

COMSAT

Technical Review

Volume 5 Number 2, Fall 1975

COMSAT TECHNICAL REVIEW

Volume 5 Number 2, Fall 1975

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

Editorial Board Pier L. Bargellini, Chairman
Robert D. Briskman
S. Browne
S. J. Campanella
William L. Cook
Denis J. Curtin
Jorge D. Fuenzalida
Joachim Kaiser
R. W. Kreutel
Terrence P. McGarty
H. J. Meyerhoff
Akos G. Revesz
Robert Strauss

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

The COMSAT TECHNICAL REVIEW is published twice a year by the Communications Satellite Corporation (COMSAT). Subscriptions including postage, \$5.00 U.S. per year; single copies, \$3.00 U.S. Bulk rates are available upon request. Remittances for subscriptions should be made payable to the Treasurer, COMSAT, and addressed to COMSAT TECHNICAL REVIEW, Room 5273, 950 L'Enfant Plaza, S. W., Washington, D.C. 20024, U.S.A.

v FOREWORD

- 211** A COMPARISON OF THE COMSAT VIOLET AND NON-REFLECTIVE SOLAR CELLS **J. F. Allison, R. Arndt AND A. Meulenberg**
- 225** ANALYSIS OF PSK SIGNAL POWER SPECTRUM SPREAD WITH A MARKOV CHAIN MODEL **C. Devieux**
- 253** RESULTS AND ANALYSIS OF WORLDWIDE ECHO CANCELLER FIELD TRIAL **H. Suyderhoud, S. Campanella AND M. Onufry**
- 275** STUDY AND SIMULATION OF A VARIABLE LENGTH CODE DPCM FOR THE LUMINANCE SIGNAL IN DITEC **I. Dinstein**
- 301** CYLINDRICAL 3-DIMENSIONAL MODEL OF A GUNN OSCILLATOR **V. Riginos AND I. Atohoun**
- 329** 28-GHZ VARACTOR CONTINUOUS PHASE MODULATOR **Y. Lee**
- 355** THE UET MULTITUBE TRANSMITTER **A. Standing**
- 367** POLARIZATION ISOLATION AND SPURIOUS SIGNAL REJECTION REQUIREMENTS FOR A DUAL-POLARIZED UNATTENDED EARTH TERMINAL **W. Sones AND M. Grossman**
- 387** CTR NOTES:
DESIGN OF A LOSSY WAVEGUIDE FILTER **M. Chen AND C. Mahle** 387
11.7- AND 20-GHZ RADIOMETRIC MEASUREMENTS AT A BRAZILIAN EQUATORIAL SITE **H. Meyerhoff** 399
- 413** TRANSLATIONS OF ABSTRACTS
FRENCH 413
SPANISH 418

Foreword

With this issue, the COMSAT TECHNICAL REVIEW marks its fifth year of publication. In the first issue of this journal in Fall 1971, a *Statement of Purpose* written by Joseph V. Charyk, President, Communications Satellite Corporation, set the following goals:

- To address a world-wide community of engineers, scientists and others whose activities are related to communications satellite technology.
- To provide a medium for the presentation of the various aspects, trends and applications of satellite communications.
- To promote the understanding and use of communications satellites, and to contribute to a broader application of their benefits to meet expanding communications requirements around the world.

Since November, 1971, when this journal was first published, the number of INTELSAT telephone circuits in use has increased almost three-fold, and a number of domestic satellite systems have been implemented. Several of these utilize leased INTELSAT satellite transponders, and more such systems are planned. An experimental broadcast satellite has been launched, and another is to be launched shortly. A first maritime satellite is in its final stage of pre-launch testing; and plans are underway for the implementation of an aeronautical satellite system.

In the context of this background and the criteria followed by the Editorial Board, it is believed that the scope and quality of the papers presented in the COMSAT TECHNICAL REVIEW have been commensurate with the above-stated goals. It is intended to maintain in the future the same high standards of technical content and presentation on satellite communications technology.

The Editorial Board
COMSAT TECHNICAL REVIEW

A comparison of the COMSAT Violet and Non-reflective solar cells*

J. F. ALLISON, R. A. ARNDT, AND A. MEULENBERG

(Manuscript received May 2, 1975)

Abstract

This paper presents a comparison of the properties of the Violet solar cell with those of the second generation improved cell designated the COMSAT Non-reflective cell. These properties include current-voltage characteristics, temperature coefficients, reflectivity vs wavelength, spectral response, output power vs 1-MeV electron fluence, output as a function of angle of light incidence, solar absorptance and thermal emissivity, and grid adhesion. The Violet cell represents an improvement of at least 30 percent in power output relative to conventional satellite cells, whereas the improvement in the Non-reflective cell exceeds 50 percent. In both cases this relative improvement is maintained under ultraviolet and penetrating space radiation.

Introduction

The COMSAT Violet solar cell announced in 1972 [1] proved the feasibility of fabricating a silicon cell which is much improved for space use. The response to the blue-violet portion of the solar spectrum was vastly improved by eliminating the dead layer and using an improved anti-reflection coating. A significant advance over the Violet cell, announced in 1974 [2], was the development of the COMSAT Non-reflective (CNR)

*A condensed version of this paper has been presented at the 10th Intersociety Energy Conversion and Engineering Conference, August 17-22, 1975, Delaware.

cell, primarily including a pyramidal surface structure which improved the optical coupling. The pyramidal structure also provided for higher generation rates of photocarriers in the proximity of the junction, resulting in improvement of the cell red response and the radiation hardness. The Non-reflective cell in a 2- × 2-cm size produces 82–86 mW under simulated AM0 illumination.

This paper will report on some of the properties and characteristics of the Violet and Non-reflective cells, including I-V curves, reflectivity as a function of wavelength, spectral response, output power after 1-MeV electron irradiation, output vs angle of incident light, thermal properties, and grid adhesion.

Background

The Violet solar cell research effort has shown that the short-wavelength response can be increased by reducing the junction depth and eliminating the diffusion-produced dead layer. It has also been found that the junction quality and curve fill factor can be improved by carefully controlling the diffusion conditions. The reduction in junction depth with its associated increase in sheet resistivity has necessitated the use of a fine contact grid geometry of 10 to 30 lines per centimeter. The increased short-wavelength response has resulted in the initiation of research on the properties of antireflection (AR) coatings. New antireflection coatings of tantalum pentoxide and niobium pentoxide with higher indices of refraction and freedom from short-wavelength absorption have been developed [3]. These developments combined to form the Violet cell now being manufactured by OCLI.* Recently, cells based on these and parallel developments but showing somewhat less improvement have been announced by several other groups [4]–[6].

The improved antireflection coatings of tantalum oxide and niobium oxide can provide near-perfect optical coupling of the light into a covered cell at any selected wavelength. Despite this near-perfect coupling at a single wavelength, a 10-percent loss in cell power results from reflection over the solar spectrum. This loss has been reduced significantly by the development of a surface completely covered with a multitude of pyramids of uniform shape and orientation. Typically the pyramids are two to four micrometers across the base, with a density of $\sim 10^8$ per square centimeter.

*Optical Coating Laboratories Incorporated, PhotoElectronics Group, 15251 E. Don Julian Way, Industry, California 91746.

The faces of the pyramids are $\langle 111 \rangle$ planes formed by an anisotropic etching process on the $\langle 100 \rangle$ face of a silicon slice. This surface acts similarly to an anechoic chamber in that it traps virtually all incident light by multiple reflection.

Optical properties

The properties of the non-reflective surface have been extensively investigated [7]. These results indicate that the reflectance from such a surface is R^2 , where R is the reflectance from a plane surface. A single-layer antireflection coating of tantalum or niobium oxide further reduces the reflection from the pyramidal surface. The dominant reflection loss remaining in a completed CNR cell is that due to the mismatch of the magnesium fluoride on the cover glass and to the silver contact fingers.

The reflection as a function of wavelength of typical Non-reflective, Violet, and conventional cells is shown in Figure 1. The improvements

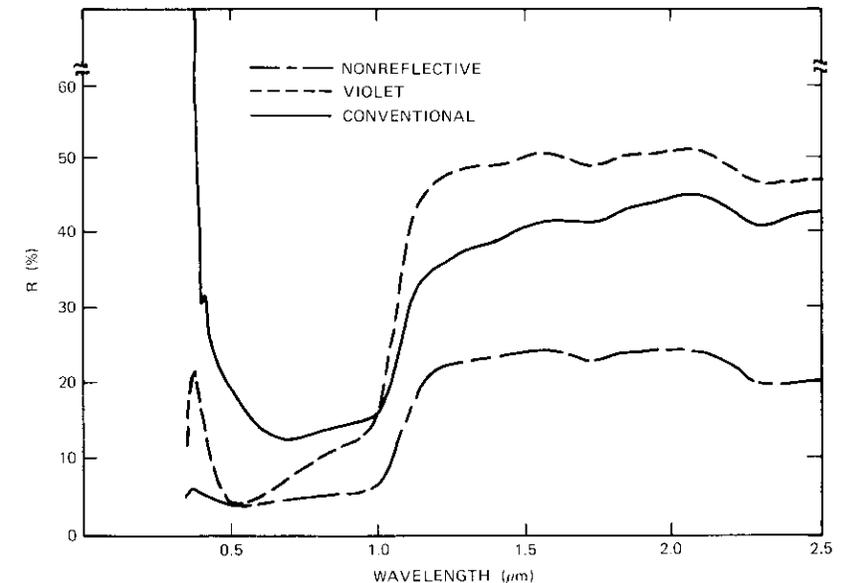


Figure 1. Reflectivity as a Function of Incident Light Wavelength for Non-reflective, Violet, and Conventional Cells Measured in an Integrating Sphere

in optical matching of the Violet cell with respect to the conventional satellite cell, and the CNR cell with respect to the Violet cell over the operating spectrum of 0.3 to 1.0 μm are readily apparent. The cells were covered with fused silica cover slides with blue filters such as those used for flight (0.41- μm cut-on filters for the conventional cells and 0.35- μm cut-on filters for the Violet and Non-reflective cells). The optical mismatch of the cover slide with its MgF_2 antireflective coating reflects 2-3 percent of the incident light. The remaining losses shown in Figure 1 are due to the silver contact fingers.

An experimental investigation of the ultraviolet degradation characteristics of the Ta_2O_5 antireflection coatings was conducted on a sample of covered and uncovered Violet cells. The primary considerations were to keep the cells under high vacuum* during the complete test and measurements, to expose the samples to a realistic solar ultraviolet source for a period long enough to show saturation effects (1,500 hr at one solar UV constant), and to provide control cells for comparison of the Violet cells with flight cells. The results indicated that Violet cells, using Ta_2O_5 and a 0.35- μm cut-off filter, degraded at about one-half the rate of flight-qualified conventional cells with SiO_x † AR coatings. At the end of the 1,500-hr test, the Violet cells, bare and covered, had degraded by 1.5 ± 0.5 percent and the conventional cells by 3 ± 1 percent. Linear extrapolation of the short-circuit current on a semi-log time plot indicated a maximum 5-year degradation, due to ultraviolet damage, of 2.5 ± 0.8 percent for the Violet cells, and 5 ± 2 percent for conventional cells.

Electrical characteristics

The electrical performance of typical laboratory-produced Violet and Non-reflective cells is compared to that of a conventional satellite cell in Figure 2. The 2- x 2-cm cells have been illuminated with simulated AM0 sunlight produced by the COMSAT Spectrolab X-25L solar simulator. The spectrum has been measured and adjusted to be as close as possible to

*To avoid any possibility of organic surface contamination of the system and cells, an oil-free pumping system was used to evacuate the chamber to the mid 10^{-7} torr range.

