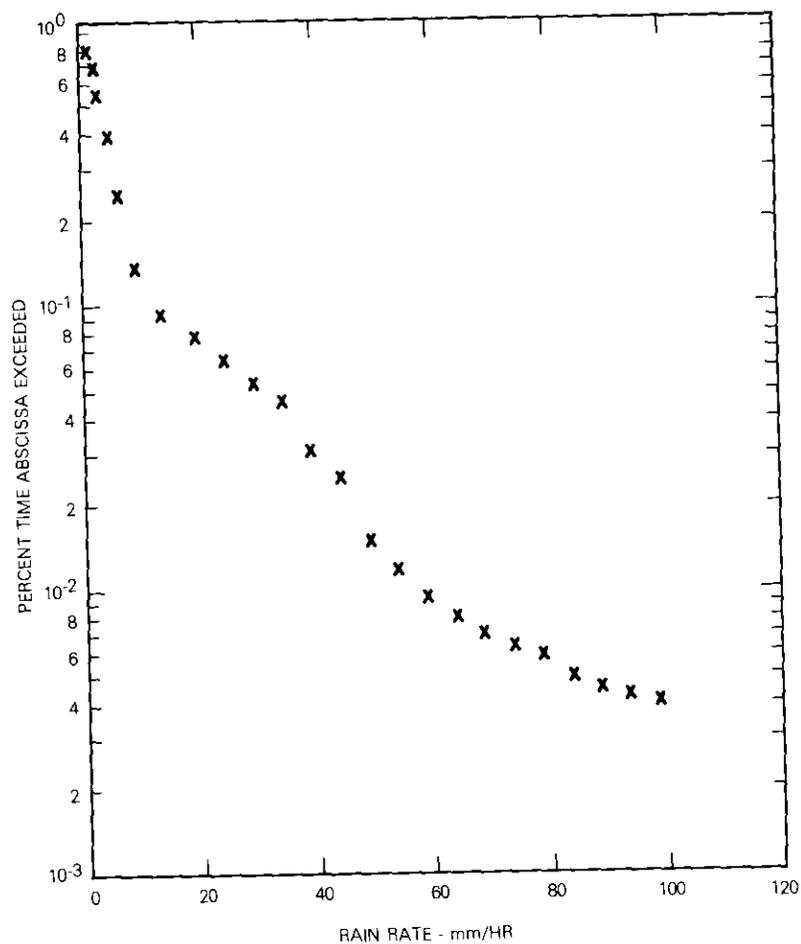


Figure 4. Cumulative On-Site Rain Rate Statistics from Mid-July 1976 to Mid-January 1977

day. Each distribution has two distinct peaks, one in the early morning (8–12 GMT), and the other in the late afternoon (21–23 GMT). The differential magnitude of these peaks can be explained in terms of the meteorological phenomena, since most of the early morning fading occurred in the non-summer months, while most of the late afternoon fading occurred during July and August. Late afternoon summertime fades were generally

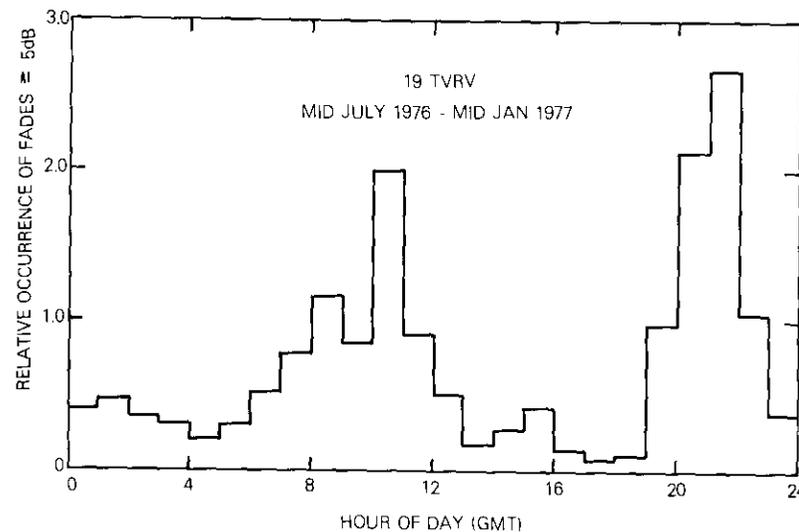


Figure 5a. Diurnal Distribution of Fades ≥ 5 dB at 19.04 GHz

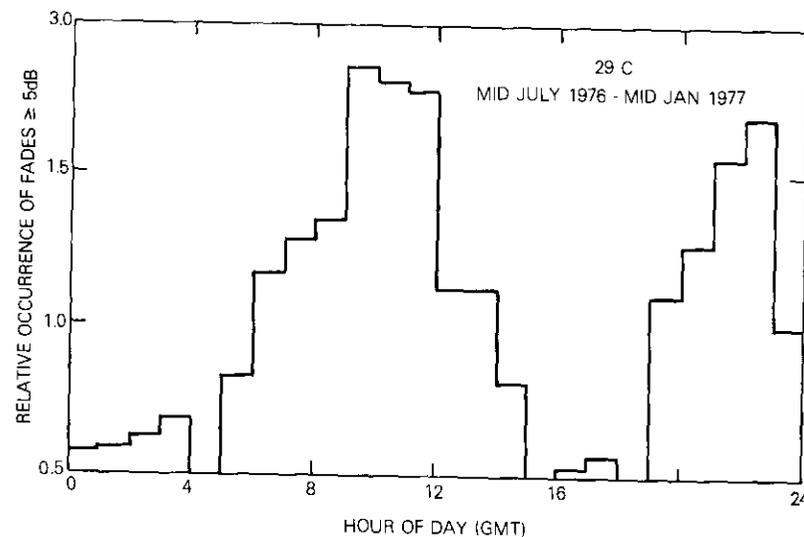


Figure 5b. Diurnal Distribution of Fades ≥ 5 dB at 28.56 GHz

deeper and of shorter duration due to the increased occurrence of convective rainstorms characterized by intense rainfall and higher altitude melting layers.

The statistical correlation of the cumulative fading at the two frequencies is indicated in Figure 6 (solid line with circles), which shows the ratio of 29- to 19-GHz fade levels for the same percentages of time. The 29/19-GHz fade ratio is a decreasing exponential function of the 19-GHz attenuation, with the maximum ratio occurring at low fade levels. It is approximately 2.5 at the 1-dB level and decreases to about 2.0 at the 20-dB level. The well-behaved single valuedness of this relationship indicates that some path-related parameter is responsible for the decreasing ratio with increasing attenuation. The negative slope of this relationship also suggests that small diameter hydrometeors, which compose the light rain or fog and produce low-level fading, affect the 29-GHz signals much more severely than 19-GHz signals. As the hydrometeors become larger the fade ratio decreases.

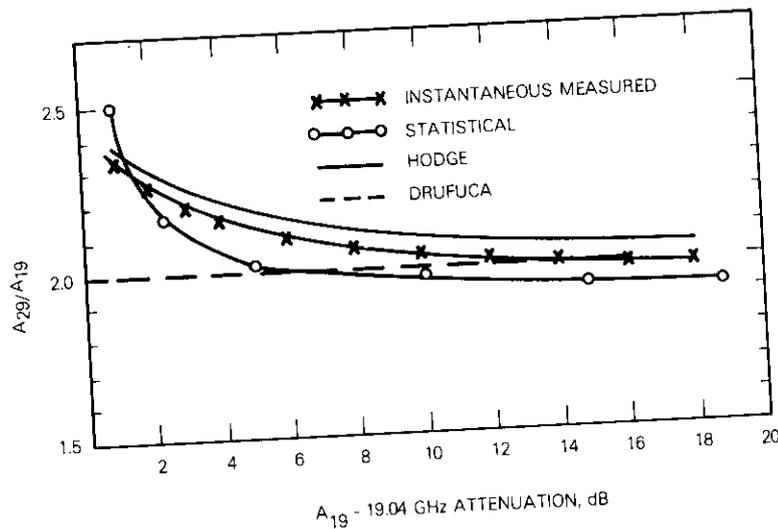


Figure 6. 28.56-GHz to 19.04-GHz Fade Ratio

The dashed line shown in Figure 6 is the result derived from the empirical equation given by Drufuca [3],

$$\frac{A_f(\text{dB})}{A_{f_0}(\text{dB})} = \left(\frac{f}{f_0}\right)^{1.72} \quad (1)$$

where A is attenuation and f is signal frequency. For $f_0 = 28.56$ GHz and $f = 19$ GHz, this equation yields a value of about 2, which matches measured data only at cumulative fade levels above about 8 dB. It should be noted, however, that Drufuca's result was derived primarily for application between 11 and 19 GHz.

Correlation of simultaneous samples of the attenuation at 19 and 29 GHz was also found to be quite good. Several 1-hour blocks of data, sampled at 0.5-minute intervals to give 120 points, were used in the event analysis. Scattergrams (plots of the simultaneous decibel values of 19- and 29-GHz signal fade levels), ratiograms (plots of the ratios of the decibel values of these signals), and least-mean-square fits were developed from each data block to yield the numerical values of α and β for the relationship

$$A_{29} = \alpha A_{19}^{\beta} \quad (2)$$

where A_{29} and A_{19} are in dB.

Figures 7a and 7b show typical event recordings for 19-TVrv and 29-C signals, respectively. The correlation of the attenuation is evident. However, the ratiogram (Figure 8a) and scattergram (Figure 8b) indicate that the instantaneous relationship is a nonstationary process, probably due to time variations of the medium such as changes in the distribution of drop size or drop canting. As in the statistical correlation discussed above, these figures indicate a decreasing fade ratio with increasing fade level. For the several events analyzed, computed correlation coefficients were found to be greater than 0.95. Values for α and β ranged from 1.5 to 2.5 and 0.8 and 1.0, respectively.

A composite scattergram for these several events, shown in Figure 9, yielded

$$A_{29} = 2.35 A_{19}^{0.94} \quad (3)$$

This composite scattergram included both early morning and late afternoon fade events. The simultaneous fade ratio derivable from this equation is given in Figure 6 (solid line with '+'s). The expected attenuation ratio can also be calculated using an empirical formula which relates the total attenuation to a uniform effective rain rate and effective path length, L :

$$A(\text{dB}) = aR^bL \quad (4)$$

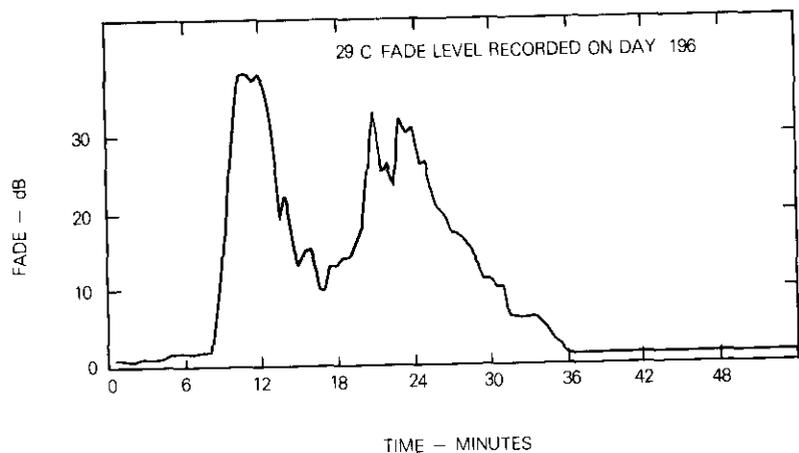


Figure 7a. Time Record of 29-GHz Fade Recorded on July 14, 1976

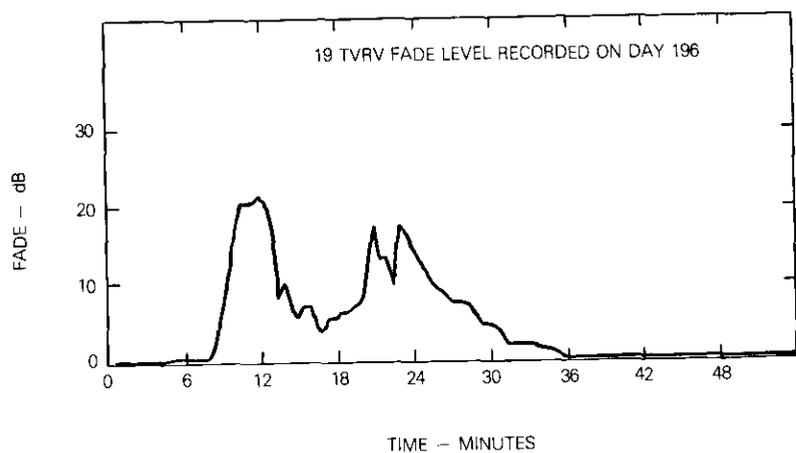


Figure 7b. Time Record of 19-GHz Fade Recorded on July 14, 1976

For the Laws and Parsons drop size distribution, constants a and b for 19 and 29 GHz are found from a best fit curve of Medhurst's data [4] on specific attenuation:

$$A_{29} = 0.155R^{1.04}L \quad (5a)$$

$$A_{19} = 0.059R^{1.1}L \quad (5b)$$

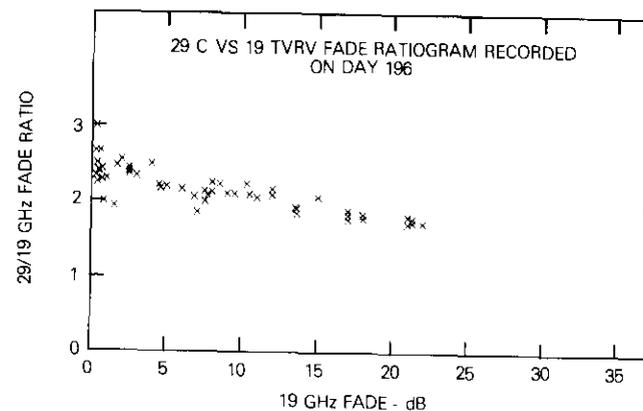


Figure 8a. Ratiogram of Event given in Figure 7

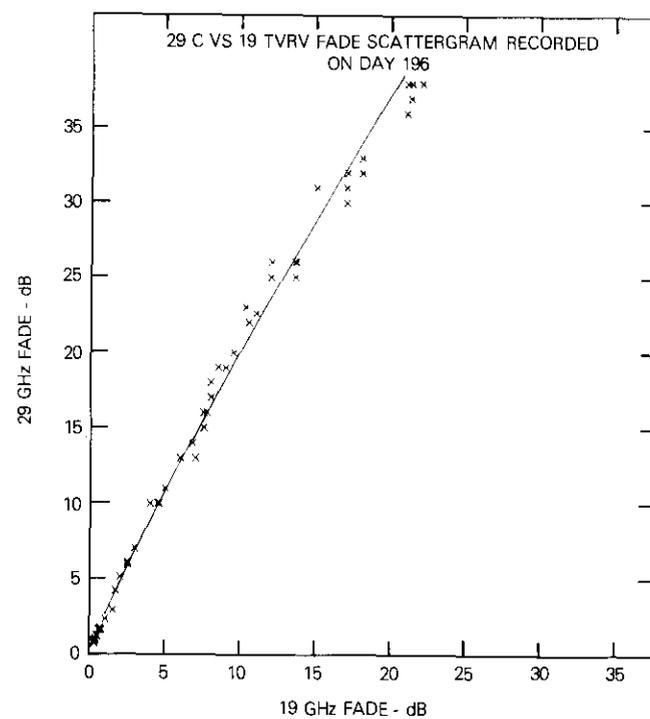


Figure 8b. Scattergram of Event given in Figure 7 (least-square-fit:
 $A_{29} = 2.35$, $A_{19} = 0.94$)

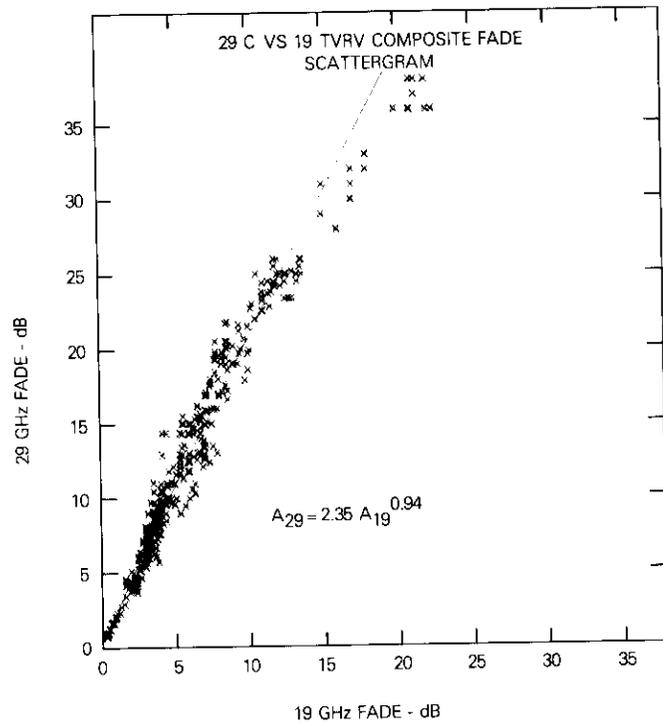


Figure 9. Composite Scattergram of Several Fade Events (least-square-fit: $A_{29} = 2.35, A_{19} = 0.94$)

Hence, the fade ratio, A_{29}/A_{19} , is

$$\frac{A_{29}}{A_{19}} = 2.63R^{-0.06} \tag{6}$$

which, of course, depends only on the path rain rate.

Hodge [5] recently showed that a significant improvement in frequency scaling could be obtained by introducing an inhomogeneity and correlation correction into the theoretical calculation of rain attenuation. Using equation (10) in Reference 5 and the values of a and b given in equations (5a) and (5b) one obtains

$$A_{29} = 2.38A_{19}^{0.95} \tag{7}$$

Equation (7) agrees closely with the measured data plotted in Figure 6 and shows remarkably good agreement with the measured instantaneous ratios given by equation (3).

Comparison of theory and measurements from equations (3) and (6) yields a relationship between A_{19} and rain rate, R ,

$$2.35A_{19}^{-0.06} = 2.63R^{-0.06} \tag{8}$$

Substitution from equation (5b) into equation (8) results in an empirical relationship between the so-called uniform effective path length, L , and the uniform effective rain rate, R :

$$L(R) = 2.6R^{-0.1} \tag{9}$$

Equation (9), plotted in Figure 10, models the physical meteorological condition of increasing path length with decreasing effective rain rate. Such an analysis results in path length and rain rate parameters which fit the measured data at both 19 and 29 GHz and thus may be frequency independent. The conventional technique of comparing measured rain

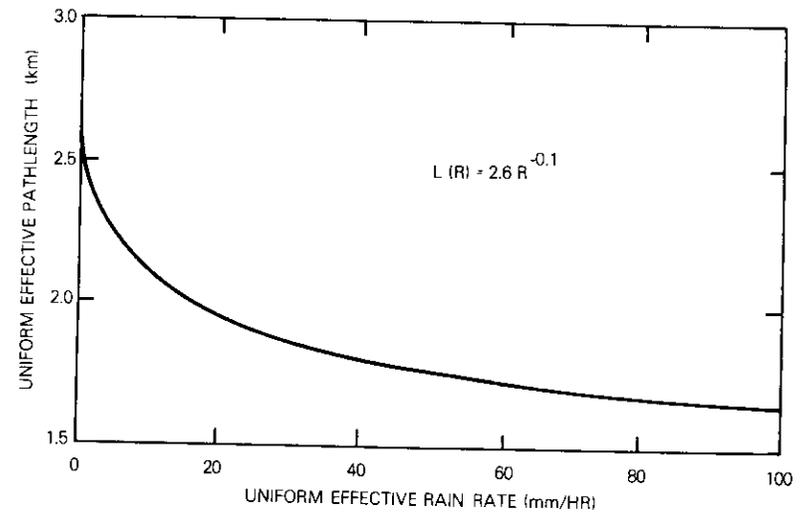


Figure 10. Variation of Uniform Effective Path Length with Uniform Effective Rain Rate

rate and attenuation statistics results in a frequency-dependent path length larger than that given by equation (9).

Depolarization data

The polarization switching at 19.04 GHz allows direct measurement of the power increase received in an orthogonal linear polarization due to path-induced wave depolarization. The satellite beacon transmitter is configured so that at Clarksburg, Maryland, the electric field vector of one signal (19 TH) is nearly parallel to the local horizontal and that of other (19 TV) is about 21° off the local vertical. Under clear skies, cross-polarization signal levels are more than 30 dB below the copolarization signal levels. Measurements of cross-polarization power alone do not directly yield cross-polarization isolation because the simultaneously occurring distributed attenuation tends to mask depolarization along the path. Nevertheless, careful interpretation of the copolarized, cross-polarized, and differential phase data can be used to estimate the isolation values. Several events which typify the depolarization data measured to date are discussed in the following paragraphs.

The event of Figure 11, recorded on September 16, 1976, gives the 19 TVRV, 19 THRH, and 29 TVRV attenuation; 19 THRV cross-polarized signal power; computed cross-polarization isolation; and rain rate. The peak copolarized signal fading levels were about 9.3 and 9.8 dB for the 19 TVRV and 19 THRH signals, respectively. The maximum on-site rain rate was 21 mm/hr. Cross-polarized signal power increased slightly prior to the deep fade which occurred at 0554 GMT. At 0546 attenuation began to dominate and continued to do so until about 0609, when the cross-polarized signal reached its highest level. The hump in the cross-polarized signal level is thought to be the result of differential phase induced by non-absorbing hydrometeors on the path, since no significant change in copolarized signal attenuation occurred simultaneously. If it is assumed that the cross-polarized signal is attenuated by an amount equal to that of the copolarized signal, then the degradation in cross-polarization discrimination can be estimated. This degradation of the clear weather XPD (lower trace, Figure 11) is given by the sum of the copolarized signal fade and the increase in cross-polarized signal power, all in dB. Thus, for this event, the cross-polarization discrimination degraded to about 16 dB at 0554 and 21 dB at 0612.