†The SiO_x notation indicates the delicate balance required between the use of SiO_2 , which is relatively insensitive to ultraviolet degradation but is not a good optical match for covered cells, and SiO (or more accurately $\text{SiO}_2 + \text{Si}$), which has a higher index of refraction (to improve the optical coupling) but also shows strong ultraviolet degradation characteristics.

the AM0 spectrum [8], [9]. In Figure 2, the maximum power output is 84 mW for the Non-reflective cell, 76 mW for the Violet cell, and 55 mW for the conventional cell. The corresponding efficiencies based on a total area of 4 cm^2 and 135.3-mW/ cm^2 solar intensity are 15.5 percent for the

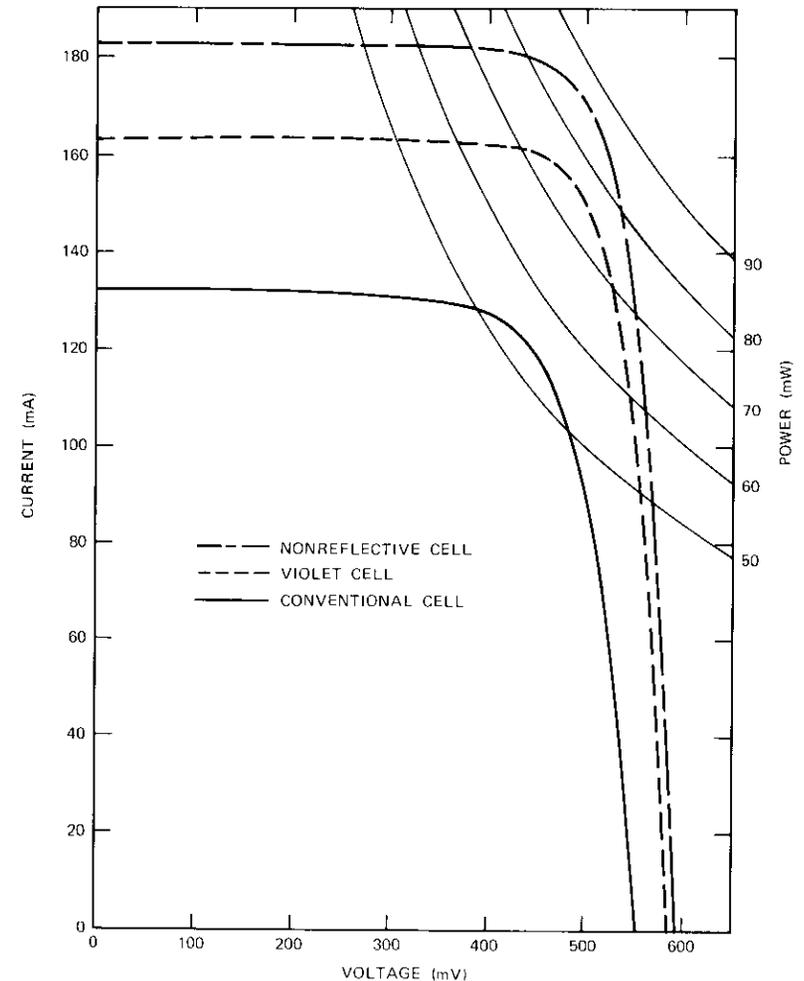


Figure 2. Current-Voltage Characteristics of Non-reflective, Violet, and Conventional Cells

Non-reflective cell, 14.0 percent for the Violet cell, and 10.1 percent for the conventional cell. Typical values for laboratory-produced Non-reflective and Violet cells range from 82 to 86 mW and 72 to 78 mW, respectively. The increased power output of the Non-reflective cell relative to the Violet cell is primarily a result of reduced reflection losses. Typical short-circuit currents are 182 mA, 162 mA, and 137 mA, respectively. The cells have been covered with fused silica with a $0.35\text{-}\mu\text{m}$ cut-on filter on the high-efficiency cells and a $0.41\text{-}\mu\text{m}$ filter on the conventional cell. (The filters are used to provide a compromise between prevention of ultraviolet damage to the adhesive and AR coating and loss of cell output.)

Some meaningful data in terms of engineering design on spinning satellites are shown in Figure 3, which compares the maximum power of covered Non-reflective, Violet, and conventional cells as a function of angle of incidence. Within the limits of the measurement, the maximum power and the short-circuit current follow the expected cosine law (for angles less than 70°) in all cells. The power and current improvements are therefore maintained over the angular variation.

Damaging space particulate radiation consists primarily of penetrating electrons and solar flare protons. The Violet cell has been tested extensively after electron, proton, and neutron irradiation [10]–[12]. The characteristics of the CNR cell after electron radiation have also been discussed [13]. Figure 4 compares the average data on five or more of each type of cell after 1-MeV electron irradiation for a fluence up to $2 \times 10^{15} \text{ e}^-/\text{cm}^2$ and shows that the relative performance increase of the advanced cells is maintained after irradiation. The curves for the CNR and the Violet cells are parallel, indicating a relative improvement for the Non-reflective cells over the Violet cell at increased radiation dose levels. This is due primarily to the refraction of light by the angular surface, which results in a higher carrier generation nearer the junction than would occur in the conventional cell.

Figure 5a is a plot of spectral response for the same three types of uncovered cells prior to irradiation. Note that the Non-reflective cell response has been improved over the entire solar spectrum. The spectral response [13] of the same cells after exposure to a 1-MeV electron fluence of $3 \times 10^{14} \text{ e}^-/\text{cm}^2$ and covering as above is shown in Figure 5b. The increased response in the long-wavelength end of a spectrum resulting from the optical refraction is maintained at a higher relative level even after irradiation. This sustained improvement is in marked contrast to the radiation-sensitive enhancement in red response commonly achieved by increased lifetime or use of p^+ back surface field contacts.

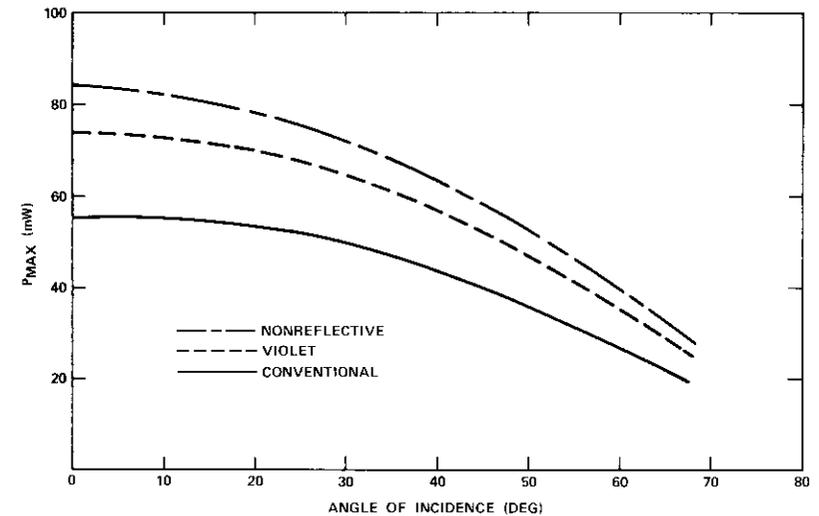


Figure 3. Maximum Power Output as a Function of Angle of Incidence for Non-reflective, Violet, and Conventional Cells

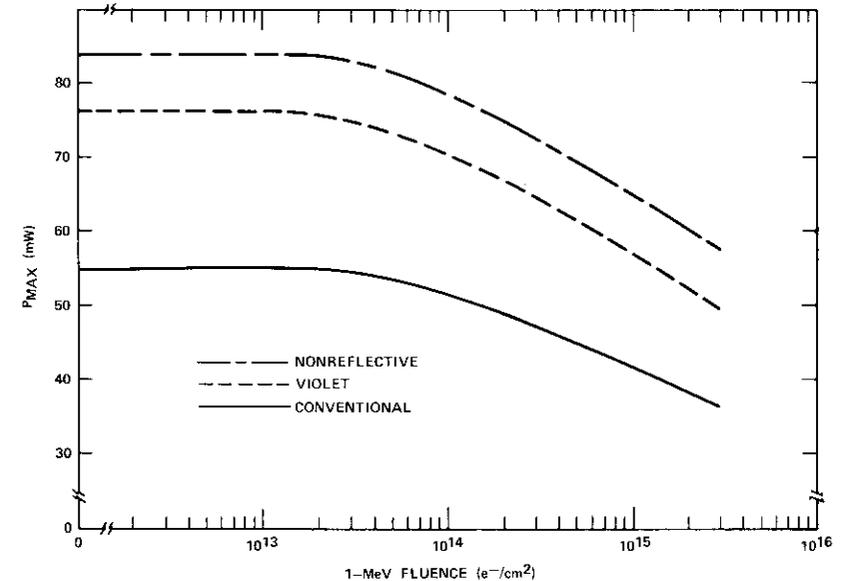


Figure 4. Maximum Power Output of Non-reflective, Violet, and Conventional Cells as a Function of 1-MeV Electron Irradiation Fluence

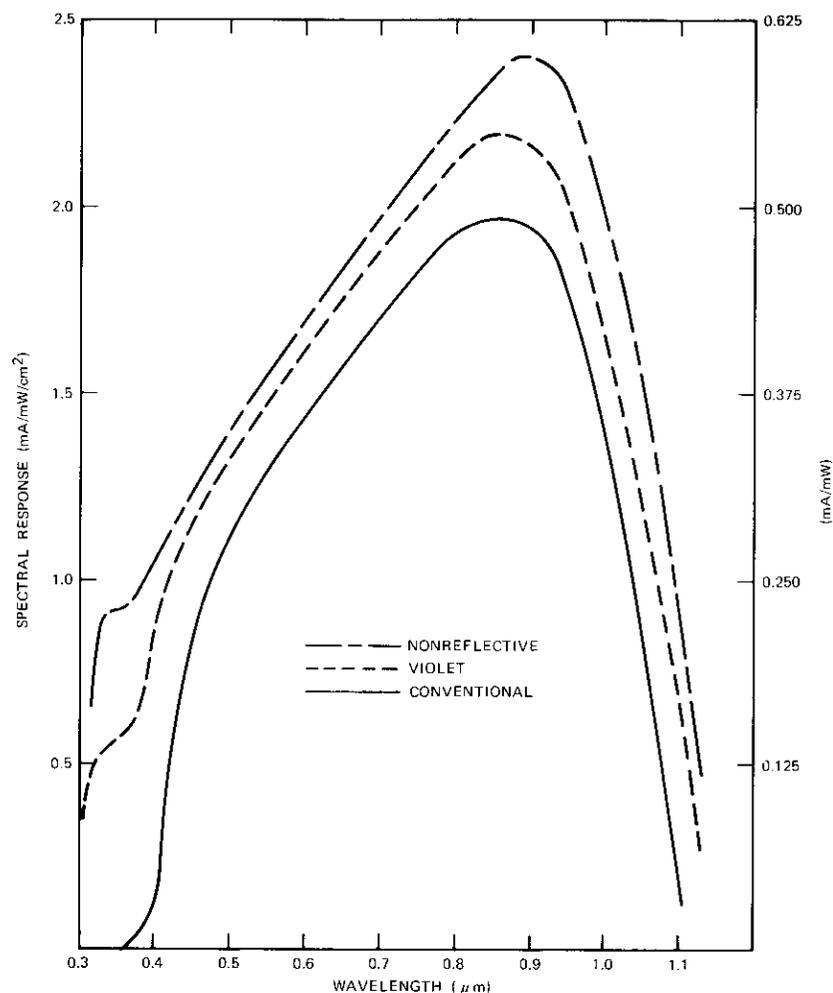


Figure 5a. Spectral Response of Non-reflective, Violet, and Conventional Cells

Thermal and mechanical properties

The I-V characteristics of a CNR cell at various temperatures are shown in Figure 6. The temperature coefficients calculated from these data and those obtained from similar cells are compared with those obtained for Violet and conventional cells in Table 1. The higher short-circuit current

temperature coefficient for the CNR cells is within the measured range for conventional cells [12].