Figure 12 was also recorded on September 16, 1976. As in the previous

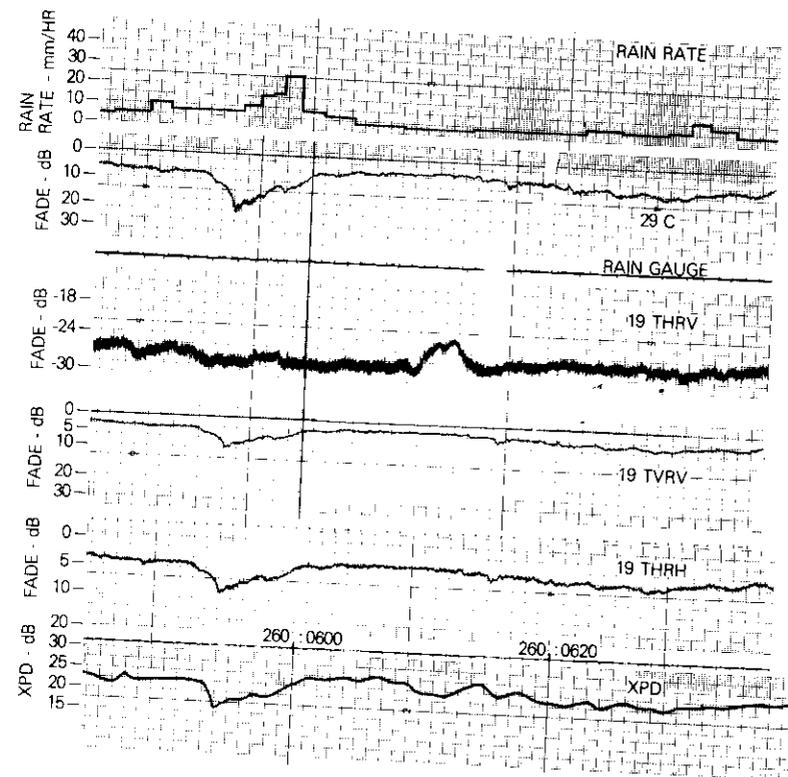


Figure 11. Stripchart Record of Depolarization Event Recorded on September 16, 1976

event, the maximum copolarized signal fading was less than 10 dB at 19 GHz. In this case, however, the on-site rain rate was much higher, eventually reaching about 50 mm/hr. In this cross-polarized signal recording, high-frequency fluctuations occurred between 1050 GMT and 1140 GMT. The occurrence of these fluctuations is distinctly correlated with the on-site rainfall. Path-induced depolarization is confirmed by the cross-polarized signal levels between 1125 GMT and 1135 GMT, which show increases corresponding to the copolarized signal fading. In addition, from 1140 GMT onward significant path-induced cross-polarized signal fading was measured with a relatively low copolarized signal attenuation (less than 2 dB).

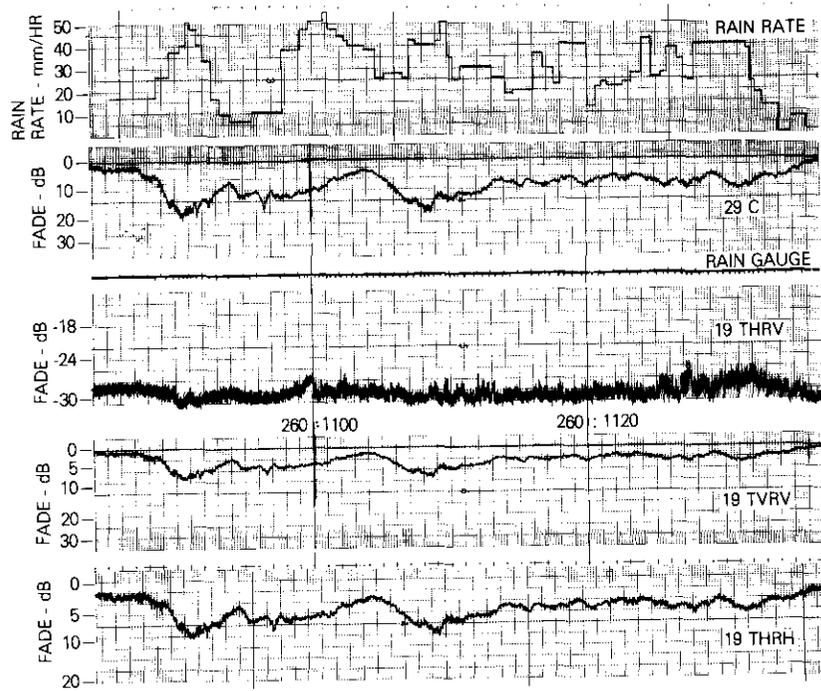


Figure 12. Stripchart Record of Depolarization Event Recorded on September 16, 1976

Other path-induced depolarization events of much longer duration and higher magnitude have been recorded. For instance, a depolarization event in which the clear-weather cross-polarization discrimination decreased to as low as 10 dB is shown in Figure 13. Again, the physical mechanism appears to be differential phase shift, since copolarized signal attenuation was relatively low (less than 5 dB). In addition, the depolarization does not appear to occur along the same segment of the path as the attenuation, since fading of nearly equal magnitude is induced on both the copolarized and cross-polarized signal amplitudes. This latter fact suggests that the differential phase shift may be due to ice crystals aloft or possibly to effects of the melting layer interface that can be described in terms of the Fresnel transmission coefficients.

The existence of ice crystals as a mechanism is further supported by the abrupt decrease in cross-polarized power just prior to the detection of

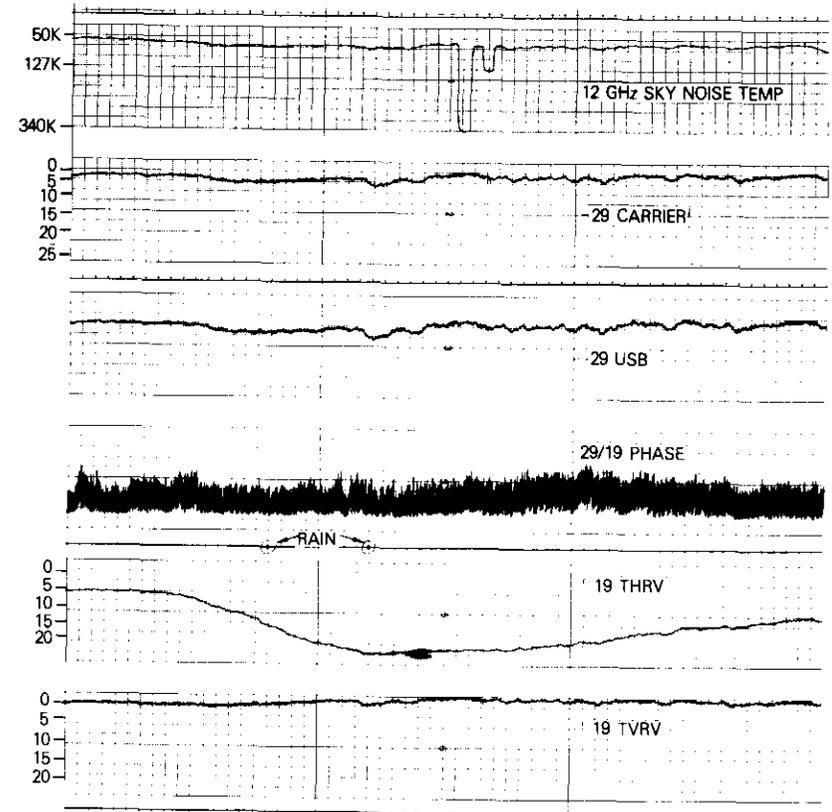


Figure 13. Stripchart Record of Depolarization Event with Very Poor XPD

rain (on-site), which might coincide with melting of the ice in a quasi-stationary rain cell or passing of a precipitating cell at a higher equilibrium temperature. By far the highest levels of cross-polarized signal amplitude are encountered in this type of depolarization event due to the high attenuation encountered in rain-induced depolarization events. (If similar results were to occur at 29 GHz, then this mechanism would be a serious cause of up-link cross-polarization interference in a dual-polarized system.)

Figure 14 consists of scattergrams relating the 19 TVRV and 19 THRH fading and differential attenuation for a typical event. A best least-mean-square fit performed from this data gave a correlation coefficient greater

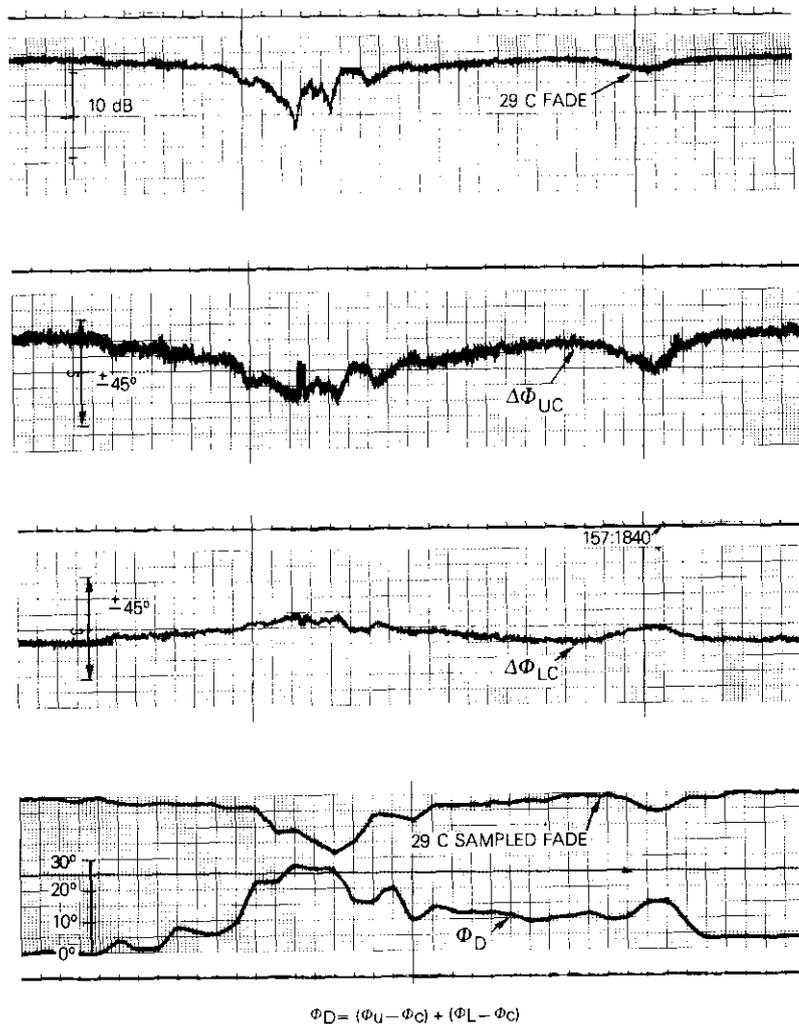


Figure 15. Stripchart Record and Data Analysis of Differential Phase Event

close correspondence and a frequency-independent relationship between effective path length and effective uniform rain rate has been derived. Finally, differential phase at ± 264 MHz from the carrier has been measured.

Acknowledgments

The work described herein represents the results of efforts by many people. The Microwave Laboratory has provided much of the equipment used in the measurements. The antenna feed OMT and polarizer design and subsequent antenna measurements were performed by R. Gruner of the Antenna Department and the receiver was built by A. Standing, K. Beta-haron, and C. Mahle of the Transponder Department. Within the Propagation Studies Department much support was received from K. Mekhayarajanonth, who performed the data reduction; from J. Talbot, who operated the receiver at all hours of the night and day; and from W. Mercer and C. Hart, who supported the installation and operation efforts tirelessly. This paper would not have been written except for the patient encouragement of P. L. Bargellini and L. Pollack.

References

- [1] L. Pollack, "Centimeter Wave Beacons for the COMSTAR Satellites," *COMSAT Technical Review*, Vol. 7, No. 1, Spring 1977, pp. 103-108.
- [2] D. C. Cox, "Design of the Bell Laboratories 19 and 28 GHz Satellite Beacon Propagation Experiment," International Conference on Communications, 1974, Paper 27-E.
- [3] G. Drufuca, "Rain Attenuation Statistics for Frequencies above 10 GHz Rain Gauge Observation," *I.U.C.R.M. Colloquium on the Final Scale Structure of Precipitation and EM Propagation*, Nice, France, October 1973.
- [4] D. B. Hodge, "Frequency Scaling of Rain Attenuation," *IEEE Transactions on Antennas and Propagation*, AP-25, No. 3, May 1977, pp. 446-447.
- [5] R. G. Medhurst, "Rainfall Attenuation of Centimeter Waves: Comparison of Theory and Measurements," *IEEE Transactions on Antennas and Propagation*, AP-13, No. 4, July 1965, pp. 550-563.
- [6] L. S. Ippolito, "20- and 30-GHz Millimeter Wave Experiments with the ATS-6 Satellite," NASA TN D-8197, April 1976.



J. M. Harris was born in Sardis, Mississippi, on August 27, 1951. He received a B.S.E.E. degree (1973) from the University of Mississippi, Oxford, Miss., and an M.S.E.E. degree (1974) from the California Institute of Technology, Pasadena, Cal., where he was an Earle C. Anthony Fellow. Mr. Harris' current interests are electromagnetic field interactions with random and turbulent media and electromagnetic modeling of distributed periodic structures.

Mr. Harris joined COMSAT Laboratories in 1974, where he is primarily engaged in research on electromagnetic wave propagation through the atmosphere at centimeter and millimeter wavelengths.

Geoffrey Hyde is a native of Toronto, Ontario, Canada. He received his B.A.Sc. in engineering physics and M.A.Sc. in electrical engineering from the University of Toronto in 1953 and 1959, and his Ph.D. in electrical engineering from the University of Pennsylvania in 1967.



Prior to joining COMSAT Laboratories in July 1968, Dr. Hyde worked on antennas, microwaves, and propagation at RCA at Moorestown, N. J., at Avro Aircraft Co., and at Sinclair Radio Labs in Canada. At COMSAT Laboratories prior to 1974 he was concerned with the development of the torus antenna, a general antenna analysis computer program (GAP), and related areas. Since February 1974 he has been Manager, Propagation Studies Department, where his work included the ATS-6 13/18-GHz COMSAT Propagation Experiment and a wide variety of efforts in propagation measurements and analysis.

Honors include David Sarnoff Fellowships (1965 and 1966) and the IEEEG-AP award for best paper, 1968 (jointly with Dr. Roy C. Spencer). He is a member of IEEE and the IEEE Antenna and Propagation Society, Adcom, URSI Commissions B and F, and U.S. C.C.I.R. Study Group 5, and is a registered professional engineer in Ontario, Canada.

CTR Notes

Characteristics of PCM telephone voice channels

A. G. GATFIELD

(Manuscript received July 7, 1977)

Introduction

In recent years, digital transmission with time-division multiplexing has been gradually but extensively introduced into national telephone networks. Leased or dialed long-distance channels often contain one or more PCM links. Communications engineers who have long understood degradation in analog FDM voice channels must now consider the peculiar characteristics of PCM voice channels.

Each PCM link introduces impairments such as amplitude, delay distortion, and white noise. These impairments, resulting from the analog portion of the link, are usually small and therefore do not constitute controlling factors.

Controlling factors for a PCM link are quantization noise, clipping noise, and bit-error noise. This note considers the combination of these three impairments and presents in graphical form the result of a study of the effect of bit errors on PCM signal quality.

This note is based on work performed under the sponsorship of the International Telecommunications Satellite Organization (INTELSAT). Views expressed in this paper are not necessarily those of INTELSAT.

A. G. Gatfield is Manager, Image Processing Department, at COMSAT Laboratories.

Since information on clipping and quantization error has been published [1], the most significant contribution of this note is the effect of bit errors. Two forms of graphical presentation are employed. Total weighted noise power resulting from clipping error, quantizing error, and bit-errors is plotted as a function of input signal level. Families of five curves are presented for bit-error rates from 0 to 10^{-3} . A second set of curves intended for easy comparison with C.C.I.T.T. Recommendation G.712 presents the same information in the form of unweighted, band-limited signal-to-noise (S/N) ratio. Exponential signal distribution reveals the effect on voice. Sine wave and Gaussian signal distributions indicate the expected results from the two common measurement techniques.

Although the equations for the three basic types of distortion or noise given in the following section are difficult to handle analytically, they have been numerically evaluated by computer to produce the curves presented here. However, the rms addition of three different signal-dependent forms of noise having different distributions with time is inadequate to establish a measure for subjective evaluation ranking. A program of subjective evaluation is therefore recommended.

Basic equations

Quantizing noise is expressed as [2]

$$N_q^2 = \frac{2h^2}{3 \times 2^{2m}} \int_0^h p(x) \left(\frac{dx}{dy}\right)^2 dx$$

where N_q^2 = quantizing noise power

h = clipping level

m = bits/sample

x = signal amplitude

$p(x)$ = probability density function of the signal.

$y = F(x)$ = output of the compressor.

Clipping noise is expressed as [2]

$$N_c^2 = 2 \int_h^\infty (x - h)^2 p(x) dx$$

For bit-error noise without differential encoding x'_i is defined as the sample value at the expander output if an error occurs in bit i . Then $x - x'_i$ is the error resulting from a bit error in bit i .

Assume that P_e (bit-error rate) is small and that only one error exists in any given code word, which is a reasonable assumption for code words of 8 bits or less which are common in PCM systems. Then N_e^2 , the mean square bit-error noise, is

$$N_e^2 = 2 \int_h^\infty \left[\sum_{i=1}^m \frac{(x - x'_i)^2}{m} \binom{m}{1} P_e (1 - P_e)^{m-1} \right] p(x) dx$$

For very small P_e ,

$$N_e^2 = 2P_e \int_0^h \sum_{i=1}^m (x - x'_i)^2 p(x) dx$$

In the normal case of nonuniform quantizing, a complete mathematical expression would be difficult to obtain. However, the equation has been evaluated by computer for various companding laws, binary codes, and signal probability distributions.

With differential coding, bit errors occur in pairs and the equation is modified as follows:

$$N_e^2 = P_{ed} \int_0^h \sum_{i=1}^m [x - x'_{i,j}]^2 p(x) dx$$

where P_{ed} = channel bit-error rate including the effect of differential encoding

$x'_{i,j}$ = sample value at the expander output when an error occurs in bit i , and consequently the differential decoder produces an error in bit j .

This computer evaluation depends not only on the companding law and binary code, but also on the symbol structure of the differential decoder.

Noise versus signal level

The curves of Figures 1 and 2 help to explain the effect of the three kinds of noise and the manner in which they combine. The vertical axis represents total noise power in dBm0p. On this scale, 1 pW is -90 dBm0 so that the curves range from 100 mW (-10 dBm0) to 10 pW (-80 dBm0).

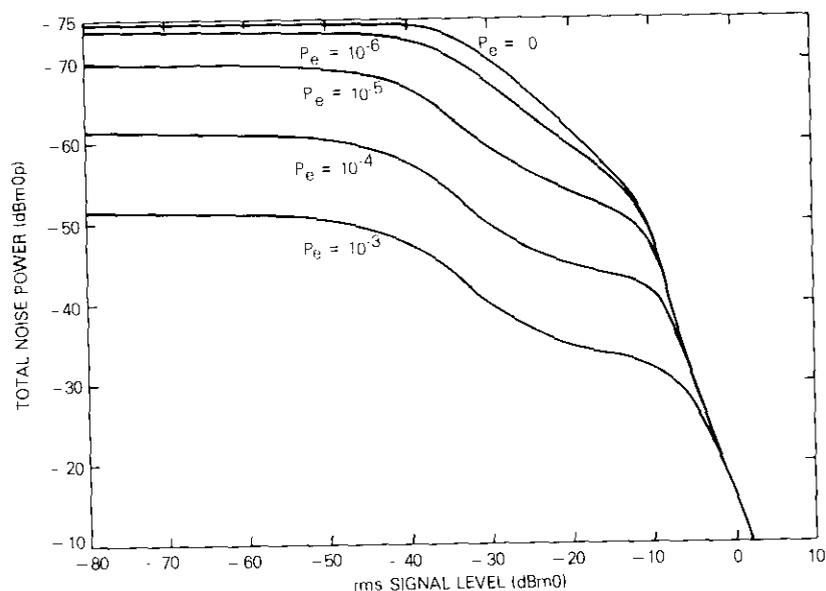


Figure 1. 13-Segment A-law (G.711) Folded Binary Code, Exponential Signal

The conventional method of combining noise power levels is an rms addition of voltages; however, this method presents some problems which will be discussed later.

For the μ -law and A-law encoders specified in C.C.I.T.T. Recommendation G.711 [1], the signal varies from zero to the clipping level at 3.14 dBm0 rms sine wave or 6.14 dBm0 absolute. The horizontal axis extends to -80 dBm0, which is low enough to be considered an idle channel. The normal active speech level is usually about -15 dBm0.

Figures 1 and 2 contain families of curves, one curve for each bit rate. The top curve ($P_e = 0$) corresponds to zero errors and thus presents only quantizing noise and clipping noise. The steep slope at the right of each figure where the curves seem to blend is the region in which clipping noise predominates. The signal for these curves has an exponential distribution to approximate speech; the slope is a function of this distribution. The second segment of the $P_e = 0$ curve slopes upward at approximately 45° , representing a nearly linear variation of quantizing noise with signal level, which is the intent of the μ -law or A-law compression schemes. Such a variation of noise with signal level results in a nearly constant S/N

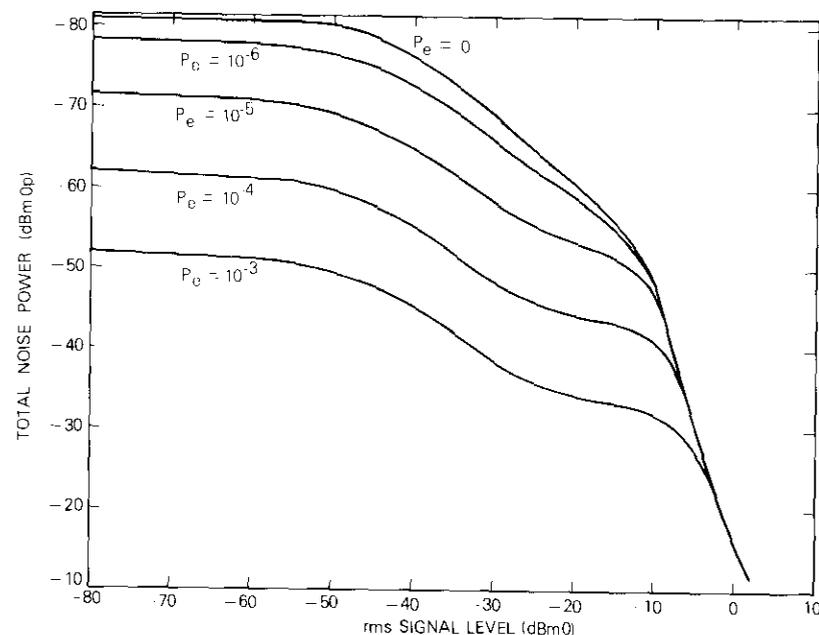


Figure 2. 15-Segment μ -law (G.711) Folded Binary Code, Exponential Signal

ratio. The final section of each curve at the lower signal levels is nearly horizontal and represents idle channel noise, which is a special case of quantizing noise with near-zero input signal.