The solar absorptance (α_s) and normal thermal emittance (ϵ_n) have been measured. The comparative results are listed in Table 2 along with the α_s/ϵ_n ratio. In actual operation, part of the absorbed solar power is

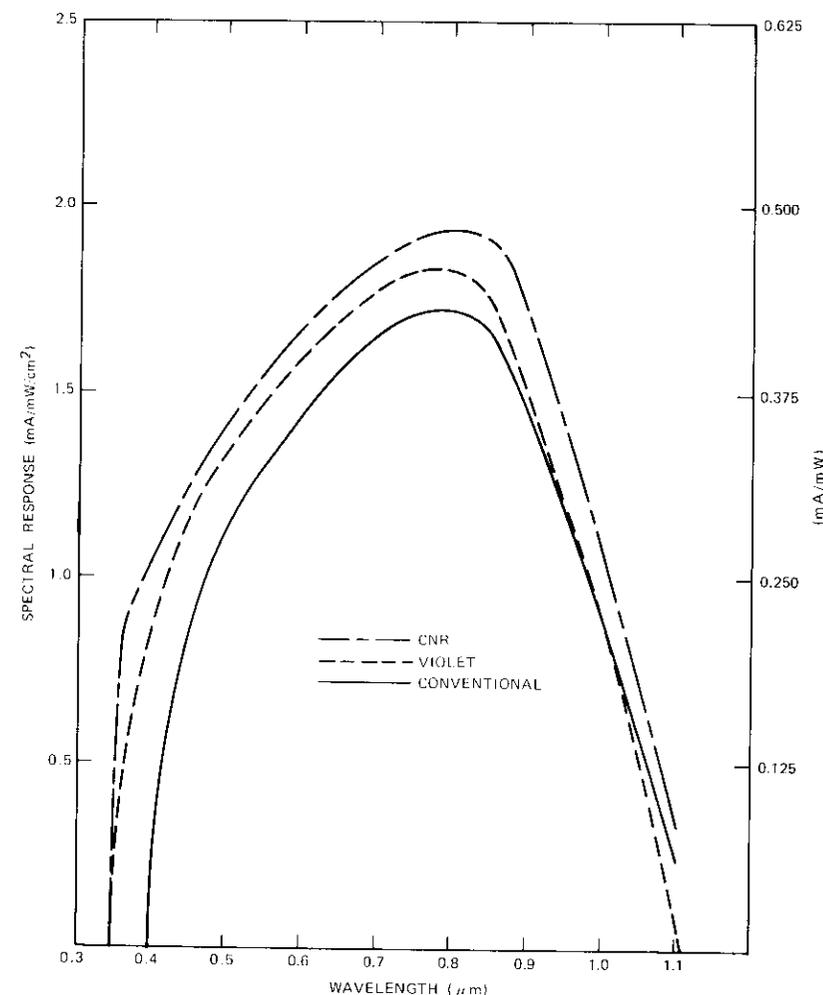


Figure 5b. Spectral Response of Non-reflective, Violet, and Conventional Cells after Exposure to $2 \times 10^{15} e^-/cm^2$ of 1-MeV Energy (also shows the effect of a UV filter on short wavelengths)

TABLE 1. AVERAGE TEMPERATURE COEFFICIENTS OF MAXIMUM POWER, CURRENT, AND VOLTAGE FOR CNR, VIOLET, AND CONVENTIONAL CELLS (BETWEEN 10°C AND 70°C)

	CNR (10 cells)	Violet (10 cells)	Conventional [12]
$\frac{1}{P_{\max(25^{\circ}\text{C})}} \frac{dP_{\max}}{dT} (\%/^{\circ}\text{C})$	-0.41 ^a	-0.43 ^a	-0.54 ^b
$\frac{1}{I_{sc(25^{\circ}\text{C})}} \frac{dI_{sc}}{dT} (\%/^{\circ}\text{C})$	0.07	0.05	0.05
$\frac{1}{V_{oc(25^{\circ}\text{C})}} \frac{dV_{oc}}{dT} (\%/^{\circ}\text{C})$	-0.37	-0.36	-0.41

^a After irradiation to 2×10^{15} 1-MeV electrons/cm², these values are still -0.43 ± 0.03 .

^b After irradiation to 2×10^{15} 1-MeV electrons/cm², this value drops to -0.46 ± 0.03 .

dissipated as electrical power away from the cells; hence $(\alpha_s - \eta)/\epsilon_n$, where η is the cell efficiency, is a more realistic ratio than α_s/ϵ_n for determining cell temperature. These ratios and cell efficiencies are also listed in Table 2.

The higher values of $(\alpha_s - \eta)/\epsilon_n$ for the Violet and CNR cells indicate that the improved collection efficiency of these cells results in a higher cell temperature. Use of cover slides with infrared reflection filters and/or greater emissivity (e.g., ceria doped microsheet with $\epsilon_n = 0.843$) would

TABLE 2. EFFICIENCY AND THERMAL PROPERTIES OF NON-REFLECTIVE, VIOLET, AND CONVENTIONAL CELLS*

Cell	Efficiency, η	Solar Absorp- tance, α_s	Normal Emittance, ϵ_n	α_s/ϵ_n	$(\alpha_s - \eta)/\epsilon_n$
Non-reflective	0.155	0.906	0.803	1.128	0.935
Violet	0.140	0.810	0.803	1.009	0.834
Conventional	0.101	0.725	0.803	0.903	0.777

* Non-reflective and Violet cells covered with fused silica 0.006 in. thick with 0.35- μm blue filter and MgF₂ coating; conventional cell covered with fused silica 0.012 in. thick with 0.41- μm blue filter and MgF₂ coating.

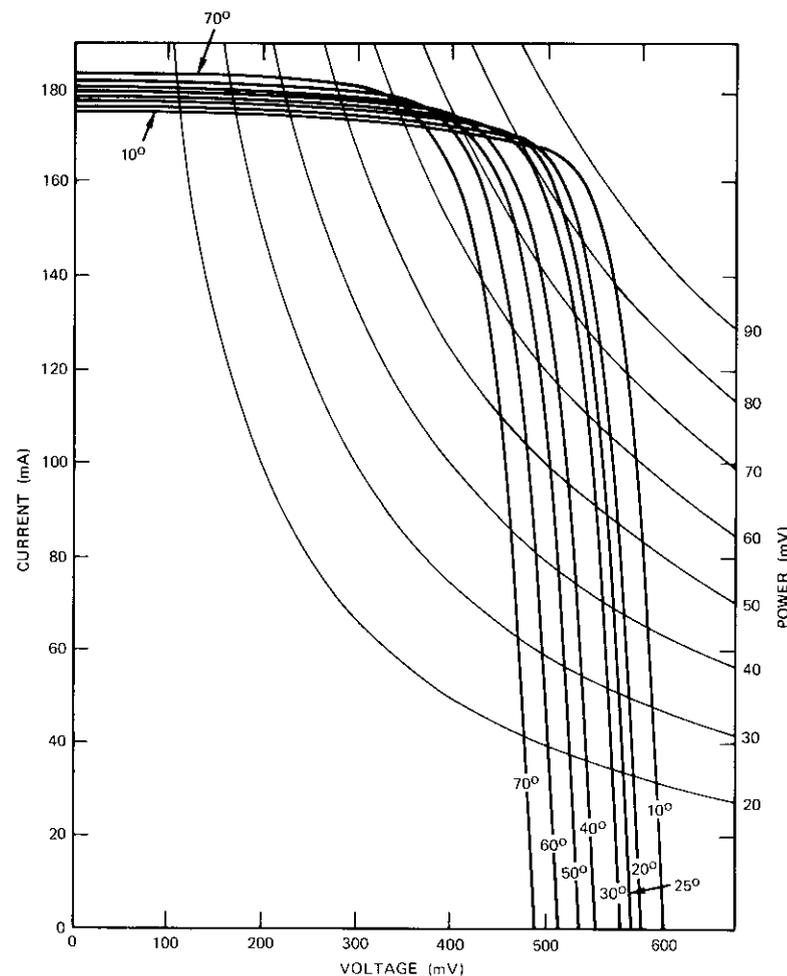


Figure 6. Current-Voltage Characteristics of the Non-reflective Cell at Temperatures from 10°C to 70°C

reduce or eliminate this penalty. The lower temperature coefficients of P_{\max} shown in Table 1 also tend to reduce the difference.

The increased surface roughness of the Non-reflective cell raised questions concerning the grid adhesion. Standard adhesive tape peel tests revealed no weakness. In addition, silver strips were thermocompression bonded directly to the grid lines and pull tests performed. Pull tests parallel

to the grid lines and along the cell surface proved that grids in the CNR cell are more than twice as firmly attached as identically processed and sized grids on Violet cells.* Pull tests normal to the cell showed that the CNR cell has even greater adhesion, since the roughened surface does not allow a normal angle to be formed with any more than one part of the surface at any time. These latter results were independent of strip bonding, either parallel or perpendicular to the grid lines.

Conclusion

A major change in the silicon solar cell junction depth and associated changes in grid geometry produced a cell with enhanced blue-violet response. This necessitated the development of a new antireflection coating with a higher index of refraction and less absorption. The combination of these innovations resulted in the Violet cell, with at least a 30-percent improvement over typical satellite cells. Subsequently, the development of a reproducible chemical etching technique produced a multi-tetrahedral non-reflecting surface. The combination of this processing with elements of the Violet cell technology provided the basis for the development of the Non-reflective cell, which produces at least 50 percent more power than the conventional satellite cell. In both types of cell the percent gains relative to conventional cells are not lost to ultraviolet or penetrating space radiation.

Acknowledgment

The authors are indebted to Dr. E. S. Rittner, whose encouragement and support of this program have been invaluable. Many members of the Applied Sciences Laboratory contributed to the measurements and participated in the discussion.

References

- [1] J. Lindmayer and J. F. Allison, "An Improved Silicon Solar Cell—The Violet Cell," *Proceedings of the 9th Photovoltaic Specialists Conference*, Silver Spring, Maryland, May 1972; *COMSAT Technical Review*, Vol. 3, No. 1, Spring 1973, pp. 1–22.

*Contact breakage in both types of cell during this test often resulted in rupture of the silicon rather than the grid bond.

- [2] J. G. Haynos et al., "The COMSAT Non-Reflective Silicon Solar Cell: A Second Generation Improved Cell," *Proceedings of the International Conference on Photovoltaic Power Generation*, Hamburg, Germany, September 25–27, 1974, pp. 487–500.
- [3] A. Revesz, "Vitreous Oxide Antireflective Films in High-efficiency Solar Cells," *Conference Record of the 10th IEEE Photovoltaic Specialists Conference*, Palo Alto, California, November 1973; *COMSAT Technical Review*, Vol. 3, No. 2, Fall 1973, pp. 449–452.
- [4] H. Fischer and W. Pschunder, "Improvements in Analyses and Technology of Silicon Solar Cells with Improved Efficiency," *Proceedings of the International Conference on Photovoltaic Power Generation*, Hamburg, Germany, September 25–27, 1974, pp. 69–84.
- [5] D. Amingual, P. Colardelle, and T. Nguyen Duy, "High Efficiency Silicon Solar Cells," *Proceedings of the International Conference on Photovoltaic Power Generation*, Hamburg, Germany, September 25–27, 1975, pp. 85–101.
- [6] E. L. Ralph and J. Scott-Monck, "Development and Space Qualification of New High Efficiency Silicon Solar Cells," *Proceedings of the International Conference on Photovoltaic Power Generation*, Hamburg, Germany, September 25–27, 1974, pp. 103–115.
- [7] R. A. Arndt et al., "Optical Properties of the Non-Reflective Cell," *Conference Record of the 11th IEEE Photovoltaic Specialists Conference*, Scottsdale, Arizona, May 6–8, 1975.
- [8] M. P. Thekaekara, "The Solar Constant and the Solar Spectrum Measured from a Research Aircraft at 38,000 ft," Goddard Space Flight Center Report No. X-322-68-304, August 1968.
- [9] "Solar Constants and Air Mass Zero Solar Spectral Irradiance," ATSM E490-73a.
- [10] R. A. Arndt, "Effects of Radiation on the Violet Solar Cell," *COMSAT Technical Review*, Vol. 4, No. 1, Spring 1974, pp. 41–52.
- [11] E. M. Gaddy, "Flight Qualification Test Results for Violet Cells," *Conference Record of the 10th IEEE Photovoltaic Specialists Conference*, Palo Alto, California, November 1973, pp. 153–162.
- [12] D. J. Curtin and R. W. Cool, "Qualification Testing of Laboratory Produced Violet Solar Cells," *Conference Record of the 10th IEEE Photovoltaic Specialists Conference*, Palo Alto, California, November 1973, pp. 139–152.
- [13] A. Meulenber et al., "Radiation Damage to the COMSAT Non-Reflective Cell," *Conference Record of the 11th IEEE Photovoltaic Specialists Conference*, Scottsdale, Arizona, May 6–8, 1975.