As the bit-error rate increases, bit-error noise becomes the controlling factor. The 10-dB spacing between curves for error rates greater than 10^{-5} indicates the relatively negligible effect of other contributions on total noise.

Signal-to-noise ratio versus signal level

While the noise power figures are useful for visualizing PCM link performance, use of the S/N ratio is more common. C.C.I.T.T. Recommendation G.712, which concerns the quality of PCM channels, states the expected performance of a PCM link in terms of S/N for two different input signals.

Figures 3 and 4 show families of curves for a Gaussian test signal and Figures 5 and 6 for a sine wave signal. In each case the noise is unweighted

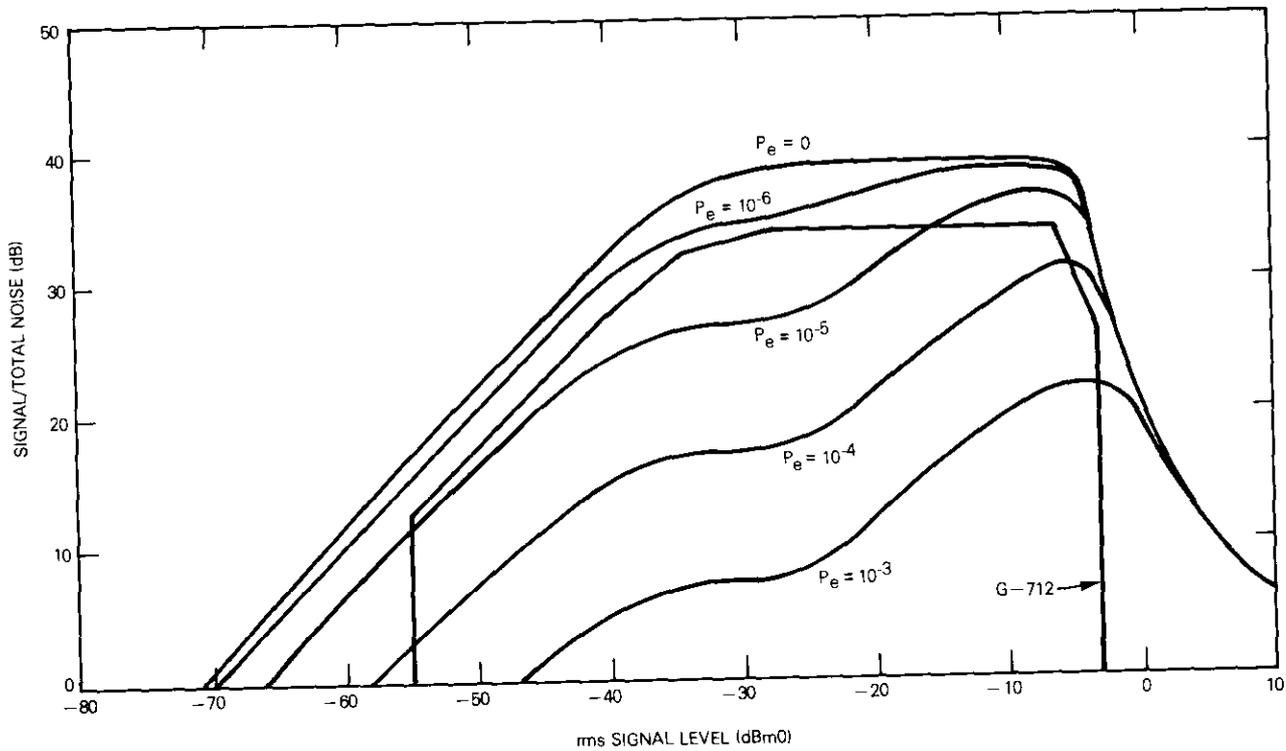


Figure 3. PCM Signal-to-Noise Ratio, G.711 A-law Coding, Gaussian Signal

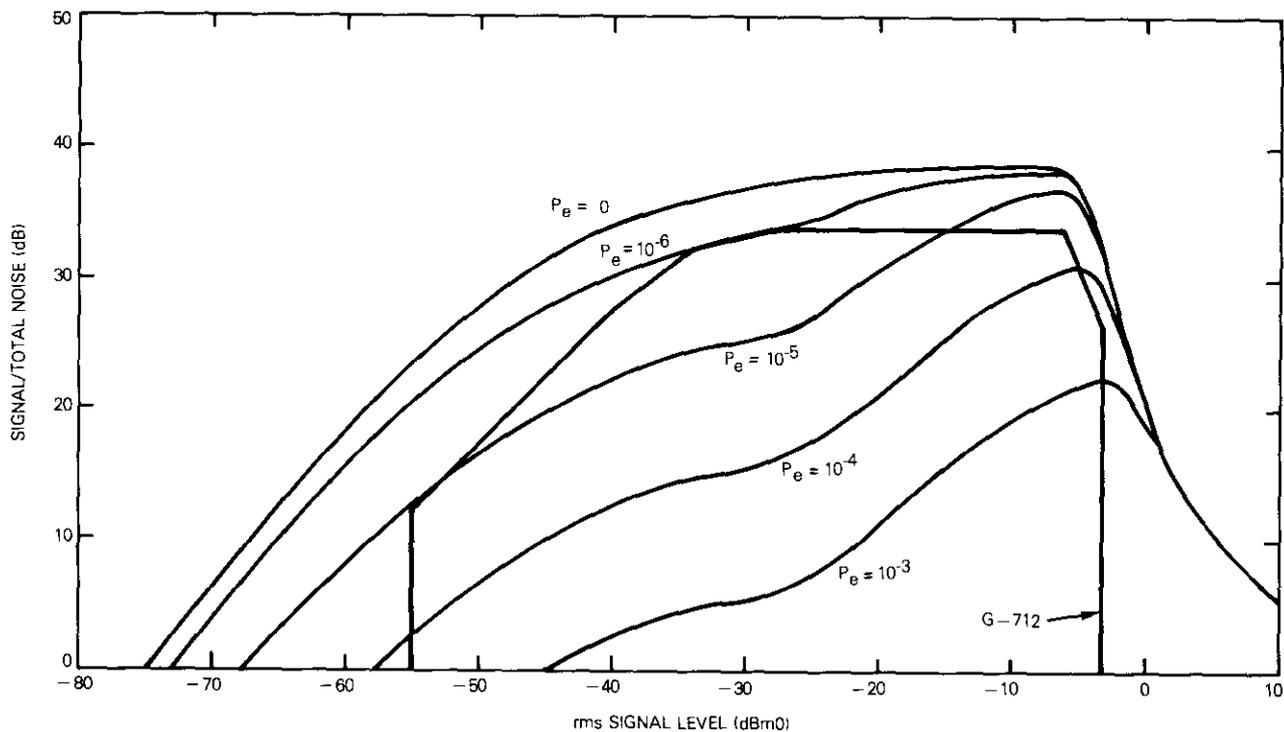


Figure 4. PCM Signal-to-Noise Ratio, G.711 μ -law Coding, Gaussian Signal

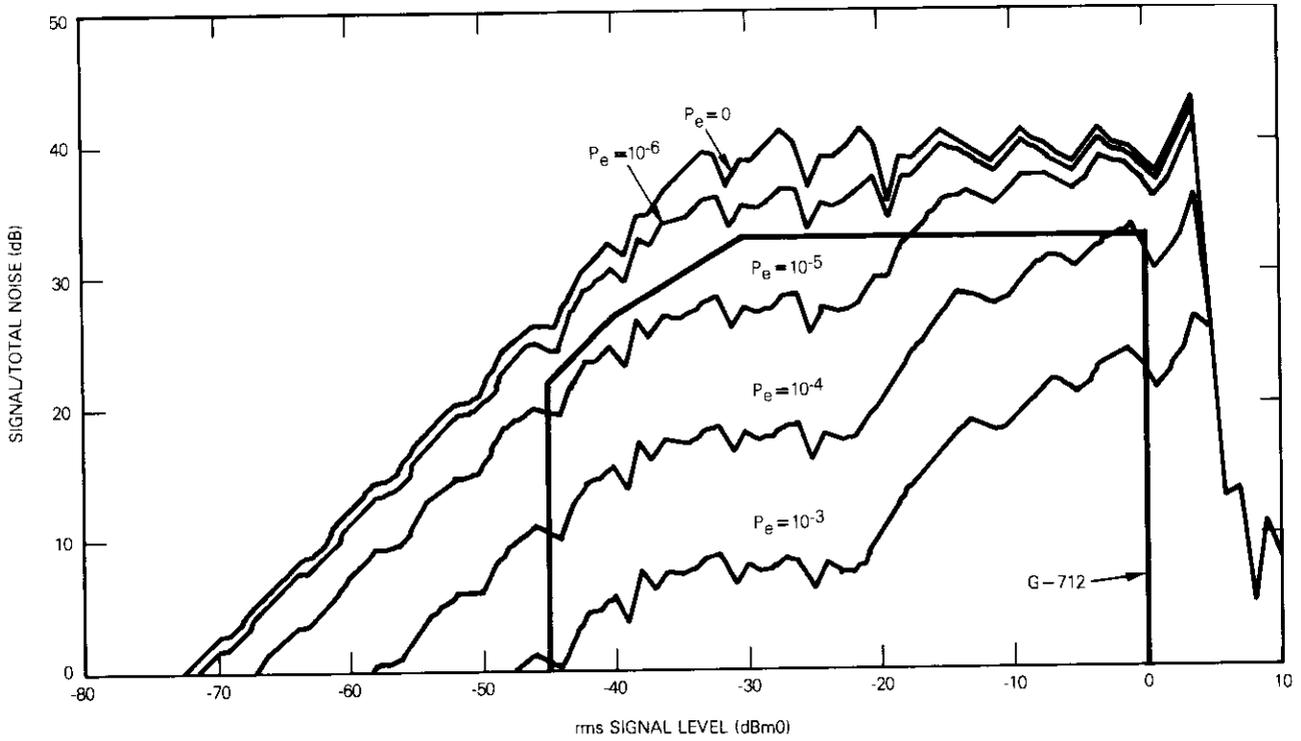


Figure 5. PCM Signal-to-Noise Ratio, G.711 A-law Coding, Sine Wave Signal

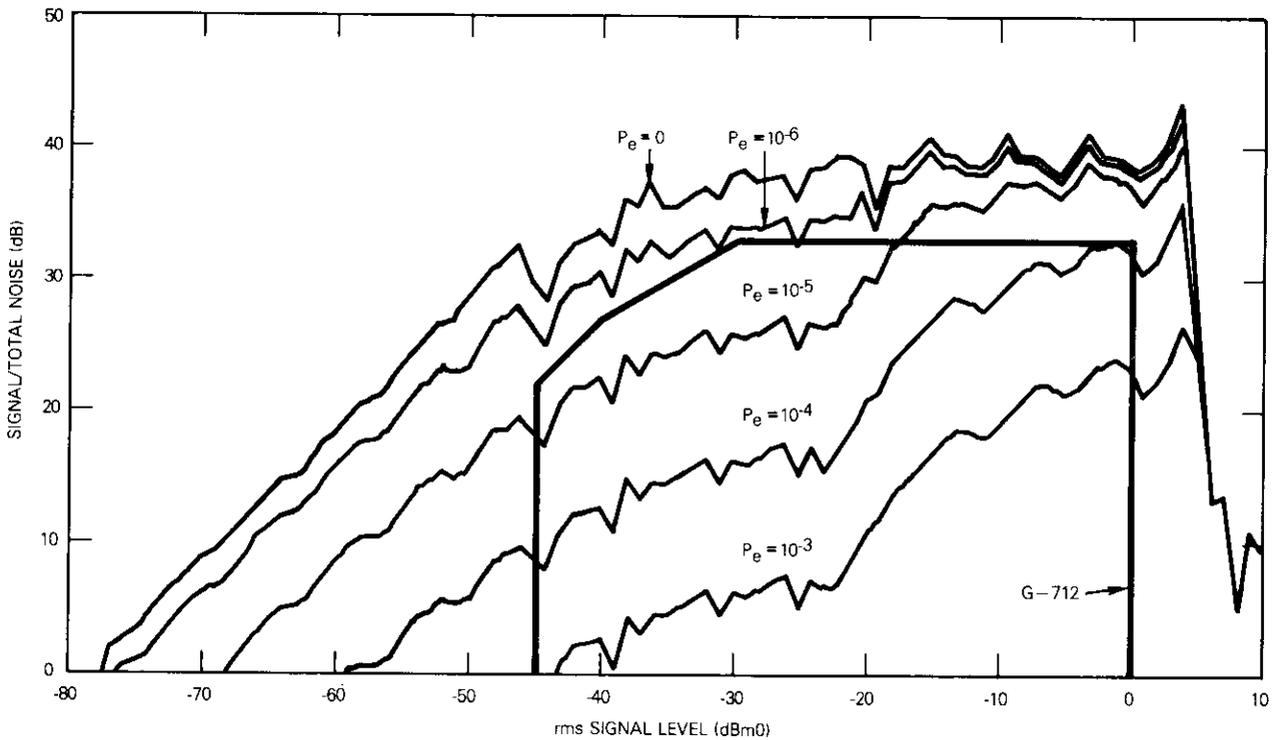


Figure 6. PCM Signal-to-Noise Ratio, G.711 μ-law Coding, Sine Wave Signal

to conform to Recommendation G.712. The effect of bit errors on curve shape and the effectiveness of the μ -law and A-law coding schemes in maintaining constant S/N ratio is evident from these figures, which also indicate the G.712 performance limits. From these plots of signal to rms noise ratio, it would appear that the highest permissible bit-error rate is 10^{-6} . This deduction conflicts with subjective evidence that an error rate of 10^{-4} is barely detectable.

Since differential coding is often used in 4-phase modems designed for satellite links, the effect of this form of coding (shown in Figure 7) on S/N is interesting. Figure 8 shows the S/N curves for 7-bit A-law PCM used in the SPADE* system. The performance limits of G.712 do not apply.

Subjective versus objective results