James F. Allison received a B.S.E.E. from Carnegie Institute of Technology (Carnegie Mellon University) in 1959 and an M.S.E.E. from Princeton University in 1962. At RCA Laboratories, David Sarnoff Research Center, for 10 years, he conducted research on thin films and early solid-state device development and received an RCA Laboratories award for outstanding achievement for the development of an integrated complementary symmetry silicon-on-sapphire transistor memory cell.

After joining COMSAT Laboratories in 1969, he built up a semiconductor technology research laboratory. Concerned with solid-state device development, he was instrumental in the development of the Violet and Non-reflective cell technology. He is presently Manager of the Semiconductor Technology Department.

Dr. Richard A. Arndt received his Ph.D. from Northwestern University in 1961. From 1961-1969 he was a staff scientist at the Brookhaven National Laboratory. Since he joined COMSAT Laboratories in 1969, he has been a member of the technical staff of the Applied Sciences Laboratory.



Andrew Meulenberg received his B.A. from the University of the South in 1962 and his Ph.D. in physics from Vanderbilt University in 1968. Since joining COMSAT in 1968 he has been involved in defining and simulating the space environment encountered by synchronous satellites, studying the effects of this environment on spacecraft components, and investigating material parameters and radiation defects in silicon solar cells. He is presently a member of the Semiconductor Technology Department.



Index: traveling wave tube amplifiers, Markov chains, phase shift keying, power spectra.

Analysis of PSK signal power spectrum spread with a Markov chain model

C. DEVIEUX

(Manuscript received February 20, 1975)

Abstract

A Markov chain model is used to calculate spectra of 2-phase PSK signals for linear channels and channels containing a memoryless traveling wave tube amplifier (TWTA) with AM/AM and AM/PM nonlinearities. The autocorrelation function is obtained from the Markov chain model, and a matrix representation which can be easily computerized is then obtained for the correlation function and the power spectrum. The model provides insight into the spectrum spreading action of the TWTA and particularly into the sensitivity of spectra to various system parameters such as AM/AM and AM/PM characteristics, data rate, and bandwidth. Extension of the model to cover M -phase PSK signals warrants further investigation.

Introduction

In bandwidth-limited communications systems it is desirable to maximize the transmission capacity in terms of voice or data channels per unit bandwidth. With PSK (phase shift keyed) modulated carriers, filtering produces a time-varying envelope and intersymbol interference. When this

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

signal is amplified in a traveling wave tube amplifier (TWTA), the amplifier nonlinearities produce a regeneration of spectrum sidebands which are caused directly by the time-varying envelope. Additional intersymbol interference is also incurred. The spectrum spreading produces adjacent channel interference and hence may cause transmission impairment in the neighboring channels. Consequently, it is important to derive a model describing the relationship between time-varying envelope statistics, TWTA gain and phase characteristics, and spectrum spreading. Moreover, the model should provide envelope statistics given the linear filtering characteristics preceding the TWTA and the bandwidth-to-baud-rate ratio (B/R).

This paper presents a Markov chain model for the envelope of filtered 2-phase PSK modulated carriers. After a review of pertinent background material, the derivation of the model is described. A formula expressing the power spectral density in matrix form is generated for linear channels, and then the matrix formulation for the spectrum density after the TWTA is presented. Results obtained from the computer calculations of this formula are given for a typical satellite communications channel and compared with a power spectrum measured in the laboratory with the INTELSAT IV simulator. The benefits of this model to the system designer are discussed and areas for future work are outlined.

Background

Figure 1 is a block diagram of a 2-phase PSK transmission chain through a communications satellite link. The data to be transmitted are randomized through a data scrambler before application to a 2-phase PSK modulator. Modem IF filter F_1 provides the proper waveform shaping. At the satellite the signal is filtered in filter F_2 and amplified through the TWTA. Filter F_3 suppresses out-of-band energy outside the allocated 40-MHz bandwidth. At the earth station receiver, noise is minimized by IF filter F_4 preceding the 2-phase PSK demodulator.

Derivation of power spectra from a Markov chain model is a common method used to obtain the baseband spectrum of complex multistate codes having certain characteristics such as a strong timing component [1]-[3]. Prabhu [4] has shown that the spectrum of m -phase PSK can be expressed as a compact Hermitian form. This paper starts from basic principles to derive a Markov chain model for 2-phase PSK signals transmitted through a linear channel. The method is extended to include the effects of the non-linear amplitude and phase characteristics of a TWTA.

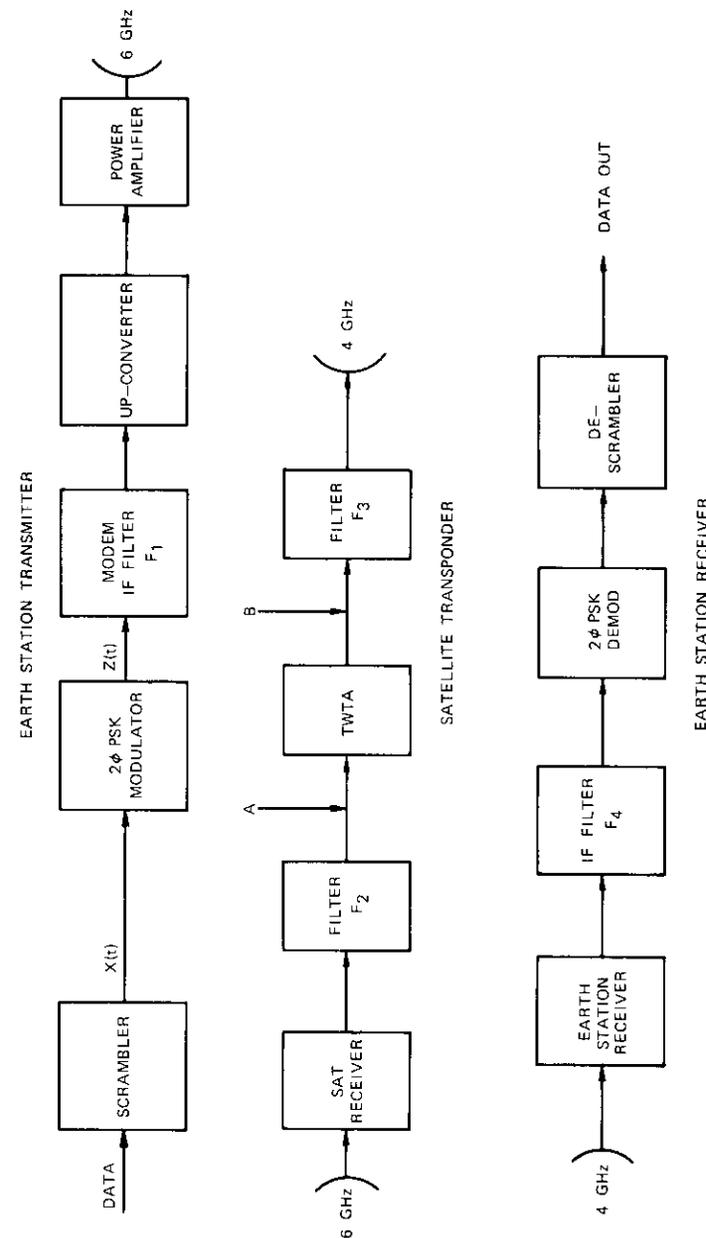


Figure 1. Transmitted Chain for 2-Phase PSK Modulated Carrier

At the modulator output a 2-phase PSK signal is expressed as

$$Z(t) = X(t) \cos \omega_c t \tag{1}$$

where the modulating signal $X(t)$ is a sequence of positive and negative square pulses,

$$X(t) = \sum_{k=-N}^N a_k P(t - kT) \tag{2}$$

$P(t)$ is the single pulse waveform, and the a_k 's (± 1) represent random bits of periodicity T .

$X(t)$ and $Z(t)$ are cyclostationary processes [5], whose power spectra $X(\omega)$ and $Z(\omega)$ are the Fourier transforms of the autocorrelation functions $R_x(\tau)$ and $R_z(\tau)$,

$$R_x(\tau) = E \left\{ \frac{1}{T} \int_{-T/2}^{T/2} X(t)X(t + \tau) dt \right\} \tag{3}$$

$$R_z(\tau) = E \left\{ \frac{1}{T} \int_{-T/2}^{T/2} Z(t)Z(t + \tau) dt \right\} \tag{4}$$

which are related as follows:

$$R_z(\tau) = \frac{1}{2} \cos \omega_c \tau R_x(\tau) \tag{5}$$

It can easily be shown that

$$R_x(\tau) = \frac{1}{T} \sum_{k=-N}^N \int_{(-T/2)+kT}^{(T/2)+kT} P(t - kT)P(t - kT + \tau) dt$$

and since

$$E \{ a_i a_j \} = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases}$$

it is possible to write

$$R_x(\tau) = \frac{1}{T} \int_{-\infty}^{\infty} P(t)P(t + \tau) dt \tag{6}$$

In other words, as is well known, for linear channels the correlation function of the process is equal to that of the single pulse $P(t)$. In the case of a rectangular pulse of amplitude B , the power spectrum of the process is

$$Z(\omega) = CT \frac{\sin^2 (\omega T/2)}{(\omega T/2)^2} \tag{7}$$

where $C = B^2/2$ is the average power of the unmodulated carrier. The power spectrum at the TWTA input is then

$$Z_A(\omega) = Z(\omega) |F_1(\omega)|^2 |F_2(\omega)|^2 \tag{8}$$

where $F_1(\omega)$ and $F_2(\omega)$ are the frequency responses of filters F_1 and F_2 , respectively.

At the TWTA output the expressions for linear channels are no longer applicable and a Markov chain model will be used to obtain the power spectrum. Envelope waveforms will be involved rather than the single basic pulse waveform $P(t)$.

Markov chain model

Consider, for example, the trapezoidal pulse $P(t)$ shown in Figure 2.

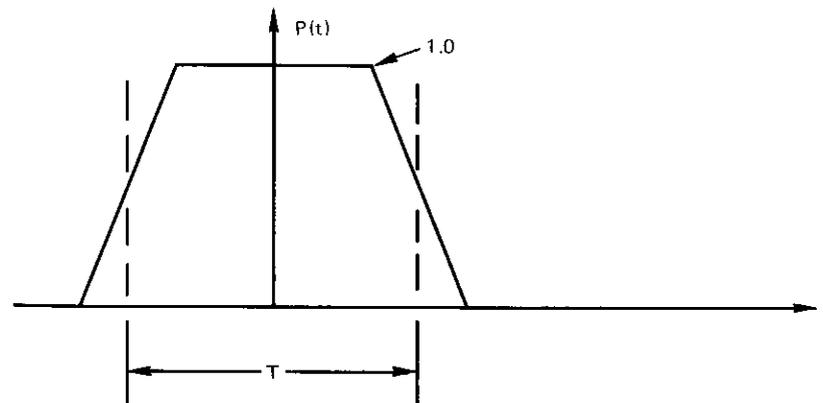


Figure 2. Trapezoidal Pulse $P(t)$

This pulse shape is a reasonable approximation of the pulse response of a typical satellite channel. The 2-phase PSK signal envelope for the trapezoidal pulses is shown in Figure 3 and four different envelope waveforms, denoted as $A_1(t)$, $A_2(t)$, $A_3(t)$, and $A_4(t)$, are shown in Figure 4.

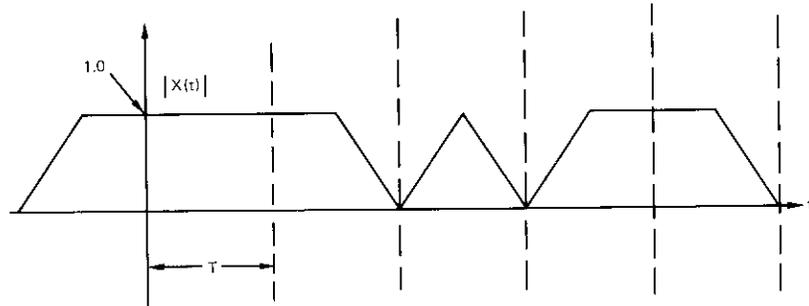


Figure 3. Typical 2-Phase PSK Envelope for Trapezoidal Pulse

The modulated carrier can be expressed as

$$Z(t) = \left[\sum_{-N}^N a_k P(t - kT) \right] \cos \omega_c t \quad (9)$$

Because pulse $P(t)$ is spread beyond the time interval T occupied by the a_0 bit, intersymbol interference is caused by the neighboring bits a_k ($k \neq 0$). It is assumed that the intersymbol interference covers $(2N + 1)$ bits (i.e., N bits toward the positive time axis and N bits toward the negative time axis). For the pulse of Figure 2, N is equal to 1 and

$$\begin{aligned} Z(t) &= \left[\sum_{-1}^1 a_k P(t - kT) \right] \cos \omega_c t \\ &= a_0 \left[\sum_{-1}^1 \left(\frac{a_k}{a_0} \right) P(t - kT) \right] \cos \omega_c t \end{aligned} \quad (10)$$

where

$$\begin{aligned} \frac{a_k}{a_0} > 0, & \quad a_k = a_0 \\ \frac{a_k}{a_0} < 0, & \quad a_k = -a_0 \end{aligned}$$

and

$$\left| \frac{a_k}{a_0} \right| = 1$$

The signal envelope $A(t)$ can then be defined as

$$A(t) \equiv \sum_{-1}^{+1} \frac{a_k}{a_0} P(t - kT) \quad (11)$$

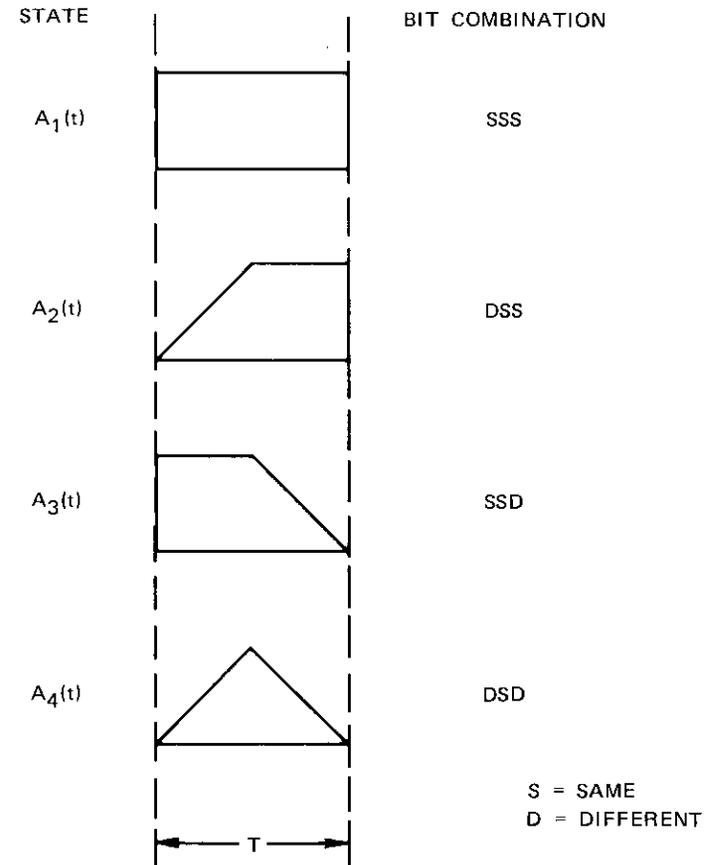


Figure 4. Envelope States and Bit Combinations Which Produce Them

The shape of $A(t)$ over the time interval $-T/2 < t < T/2$ depends upon the values of a_{-1} and a_1 , i.e., whether they are the same as or different from a_0 (see Figure 5). Four combinations are possible as indicated in Figure 4. (In the general case of intersymbol interference covering $2N + 1$ bits, there will be 2^N envelope shapes.)

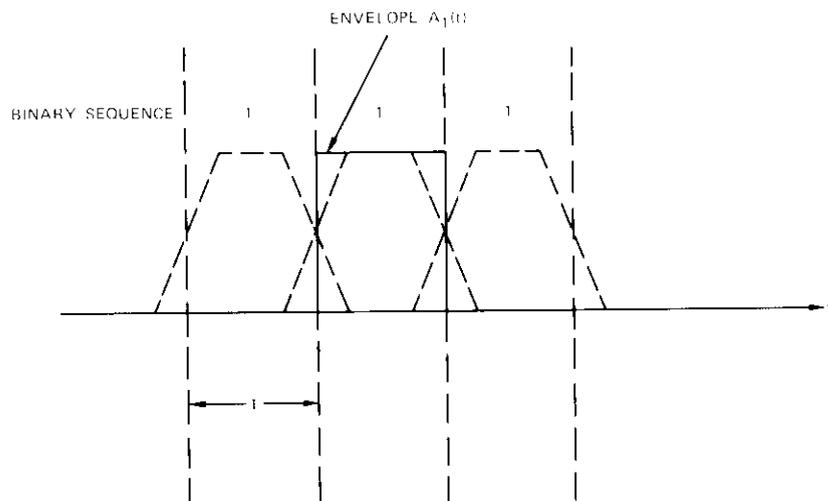


Figure 5. Formulation of Envelope Waveform from Neighboring Pulses

A computer program can easily generate the four envelope waveforms. These waveforms are not independent but are generated by the Markov chain model of Figure 6 with the transition probability matrix

$$P = \begin{matrix} & \begin{matrix} A_1 & A_2 & A_3 & A_4 \end{matrix} \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{matrix} & \begin{bmatrix} 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 \end{bmatrix} \end{matrix} \quad (12)$$

where P_{ij} is the probability of moving to state A_j starting at state A_i .

The model is obtained as follows: Assume a center bit a_0 at time $t = 0$. When the preceding and following bits, a_{-1} and a_{+1} , have the same polarity as a_0 , an SSS pattern will be produced corresponding to the envelope waveform $A_1(t)$. Consider next the envelope waveform which can occur at

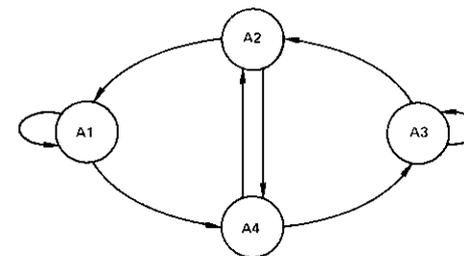


Figure 6. Markov Chain

time $t = \tau$. The triplet is now $(a_0 \ a_1 \ a_2)$, and since a_0 and a_1 have the same polarity, the pattern at time T will be SSx , where $x = S$ if a_2 and a_1 have the same polarity and $x = D$ otherwise. Therefore the two possible envelope waveforms are SSD and SSS , i.e., $A_3(t)$ and $A_1(t)$. The transition probability $P_{13} = P\{a_2 \neq a_1\} = 1/2$. Similarly, $P_{21} = (1 - P_{13}) = 1/2$. The transition probability $P_{14} = 0$ since it is impossible to pass from waveform $A_1(t)$ to waveform $A_4(t)$ in one step. The transition probabilities for other states are obtained in the same way.

Let π_i^n be the probability of being in state i after n timing clock shifts. Let π^n be the vector defined by the following column matrix:

$$\pi^n = \begin{bmatrix} \pi_1^n \\ \pi_2^n \\ \pi_3^n \\ \pi_4^n \end{bmatrix} \quad (13)$$

Then

$$\pi^{n+1} = P\pi^n$$

and

$$\pi^{n+2} = P^2\pi^n$$

where

$$P^2 = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix} \quad (14)$$

From any initial state the probability of arriving at any state after two clock shifts is independent of the initial state for the assumed pulse.

Derivation of the autocorrelation function from the Markov chain model

The autocorrelation, $R(\tau)$, of equation (4) can now be calculated from the envelope based on the Markov chain model:

$$R(\tau) = \frac{\cos \omega_c \tau}{2T} E \left\{ \int_{-T/2}^{T/2} [A_i(t) \cos \phi(t) \cdot A_j(t + \tau) \cos \phi(t + \tau)] dt \right\} \quad (15)$$

when frequency terms at $2\omega_c$ are eliminated. The $\cos \phi(t)$ term produces a sign change over the T interval [since $\cos \phi(t) = \pm 1$], but the envelope shapes A_i, A_j depend only on the polarity of the neighboring bits with respect to the center bit as explained previously.

Let $x(t) = A(t) \cos \phi(t)$. Figure 7 shows an envelope sequence $A(t)$ and the shifted envelope $A(t + \tau)$ together with the binary sequence. The time interval I , where $-T/2 < t < T/2$, will be considered since this is the interval over which the time correlation integral is obtained. Over that interval the envelope is one of the envelope shapes $A_i(t)$. Since they are equally likely the probability is $1/4$ that the initial state over this interval is $A_i(t)$. In addition to the envelope, it is necessary to know the phase, $\phi(t)$, over this interval. The two phase possibilities, $\phi_1 = 0$ and $\phi_2 = \pi$, corresponding to $a_0 = +1$ and $a_0 = -1$, respectively, are also equally

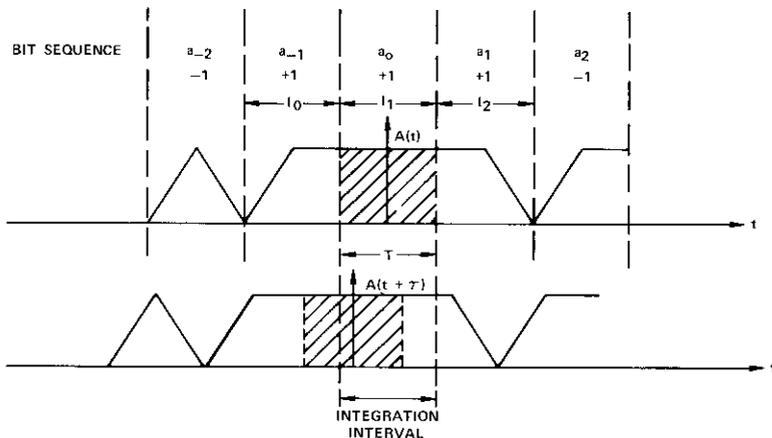


Figure 7. Integration Regions

likely; therefore, they occur over the initial interval, I_1 , with probability $1/2$, and $X(t)$ is initially determined completely by the vector (A_i, ϕ_i) . The probability of this initial state is then $1/8$ since events A_i and ϕ_i are independent over the initial interval.

Over the interval I_2 ($T/2 < t < 3T/2$), $x(t)$ is characterized by the vector (A_j^2, ϕ_j^2) , where the superscript 2 indicates an advance to the second clock interval, or from initial state S^1 to state S^2 . For a given initial state S^1 there are various possible states S^2 . For example, in Figure 7 the initial state is (A_1, ϕ_1) . At interval I_2 the state is (A_3^2, ϕ_1^1) . Another possibility over I_2 is (A_1, ϕ_1) . That is, once bits a_0 and a_1 are known, there are only two possible envelopes for interval I_2 since bit a_2 completely determines the envelope according to Figure 4. In this example the bits are $(1 \ 1 \ 1)$ and $(1 \ 1 \ -1)$, yielding possible envelope shapes A_1 (SSS) and A_3 (SSD). Then $P(S^2/S^1) = 1/2$.