Some disagreement exists between the sensitivity of PCM to bit errors as depicted by the curves and that determined by subjective listening tests. For example, a bit-error rate of 10^{-4} is far below the C.C.I.T.T. specification, as shown in Figure 3. However, subjective observations indicate that a 10^{-4} error rate is detectable only by skilled observers under ideal conditions. The discrepancy is caused by the difference between the rms addition of noise power and the perception of this noise addition by the ear. If all of the noise components were Gaussian, no difference would exist. However, each noise component is signal sensitive, and the bit-error noise is impulsive, which makes it less objectionable to the human ear. Even though the rms level may appear to be excessive, the ear is insensitive to noise confined to very short pulses.

Conclusions

The curves presented herein reveal the contribution of the three forms of noise in PCM systems to total distortion. Although information on clipping noise and quantizing noise is available, the addition of bit-error rate noise information should be valuable to digital system designers. Further work based on subjective testing is needed to establish a permissible error rate for PCM voice transmission. Additional effort will also be required to determine the effect of bit-error noise on data transmitted by conventional data modems over analog circuits containing PCM links.

*SPADE is a voice communications system developed by COMSAT Laboratories for use in the INTELSAT system. It is a single-channel-per-carrier, PCM system with multiple-access and demand-assignment features.

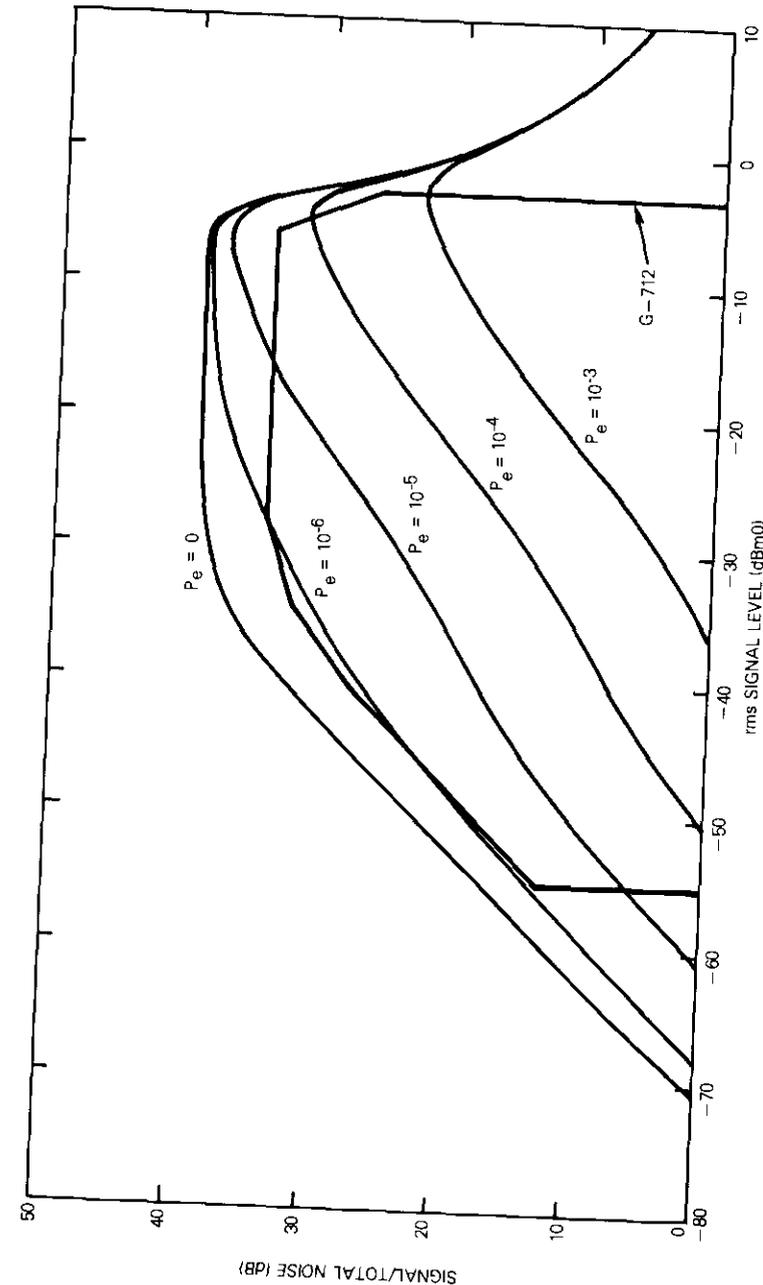


Figure 7. PCM Signal-to-Noise Ratio, G.711 A-law Coding, Gaussian Signal, 4ϕ Differential Modulation

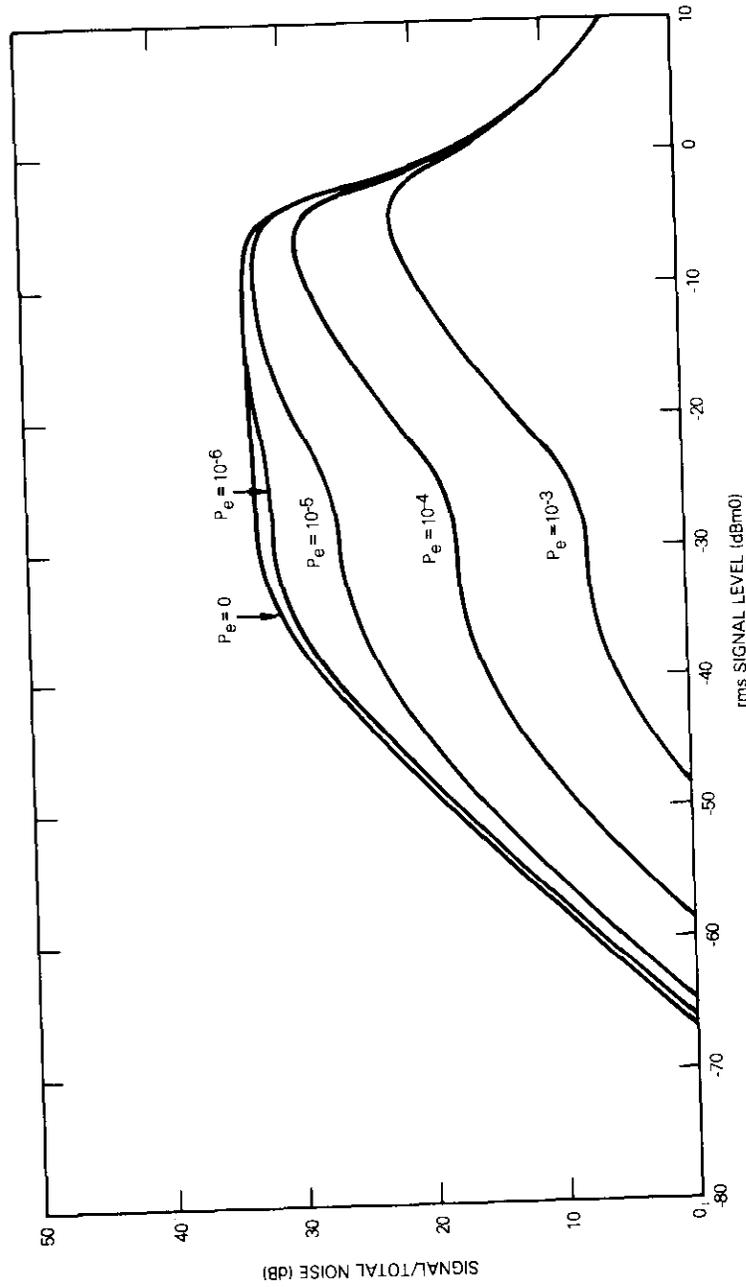


Figure 8. PCM Signal-to-Noise Ratio, SPADE 7-bit Coding, Gaussian Signal

Acknowledgment

The author wishes to acknowledge the contribution of Mr. Jesus Domingo for his early work in the development of the basic equations and initial computer programs. Dr. S. J. Campanella also contributed by emphasizing the importance of idle channel noise and by helpful discussions on the subjective effect of impulse noise.

References

- [1] International Telegraph and Telephone Consultative Committee (C.C.I.T.T.), *Green Book*, Vol. III-2, Recommendations G.711 and G.712, 1972.
- [2] K. W. Cattermole, *Principles of Pulse Code Modulation*, London: Iliffe Books Ltd., 1969.