To obtain the time correlation function first assume an initial state S^1 . The integral of $x(t)$ and $x(t + \tau)$ must be calculated for various allowable states. From Figure 7 the integration must clearly be broken into two regions. The integration is always carried over the time interval $-T/2 < t < T/2$. Observe the shifted envelope $A(t + \tau)$. Time interval I_a is denoted as

$$I_a = -\frac{T}{2} < t < \frac{T}{2} - \tau$$

and time interval I_b as

$$I_b = \frac{T}{2} - \tau < t < \frac{T}{2}$$

The integration interval, I , is the sum of the two disjoint intervals I_a and I_b :

$$I = I_a + I_b$$

Over I_a both $X(t)$ and $X(t + \tau)$ are in the same state, S^1 . Over I_b , $X(t)$ is in state S^1 and $X(t + \tau)$ is in state S^2 . Then,

$$\begin{aligned} L(\tau) &\equiv \int_I X(t)X(t + \tau) dt \\ &= \int_{I_a} X(t)X(t + \tau) dt + \int_{I_b} X(t)X(t + \tau) dt \end{aligned}$$

where $L(\tau)$ is a random variable. The ensemble averaging is to be performed to obtain $R(\tau)$, where

$$R(\tau) = \frac{\cos \omega_c \tau}{2T} E\{L(\tau)\} \quad (16)$$

Let

$$L_a(\tau) \equiv \int_{I_a}$$

and

$$L_b(\tau) \equiv \int_{I_b}$$

$$E\{L(\tau)\} = E\{L_a(\tau)\} + E\{L_b(\tau)\} \quad .$$

For any given initial state S^1 , one member of the ensemble of $L_a(\tau)$ is

$$\begin{aligned} L_{a_{ii}}(\tau) &= \int A_i(t) \cos \phi_i(t) A_i(t + \tau) \cos \phi_i(t + \tau) dt \\ &= \int A_i(t) A_i(t + \tau) dt \quad . \end{aligned}$$

Since $\cos^2 \phi_i(t) = 1$,

$$E\{L_a(\tau)\} = \sum_i P(A_i) L_{a_{ii}}(\tau)$$

where $P(A_i)$ is the probability of initially selecting A_i . It should be noted that the initial phase, ϕ_i , does not appear in this expression.

Since any one of the four states could have been selected randomly, $P(A_i) = 1/4$. Then

$$E\{L_a(\tau)\} = \sum_{i=1}^4 \frac{1}{4} L_{a_{ii}}(\tau) \quad .$$

In general,

$$\begin{aligned} E\{L_a(\tau)\} &= \sum_{i=1}^4 \sum_{l=1}^2 P(A_i, \phi_l) L_{a_{il}}(\tau) \\ &= \sum_i \sum_l P(A_i) P(\phi_l) L_{a_{il}}(\tau) \\ &= \frac{1}{4} \sum_i \sum_{l=1}^2 \frac{1}{2} P(A_i) L_{a_{il}}(\tau) \\ &= \frac{1}{4} \sum_i P(A_i) L_{a_{ii}}(\tau) \quad . \end{aligned}$$

Note that $E\{L_a(\tau)\} = 0$ for $\tau \geq T$ [since $L_{a_{ii}}(\tau) = 0$]. Similarly,

$$L_{b_{ijl}}(\tau) = \int A_i(t) \cos \phi_l(t) A_j(t' + \tau) \cos \phi_k(t' + \tau) dt$$

where $t' = t - T$. The subscript k can be omitted because, if A_i, ϕ_l , and A_j are known, ϕ_k is also known. Then,

$$E\{L_b(\tau)\} = \sum_i \sum_j \sum_l P(S^1) P(S^2/S^1) L_{b_{ijl}}$$

$$P(S^1) = P(A_i, \phi_l) = P(A_i) P(\phi_l) \quad .$$

Given state $S^2 = (A_j, \phi_k)$ the transition probability $P(S^2/S^1)$ to state $S^2 = (A_j, \phi_k)$ is the probability that ϕ_k is different from ϕ_l . However, this probability is simply the envelope transition probability, P_{ij} , given in equation (10); i.e.,

$$P(S^2/S^1) = P(A_j/A_i) = P_{ij} \quad .$$

Therefore,

$$\begin{aligned} E\{L_b(\tau)\} &= \sum_i \sum_j \sum_l P(A_i) P(\phi_l) P_{ij} \int A_i(t) A_j(t' + \tau) \\ &\quad \cdot \cos \phi_l(t) \cos \phi_k(t' + \tau) dt \end{aligned}$$

with $P(A_i) = 1/4$ and $P(\phi_l) = 1/2$.

Introducing the matrix α_{ij} defined as

$$\alpha_{ij} \equiv 2p_{ij} \cos \phi_l \cos \phi_k$$

makes it possible to write

$$E\{L_b(\tau)\} = \frac{1}{16} \sum_{i=1}^2 \sum_{ij} \int A_i(t) \alpha_{ij} A_j(t' + \tau) dt$$

where $t' = t - T$, and

$$E\{L_b(\tau)\} = \frac{1}{8} \sum_{ij} \int A_i(t) \alpha_{ij} A_j(t' + \tau) dt \quad .$$

Note that $E\{L_c(\tau)\} = 0$ for $\tau \geq 2T$, where

$$[\alpha] = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & -1 \end{bmatrix} .$$

For $\tau \geq T$ there is also a contribution $x(t + \tau)$ coming from interval I_3 , i.e., $3T/2 < t < 5T/2$, and from subsequent intervals. However, it can be shown that this contribution is equal to zero. Let $L_{cijl}(\tau)$ be the term corresponding to I_3 . Then,

$$L_{cijl}(\tau) = \int A_i(t) \cos \phi_l(t) A_k(t'' + \tau) \cos \phi_p(t'' + \tau) dt$$

where $t'' = t - 2T$. The contribution of this term is zero for $\tau < T$ since $A_k(t'' + \tau) \equiv 0$.

For $\tau \geq T$,

$$E\{L_c(\tau)\} = \sum_i \sum_j \sum_l P(S^1)P(S^3/S^1)L_{cijl}(\tau)$$

since the state over the interval I_3 is S^3 . However, state S^3 , given by (A_k, ϕ_p) , is independent of S^1 , given by (A_i, ϕ_l) . If it is assumed that (A_i, ϕ_l) is known, the envelope A_k is independent of A_i as shown by the transition matrix P^2 . Consequently,

$$P(S^3/S^1) = P(S^3) = P(A_k)P(\phi_p) = \frac{1}{8} .$$

Finally,

$$E\{L_c\} = \int E\{A_i\}E\{\cos \phi_l\}E\{A_k\}E\{\cos \phi_p\} dt = 0$$

since

$$E\{\cos \phi_l\} = E\{\cos \phi_p\} = 0 .$$

The same argument applies to all intervals I_k in which $k > 2$.

Summary of correlation function calculations

The correlation function $R(\tau)$ is obtained by calculating the contributions, R_{Ik} , of all intervals I_k , where $-N - 1 < K < N + 1$ and $2N + 1$

is the number of bits affected by intersymbol interference:

$$R(\tau) = \sum_{k=-N}^N R_{Ik} . \quad (17)$$

In the interval I_k , the generic term P_{Ik} must be calculated to obtain R_{Ik} :

$$P_{Ik} = \int_{-T/2}^{T/2} A^t \beta_k G_k dt$$

where A^t = transpose of column vector A

A = column vector of waveform $A_i(t)$

β_k = matrix obtained from the transition probability matrix.

For the case studied herein, that of four envelope states, $\beta_1 = 2I$ for I_1 , $\beta_2 = \alpha$ for I_2 , and $\beta_3 = 0$ for I_k ($k > 2$), where I is the identity matrix, and G is the column vector of time shifts of envelope state:

$$\left. \begin{array}{l} G_1 = \text{vector composed of } A_i(t + \tau) \text{ for } I_1 \\ G_2 = \text{vector composed of } A_i(t - T + \tau) \text{ for } I_2 \\ G_3 = \text{vector composed of } A_i(t - 2T + \tau) \text{ for } I_3 \end{array} \right\} \text{ for } \tau \geq 0$$

$$\left. \begin{array}{l} G_2 = \text{vector composed of } A_i(t + T + \tau) \text{ for } I_2 \\ G_3 = \text{vector composed of } A_i(t + 2T + \tau) \text{ for } I_3 \end{array} \right\} \text{ for } \tau \leq 0 .$$

(The analysis for $\tau < 0$ is the same as that given for $\tau \geq 0$ except that the adjacent intervals precede I_1 .) In summary,

$$R(\tau) = \frac{\cos \omega_c \tau}{2T} \left[\frac{1}{8} \int A^t \beta_1 G_1 dt + \frac{1}{8} \int A^t \beta_2 G_2 dt + \frac{1}{8} \int A^t \beta_2 G_{-2} dt \right] .$$

In general,

$$R(\tau) = \frac{\cos \omega_c \tau}{2T} \left\{ \frac{1}{8} \left[\int A^t \beta_1 G_1 dt + \sum_{k=2}^N \int A^t \beta_k G_k dt + \sum_{k=-2}^{-N} \int A^t \beta_{-k} G_k dt \right] \right\} . \quad (18)$$

Power spectrum for linear channels

The power spectrum $Z(\omega)$ is the Fourier transform of $R(\tau)$. The integral operation involves the generic term

$$M(\tau, l) = \int A_i \sum_j \beta_{ij} A_j(t + \tau + lT) dt$$

where $l = (k - 1)$, and as expected, terms $M(\tau, -l)$. Then,

$$\mathfrak{F} \left\{ \int A(t)B(t + \tau) dt \right\} = A(\omega)B^*(\omega)$$

where $A(\omega) = \mathfrak{F}\{A(t)\}$
 $B(\omega) = \mathfrak{F}\{B(t)\}$
 $B^*(\omega) = \text{complex conjugate of } B(\omega)$.

Moreover, let $\mathfrak{F}\{M(\tau, l)\} = M(\omega, l)$; then

$$M(\omega, l) = M(\omega)e^{j\omega lT}$$

$$\mathfrak{F}\{M(\tau, l) + M(\tau, -l)\} = 2Re\{M(\omega)e^{j\omega lT}\}$$

where $Re\{ \}$ = real part of the expression

$$\mathfrak{F}\{M(\tau)\} = M(\omega) = A_i(\omega) \sum_j \beta_{ij} A_j^*(\omega)$$

$$M(\omega, l) = \left[A_i(\omega) \sum_k \beta_{ij} A_j^*(\omega) \right] e^{j\omega lT}$$

$$\mathfrak{F} \left\{ \sum_{k=2}^N \int A^k \beta_k G_k dt + \sum_{k=-2}^{-N} \int A^k \beta_{-k} G_k dt \right\}$$

$$= 2Re \sum_{k=2}^N \sum_{i,l} \beta_{ij}^k [A_i(\omega) A_j^* e^{j\omega kT}]$$

and β_{ij}^k is an element of matrix β_k .

Let $R_L(\tau)$ be the low-pass term of $R(\tau)$; i.e.,

$$R(\tau) = \cos 2\omega_c \tau R_L(\tau) .$$

For the 4-state case, the spectrum of $R_L(\tau)$ is then

$$Z_L(\omega) = \frac{1}{2T} \left\{ \frac{1}{4} \sum_i |A_i(\omega)|^2 + \frac{1}{4} R_e \sum_{ik} \alpha_{ik} A_i(\omega) A_k^*(\omega) e^{j\omega T} \right\} \quad (19)$$

where $Re()$ indicates the real part of (). Finally,

$$Z(\omega) = \frac{1}{2} [Z_L(\omega - \omega_c) + Z_L(\omega + \omega_c)] .$$

TWTA nonlinear effects

With $Z(t)$ as the input of the memoryless TWTA, the output $y(t)$ will have a reshaped envelope and an additional phase term $\theta(t)$:

$$y(t) = g[A(t)] \cos [\omega_c t + \phi(t) + \theta(t)] \quad (20)$$

where $A(t)$ = input envelope

$\phi(t) = 0$, π information phase modulation

$g[A(t)]$ = reshaped envelope caused by TWT AM/AM characteristics

$\theta(t) = f(A)$ = phase term caused by TWT AM/PM characteristics.

Figures 8 and 9 show typical AM/AM and AM/PM curves.

In general the output can be rewritten as

$$y(t) = g(A) \cos [\phi(t) + \theta(t)] \cos \omega_c t - g(A) \sin [\phi(t) + \theta(t)] \sin \omega_c t . \quad (21)$$

For 2-phase PSK modulation,

$$y(t) = g(A) \cos \theta \cos \phi \cos \omega_c t - g(A) \sin \theta \cos \phi \sin \omega_c t . \quad (22)$$

Since $\sin \phi = 0$,

$$y(t) = a(t) \cos \omega_c t - b(t) \sin \omega_c t$$

where

$$a(t) = g(A) \cos \theta(t) \cos \phi(t)$$

$$b(t) = g(A) \sin \theta(t) \cos \phi(t) .$$

the correlation function is

$$R_y(\tau) = R_a(\tau) \cos \omega_c \tau + R_b(\tau) \cos \omega_c \tau + [R_{ba}(\tau) - R_{ab}(\tau)] \sin \omega_c \tau . \quad (23)$$

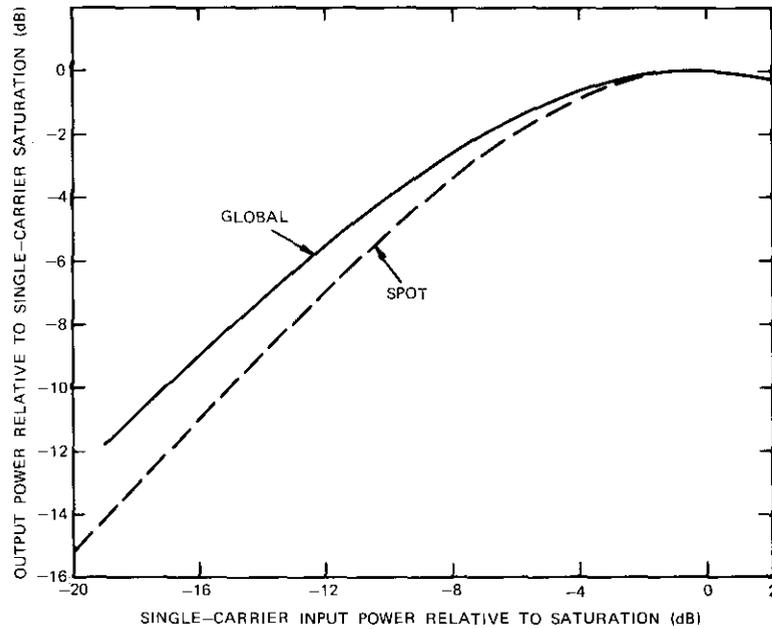


Figure 8. TWTA Relative Power vs Backoff

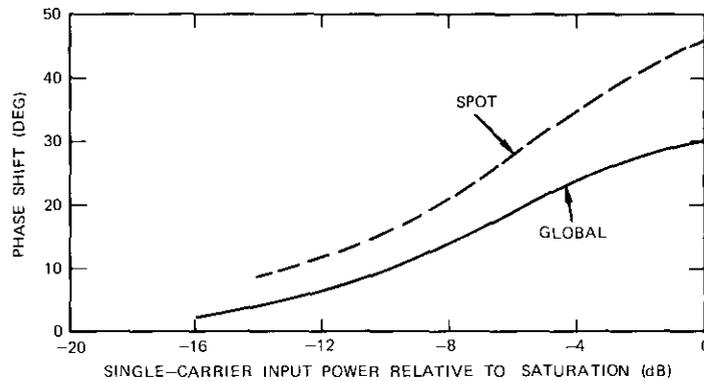


Figure 9. TWT Phase Shift vs Backoff

The power spectrum is obtained exactly as before except that in this case there are two sets of waveforms, namely,

$$C(t) = g(A) \cos \theta(t)$$

$$D(t) = g(A) \sin \theta(t) .$$

Analytical expression for the TWTA output spectrum

The total bandpass power spectrum is in general

$$P_T(\omega) = \frac{1}{2} \left[P_R(\omega + \omega_c) + P_R(\omega - \omega_c) + \frac{P_I(\omega - \omega_c) - P_I(\omega + \omega_c)}{j} \right] \quad (24)$$

where $P_R(\omega)$ is the Fourier transform of $[R_a(\tau) + R_b(\tau)]$ and $P_I(\omega)$ is the Fourier transform of $[R_{ba}(\tau) - R_{ab}(\tau)]$. From equation (17),

$$P_R(\omega) = \frac{1}{2T} \left\{ \frac{1}{4} \sum_i |C_i(\omega)|^2 + \frac{1}{4} Re \sum_{ik} \alpha_{ik} C_i(\omega) C_k^*(\omega) e^{j\omega T} + \frac{1}{4} \sum_i |D_i(\omega)|^2 + \frac{1}{4} Re \sum_{ik} \alpha_{ik} D_i(\omega) D_k^*(\omega) e^{j\omega T} \right\} \quad (25)$$

and $P_I(\omega)$ is obtained as follows. First the term $R_{ba}(\tau)$ yields the spectrum

$$\sum_i D_i(\omega) C_i^*(\omega) + Re \sum_{ik} \alpha_{ik} D_i(\omega) C_k^*(\omega) e^{j\omega T} .$$

Similarly, $R_{ab}(\tau)$ yields

$$\sum_i D_i^*(\omega) C_i(\omega) + Re \sum_{ik} \alpha_{ik} C_k(\omega) D_i^*(\omega) e^{-j\omega T} .$$

Consequently,

$$P_I(\omega) = j2Im \left[\sum_i D_i C_i^*(\omega) \right] \frac{1}{2T} \frac{1}{4} \quad (26)$$

where $I_m(\)$ indicates the imaginary part of ().

Results of computer calculations

Figure 10 shows the results of computer implementation of equations (25) and (26) for a bandwidth-to-baud-rate ratio (B/R) of about 1.3. The results are compared to the measured spectrum obtained through the

INTELSAT IV simulator with a TWTA having the characteristics of the global-beam TWTA shown in Figures 8 and 9. A more detailed discussion of filter characteristics is included in Appendix A. In the computer calculation the same TWTA characteristics are used. Both spectra are obtained at 0-dB backoff.

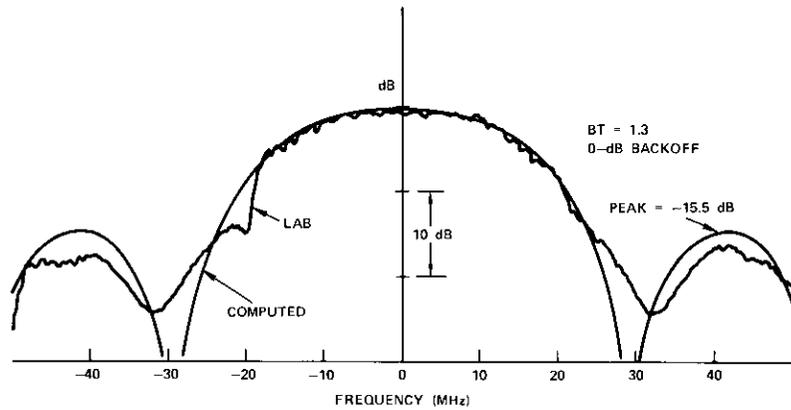


Figure 10. Comparison of Calculated and Measured 2-Phase PSK Spectrum at 0-dB TWTA Backoff

The calculated and measured spectra show good agreement within the measurement accuracy. The spectrum analyzer measurement accuracy around the regenerated sidelobes (about 16 dB down) is estimated to be about 1 dB and degrades at lower levels.

Figure 11 shows the basic single pulse response used in the calculations. The pulse is truncated at $\pm 1.5T$ for the generation of the four envelope shapes $A_i(t)$.

Figure 12 shows the calculated spectrum with the AM/PM characteristics removed. The peak sidelobe level decreases by only -0.3 dB. This confirms previously reported results [6] which indicate that the AM/PM contribution at 0-dB backoff is negligible.

Figures 13 through 16 show spectra for various BT (bandwidth \times baud duration) products from 1.0 to 5.0. The sidelobe level increases from -18.3 to -13.8 dB over this range. Note that the maximum level is -13.5 dB, which would be obtained for a square pulse (i.e., $BT \rightarrow \infty$).

The spectrum for $BT = 1.0$ is of interest. The sidelobe has a rather flat peak. Nyquist bandwidth filtering of the baseband pulses will produce this kind of spectrum.