Orbital performance of NTS-2 nickel-hydrogen battery

J. D. DUNLOP AND J. F. STOCKEL

(Manuscript received August 4, 1977)

Introduction

The U.S. Navy's Navigation Technology Satellite (NTS-2), launched from Vandenberg Air Force Base on June 23, 1977, uses a nickel-hydrogen (Ni-H₂) battery for its energy storage subsystem. Since this is the first nickel-hydrogen battery to be flight tested, its success constitutes a significant step toward the application of this new battery technology to future communications satellites.

Presently, nickel-cadmium (Ni-Cd) batteries are used for energy storage in space applications. These batteries, which are one of the major life-limiting elements, have an orbital life expectancy of approximately seven years. The Ni-H₂ battery has an orbital lifetime greater than 10 years and offers the added benefit of a significant weight reduction.

COMSAT Laboratories pioneered the development of the nickel-hydrogen battery under an R&D program sponsored by INTELSAT. This note briefly describes the battery design and orbital performance during the first month after launch.

This note is based on work performed under the sponsorship of the International Telecommunications Satellite Organization (INTELSAT). Views expressed in this note are not necessarily those of INTELSAT.

J. D. Dunlop is Manager, Electrophysical Devices Department, at COMSAT Laboratories.

J. F. Stockel is a Member of the Technical Staff, Electrophysical Devices Department, at COMSAT Laboratories.

Satellite description

The NTS-2 is the first satellite launched in the Joint Services NAVSTAR Global Positioning System. It was built by the U.S. Naval Research Laboratory (NRL) to operate for a minimum of three years at full power and to have a design life of five years.

The overall configuration of the satellite is shown in Figure 1. The body has the shape of a right octagonal prism, approximately 1.62 m (64 in.) wide and 0.79 m (31 in.) high, with a center tube 0.76 m (30 in.) in diameter. The satellite is gravity-gradient stabilized and uses reaction wheels for damping and yaw control. The solar array is oriented toward the sun by yaw control of the spacecraft and rotation of the array panels. The satellite is rotated 180° twice each orbit so that three of its rectangular surfaces continuously view deep space. One side of the satellite is never exposed to the sun.

Two eclipse seasons approximately 30 days long occur each year. The first eclipse season began on July 10, 1977. Since the eclipse occurs twice

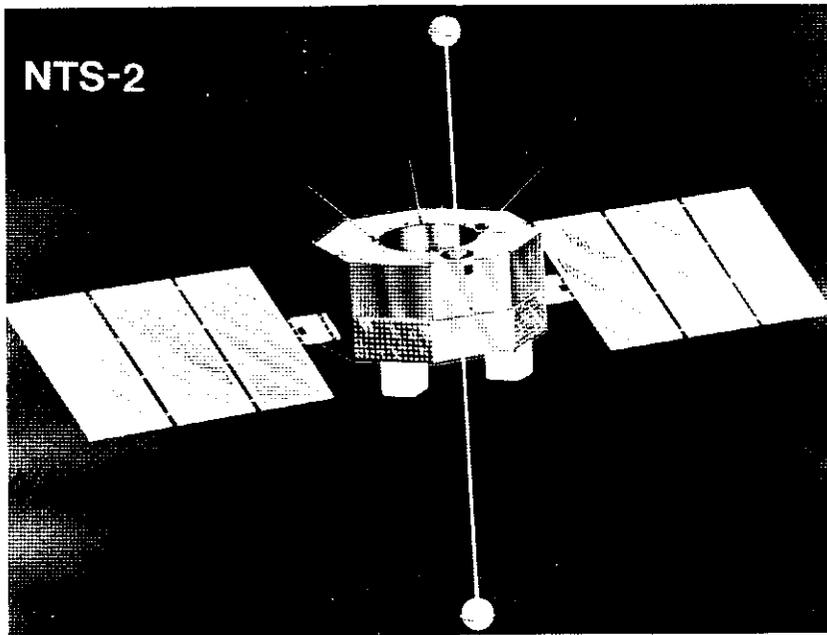


Figure 1. NTS-2 Satellite

every 24 hours, the battery is cycled twice a day, or 60 times during each eclipse season. The longest shadow period, which is in the middle of the season, lasts 0.94 hour as illustrated in Figure 2.

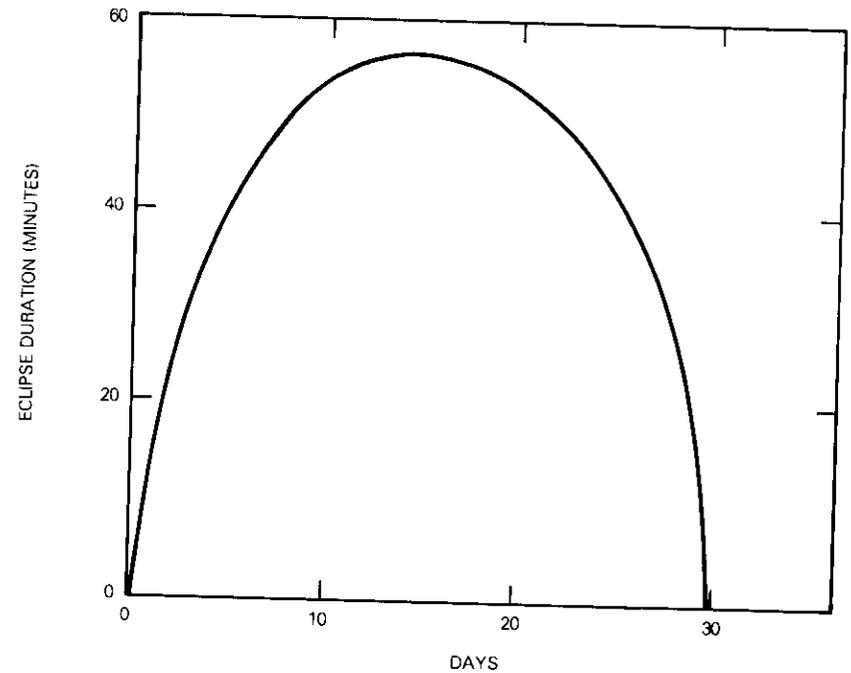


Figure 2. NTS-2 Eclipse Season (two seasons per year, two eclipses per day)

Battery design

SIZING

The nominal satellite design load [1] is 325 W. During eclipse periods the battery is discharged through a boost regulator which has an efficiency of 87 percent. This regulator maintains the bus voltage at 27 ± 1 V and increases the battery load to approximately 374 W. For the longest shadow period, the energy requirement for the battery is 354 Wh. To meet this energy requirement, the battery has fourteen 35-Ah Ni-H_2 cells (shown in Figure 3) and is discharged to approximately 60 percent of its available capacity.

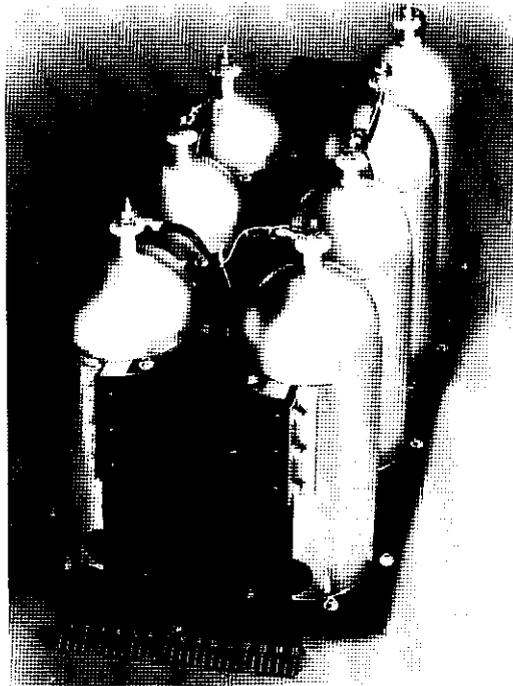


Figure 3. *Nickel-Hydrogen Battery Assembly*

Four of the cells have strain gauge bridges mounted on the domes of their pressure shells. Pressure values deduced from these strain gauge readings provide a direct measure of the cell capacity.

TEMPERATURE CONTROL

Temperature control is essential to overall battery performance and life expectancy. A direct radiator battery design concept suggested by COMSAT Laboratories was employed. NRL built the battery with INTELSAT supplied cells and integrated it into the spacecraft. Design objectives were to maintain the temperature between 0°C and 25°C during the longest eclipse day and between 0°C and 5°C between eclipse seasons.

The location of the battery assemblies within the satellite is shown in Figure 4. Each assembly is mounted on an exterior panel which continuously views deep space, providing a constant heat sink. Components mounted in the bays adjacent to the battery assemblies provide a radiative input to supplement reaction heating in the batteries.

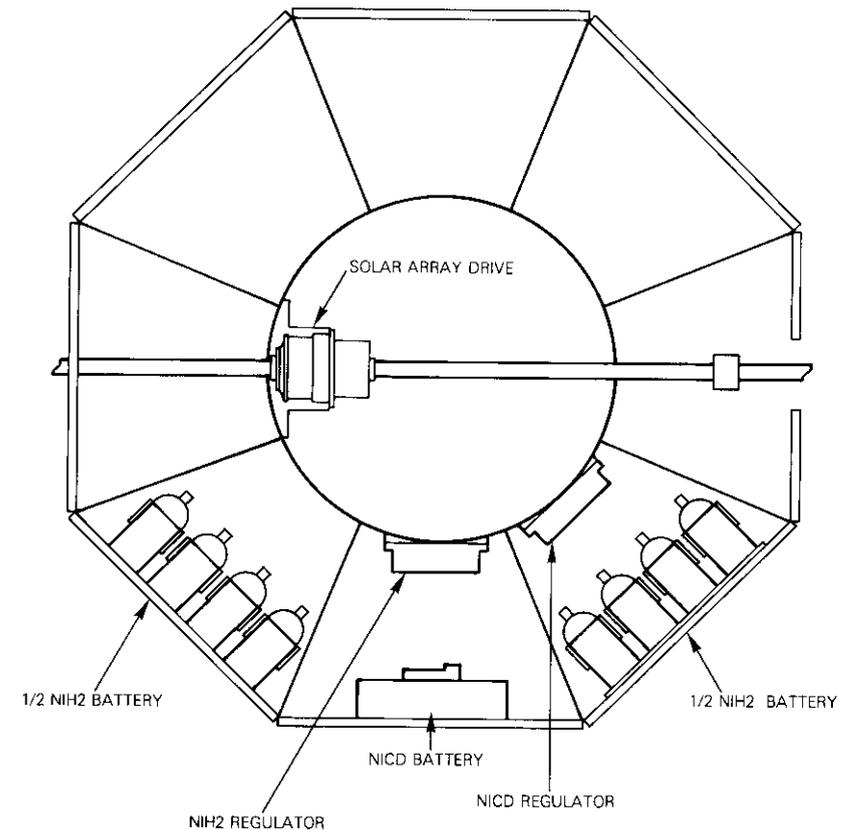


Figure 4. *NTS-2 Satellite Body Section Showing Battery Locations*

CHARGE CONTROL

Recharging the battery following eclipse discharge requires a nominal C/10 (3.6A) rate which is sufficient to fully recharge the battery during the sunlight period after the maximum eclipse period. Once the battery is fully charged, the C/10 rate would result in overheating. Based upon the thermal model temperature prediction for a mid-eclipse season orbit and upon initial thermal test data, a temperature of 15°C has been selected to terminate the main charge and to switch to a trickle charge rate of C/60 (0.6 A). To allow for temperature conditions higher than those predicted, alternate temperature set points have been provided in the 18°C to 24°C range.

Two additional operational provisions are incorporated into the system. An overtemperature limit of 35°C will switch the charger off, and an undervoltage condition of 14 V will shed noncritical loads. Both of these automatic controls can be disabled on command.

DATA ACQUISITION

In addition to operational status information, the following in-orbit battery parameters are available through telemetry: battery voltage, individual cell voltages, four temperatures per assembly, and strain gauge pressure values for selected cells.

Performance

PRELAUNCH AND ORBITAL

The Ni-H₂ battery was exercised extensively during the prelaunch operation at Vandenberg Air Force Base. State-of-charge indications from the strain gauges were a significant help during the spacecraft electrical checks. One hour prior to launch, the spacecraft began operating entirely from the battery and continued in this mode until the spacecraft first acquired the sun approximately 30 minutes after launch. The satellite was launched into an elliptical transfer orbit, during which three 30-minute eclipses were encountered. For 33 hours the satellite, which was spinning at 85 rpm, exerted a steady force of 6 g's on the battery. The satellite was then boosted into a circular orbit and despun to 40 rpm, which reduced the force on the battery to approximately 1.4 g's over the next 12 days. Finally, the satellite was despun and stabilized with its solar array deployed. Neither the dynamic forces during lift-off nor the steady load accompanying spinning had any detrimental effects on the battery.

During the final spacecraft stabilization period, the battery was inadvertently completely discharged three times, causing cell reversals. Subsequent recharge revealed no detrimental effects, illustrating dramatically the battery's tolerance to overdischarge. Since spacecraft stabilization, the battery has been operating in a 0-g environment with no adverse effects on performance.

ECLIPSE OPERATION

To illustrate battery performance during the eclipse season, telemetry data taken during the longest eclipse day (day 15) is presented. Figure 5 shows the battery voltage and cell temperature during the 0.94-hour

discharge and 11-hour recharge periods. The end-of-discharge voltage for the battery is 16.5 V or 1.18 V per cell; the corresponding discharge current is 23.6 A and the power output 388 W, which is slightly greater than the design load requirement of 374 W. As the boost regulator maintained the bus voltage at 26.6 V, the actual voltage performance was precisely as expected.

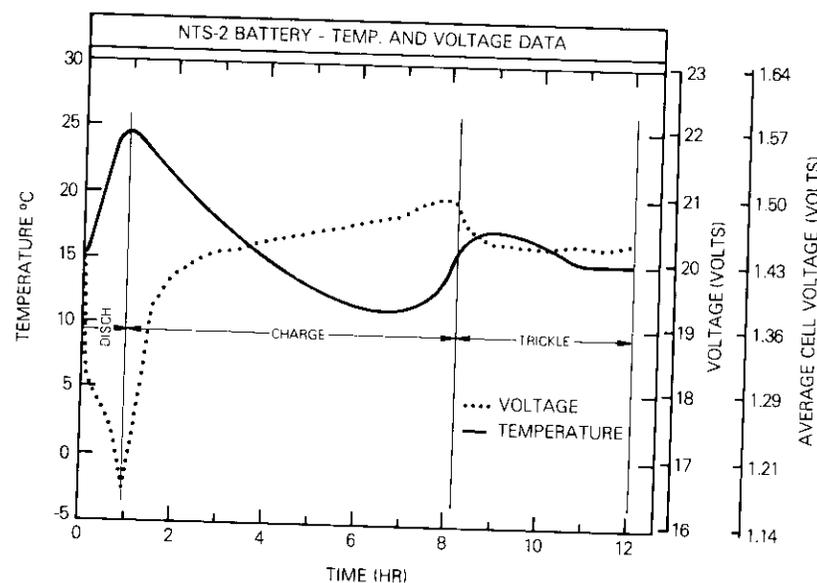


Figure 5. Battery Temperature and Voltage for a 12-Hour Cycle

The battery temperature increases to 25°C at the end of discharge due to battery heat dissipation as shown in Figure 5. This temperature value, which occurs at the end of the longest eclipse period, is the maximum battery operating temperature in accordance with design objectives. The battery is recharged at a constant current of 3.6 amperes. On recharge it is endothermic until it approaches the fully charged condition. To minimize overcharge heating, the current is reduced to a trickle rate of 0.6 ampere when the temperature reaches the set point of 15°C. (Note the initial voltage drop in Figure 5 when the current is switched to trickle charge.) Data taken during the satellite thermal vacuum tests [2] predicted the orbital performance very accurately.

Capacity and corresponding cell pressure data over the 12-hour charge-discharge cycle are shown in Figure 6. Data from the strain gauges indi-

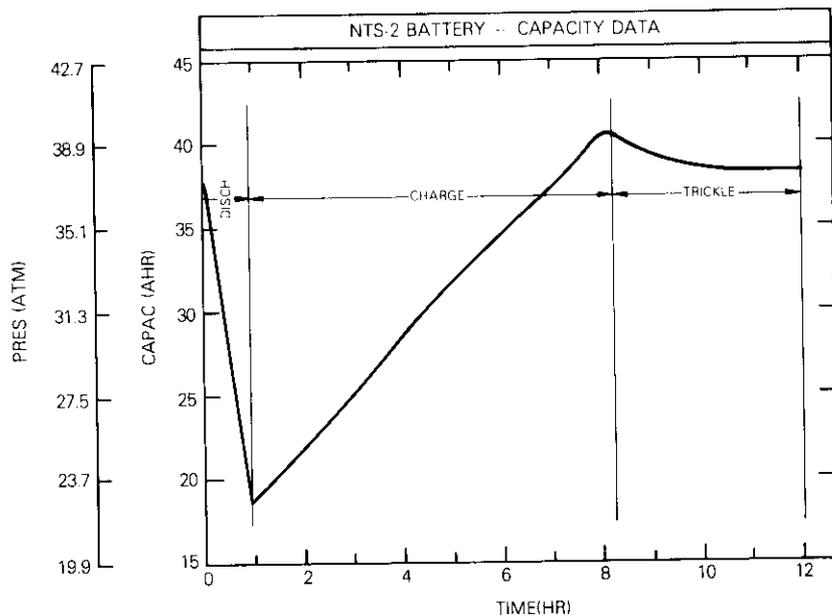


Figure 6. Battery Capacity and Pressure for a 12-Hour Cycle

cated that 19.5 Ah were removed during the eclipse period. Calculation of the capacity removed by integration of the current with time indicated that 20.7 Ah were removed, providing good agreement and verification of the strain gauge as an indicator of capacity. The battery state-of-charge at the beginning and end of the cycle is almost identical, indicating that the battery has been properly recharged.

Conclusion

A nickel-hydrogen battery was successfully built, tested, and placed into orbit. Neither the dynamic forces during lift-off, the steady load accompanying spinning, nor the 0-g environment have had any detrimental effects on battery performance. The automatic charge control system adequately recharges the battery while minimizing overcharge and internal heating of the battery. The strain gauges provide a reliable state-of-charge

indicator which is operationally useful and aids performance evaluation.

The successful flight demonstration of this new battery technology on the NTS-2 satellite paves the way for future use of the Ni-H₂ battery in operational communications satellites.

Acknowledgment

The authors would like to acknowledge the assistance of Mr. Fred Betz from NRL who provided the telemetry data and was responsible for the Ni-H₂ battery design and spacecraft implementation.

References

- [1] F. Betz, J. F. Stockel, and A. Gaudet, "Nickel-Hydrogen Storage Battery for Use on Navigation Technology Satellite-Two," *Proceedings of the 11th Intersociety Energy Conversion Engineering Conference*, September 12-17, 1976, Vol. 1, pp. 510-516.
- [2] G. Van Ommering and J. F. Stockel, "Characteristics of Nickel-Hydrogen Flight Cells," *Proceedings of the 27th Power Sources Symposium*, June 21-24, 1976, pp. 124-128.

Capacity and corresponding cell pressure data over the 12-hour charge-discharge cycle are shown in Figure 6. Data from the strain gauges indi-

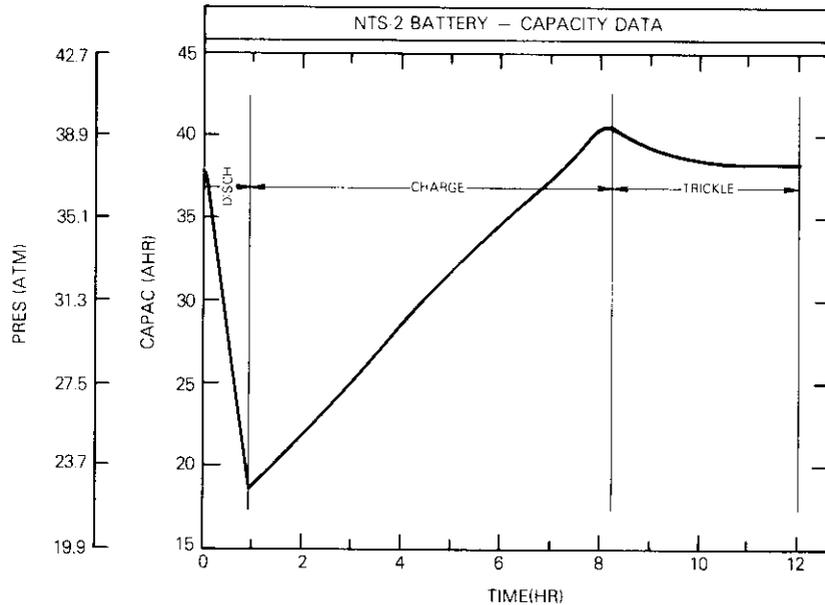


Figure 6. Battery Capacity and Pressure for a 12-Hour Cycle

cated that 19.5 Ah were removed during the eclipse period. Calculation of the capacity removed by integration of the current with time indicated that 20.7 Ah were removed, providing good agreement and verification of the strain gauge as an indicator of capacity. The battery state-of-charge at the beginning and end of the cycle is almost identical, indicating that the battery has been properly recharged.

Conclusion

A nickel-hydrogen battery was successfully built, tested, and placed into orbit. Neither the dynamic forces during lift-off, the steady load accompanying spinning, nor the 0-g environment have had any detrimental effects on battery performance. The automatic charge control system adequately recharges the battery while minimizing overcharge and internal heating of the battery. The strain gauges provide a reliable state-of-charge

indicator which is operationally useful and aids performance evaluation.

The successful flight demonstration of this new battery technology on the NTS-2 satellite paves the way for future use of the Ni-H_2 battery in operational communications satellites.

Acknowledgment

The authors would like to acknowledge the assistance of Mr. Fred Betz from NRL who provided the telemetry data and was responsible for the Ni-H_2 battery design and spacecraft implementation.

References

- [1] F. Betz, J. F. Stockel, and A. Gaudet, "Nickel-Hydrogen Storage Battery for Use on Navigation Technology Satellite-Two," *Proceedings of the 11th Intersociety Energy Conversion Engineering Conference*, September 12-17, 1976, Vol. 1, pp. 510-516.
- [2] G. Van Ommering and J. F. Stockel, "Characteristics of Nickel-Hydrogen Flight Cells," *Proceedings of the 27th Power Sources Symposium*, June 21-24, 1976, pp. 124-128.