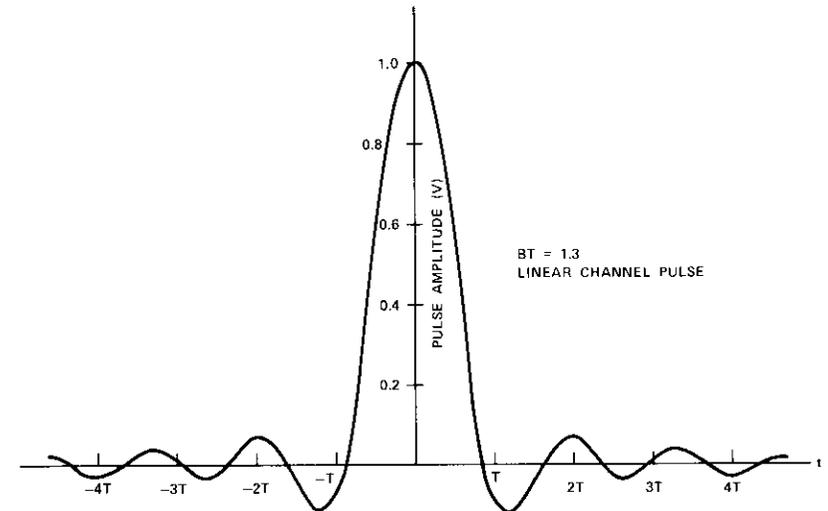


Figure 11. Single Pulse Response Used in Calculations

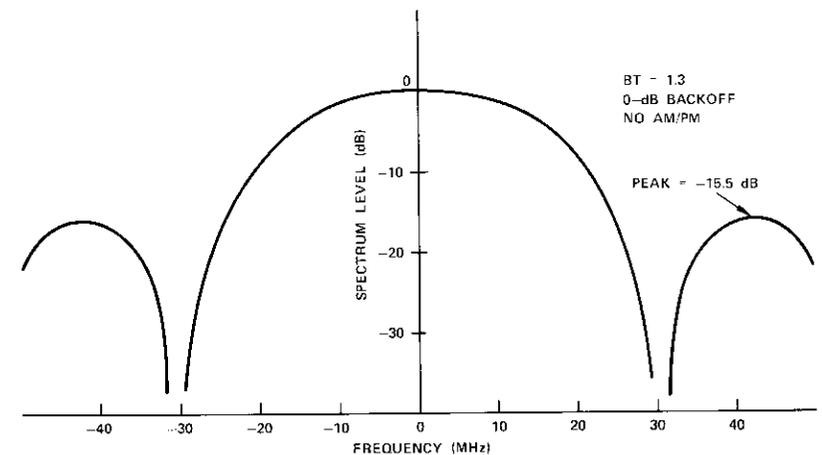


Figure 12. Calculated Power Spectrum with AM/PM Removed

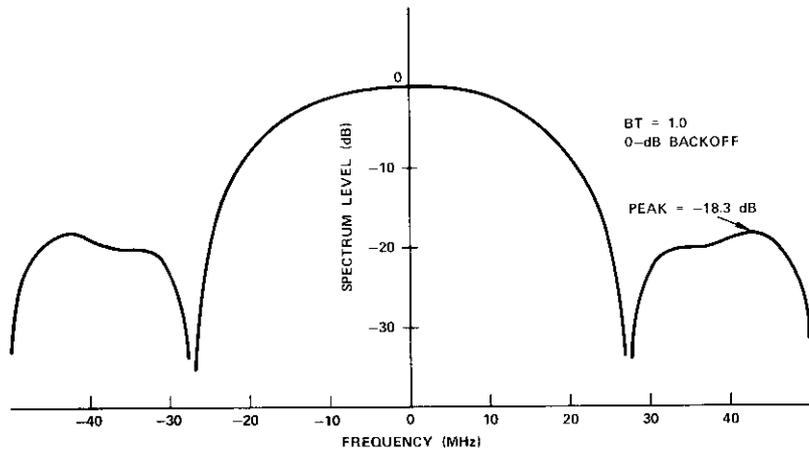


Figure 13. Calculated Power Spectrum for $BT = 1.0$

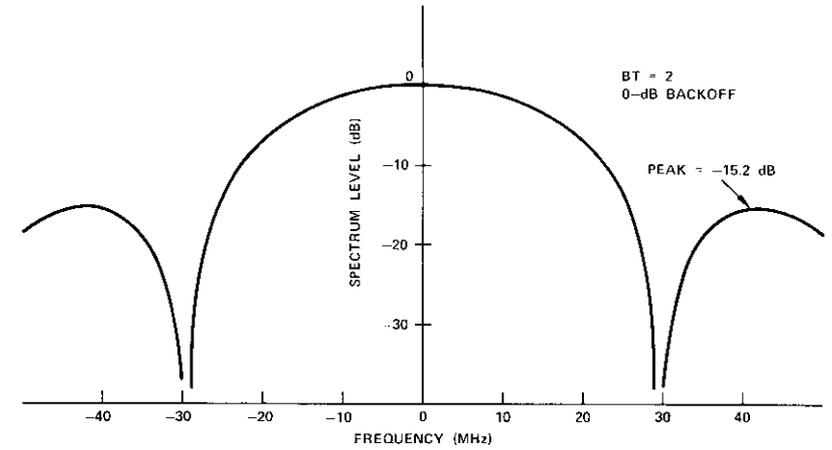


Figure 15. Calculated Power Spectrum for $BT = 2.0$

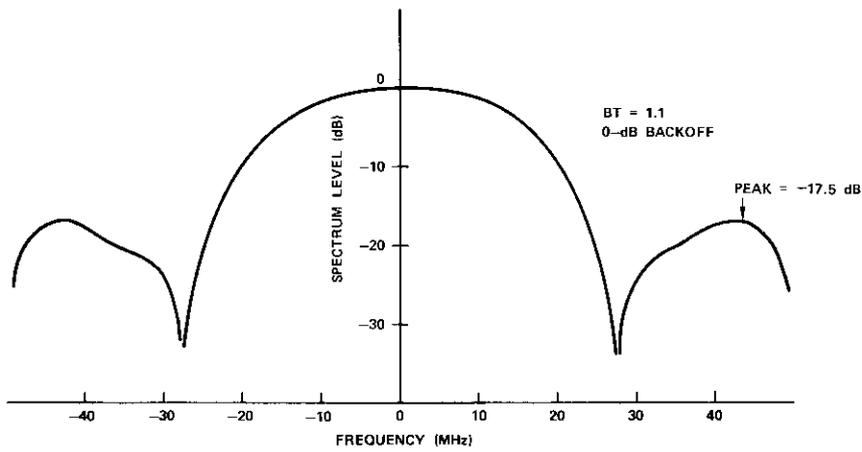


Figure 14. Calculated Power Spectrum for $BT = 1.1$

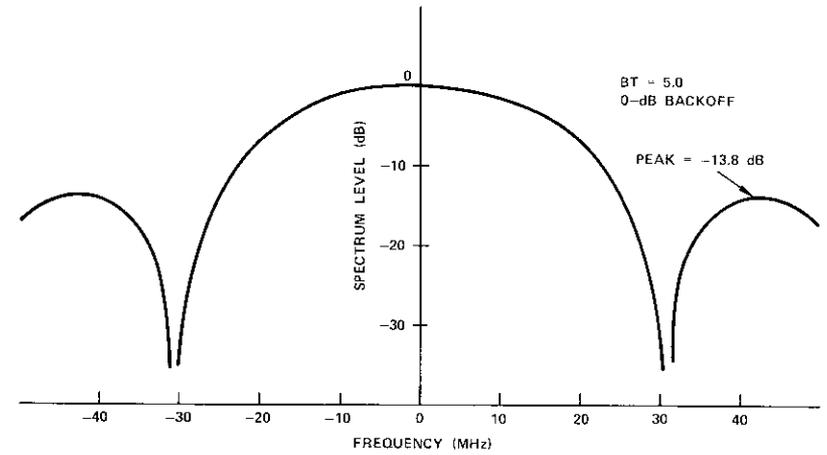


Figure 16. Calculated Power Spectrum for $BT = 5.0$

Figure 17 shows the variation of peak sidelobe levels over a practical range of BT products extending from 1 to 2.0. The spectrum peak is quite insensitive to BT in the range from 1.3 to 2.0. At BT = 1.0 the spectrum peak is about 3 dB lower than that at BT = 1.3.

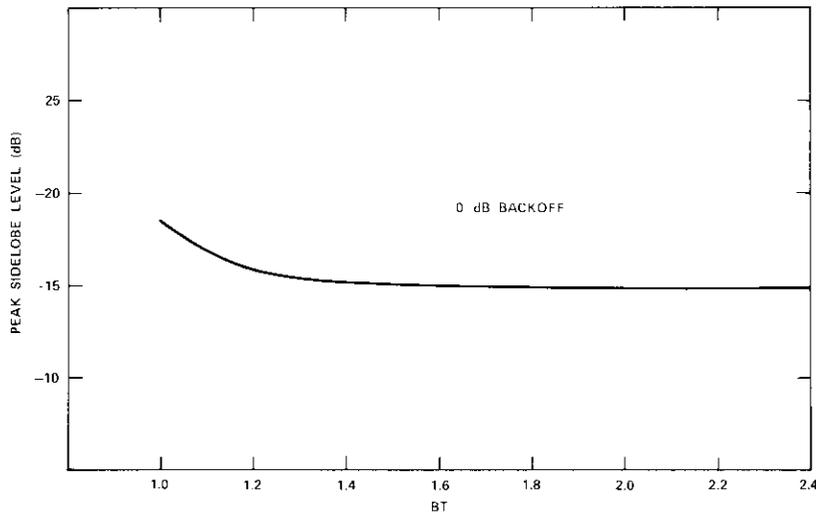


Figure 17. Variation of Peak Sidelobe Level with BT Product

Conclusions

Derivation of a Markov chain model for the envelope of 2-phase PSK signals has made it possible to calculate the correlation function and hence the power spectrum of these signals. The spectrum at the output of a TWTA was obtained through a new set of envelope shapes produced by the amplitude and phase characteristics of the TWTA. Comparison with laboratory measurements shows good agreement.

The matrix representation of the spectrum lends itself to computerization. The technique is easily extended to higher order matrices covering intersymbol interference from additional terms.

Extension of the model to 4- and M -phase PSK signals appears feasible once these signals are broken down into their quadrature components. The similarity between observed 4- and M -phase PSK signal spectra [6] and 2-phase PSK signal spectra makes such an investigation potentially interesting.

The Markov chain model is a useful tool for calculating power spectra. It constitutes a viable alternative to the other approaches for calculating the power spectrum at the output of memoryless nonlinear TWTAs. In addition, it permits observation of the direct relationship between envelope and phase distortion produced by the TWTA and the out-of-band spectrum regeneration. It also provides insight into the effect of data rate and bandwidth on the peak level of the out-of-band regenerated spectrum.

Acknowledgment

The author wishes to express his gratitude to S. J. Campanella and W. L. Cook for their helpful review and criticism of this paper.

References

- [1] M. Hecht and G. Guida, "Delay Modulation," *Proc. IEEE*, Vol. 57, No. 7, July 1969, pp. 1314-1316.
- [2] H. Yasuda and H. Inose, "Direct Calculation Method for Power Spectra of Pulse Sequences by Means of Transition Probability Matrices," *Electronics and Communications in Japan*, Vol. 53-A, No. 11, November 1970, pp. 33-41.
- [3] R. S. Bosik, "The Spectral Density of a Coded Digital Signal," *Bell System Technical Journal*, Vol. 51, No. 4, April 1972, pp. 921-932.
- [4] V. K. Prabhu and H. E. Rowe, "Spectra of Digital Phase Modulation by Matrix Methods," *Bell System Technical Journal*, Vol. 53, No. 5, May-June 1974, pp. 899-935.
- [5] W. R. Bennett, "Statistics of Regenerative Digital Transmission," *Bell System Technical Journal*, Vol. 37, No. 6, November 1958, pp. 1501-1542.
- [6] G. Robinson, O. Shimbo, and R. Fang, "PSK Signal Power Spectrum Spread Produced by Memoryless Nonlinear TWTs," *COMSAT Technical Review*, Vol. 3, No. 2, Fall 1973, pp. 227-257.

Appendix A. Laboratory measurement of 2-phase PSK power spectrum

Figure A-1 shows the amplitude response and Figure A-2 shows the group-delay response of the linear filtering preceding the TWTA, together with the approximations used for the computer calculation in the laboratory measurement test setup. This approximation has been used solely to simplify the programming. The Markov chain model is applicable to any type of filtering.

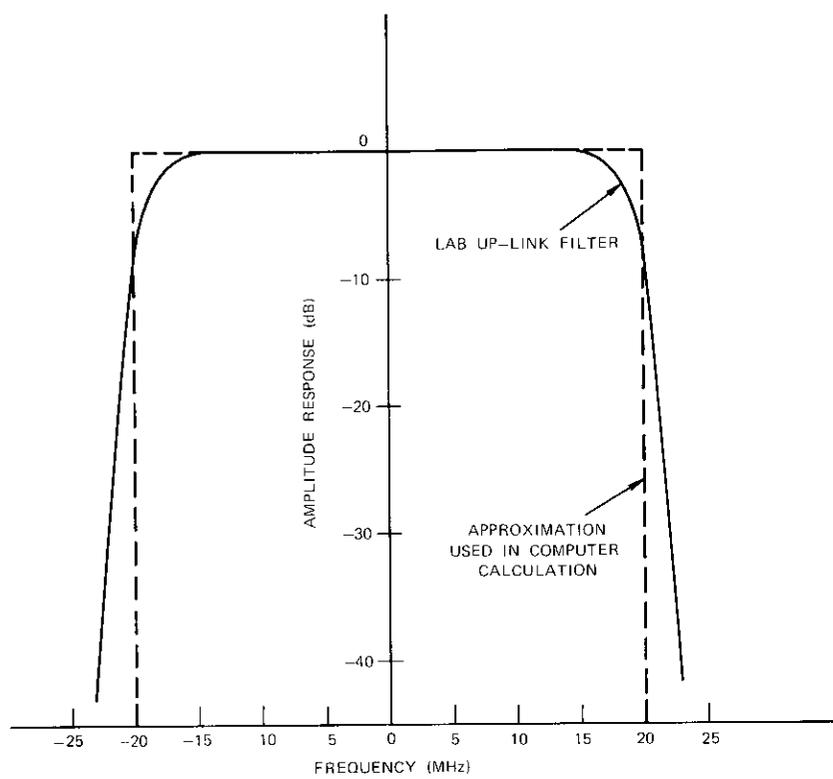


Figure A-1. Amplitude Response for Linear Filtering Preceding the TWTA

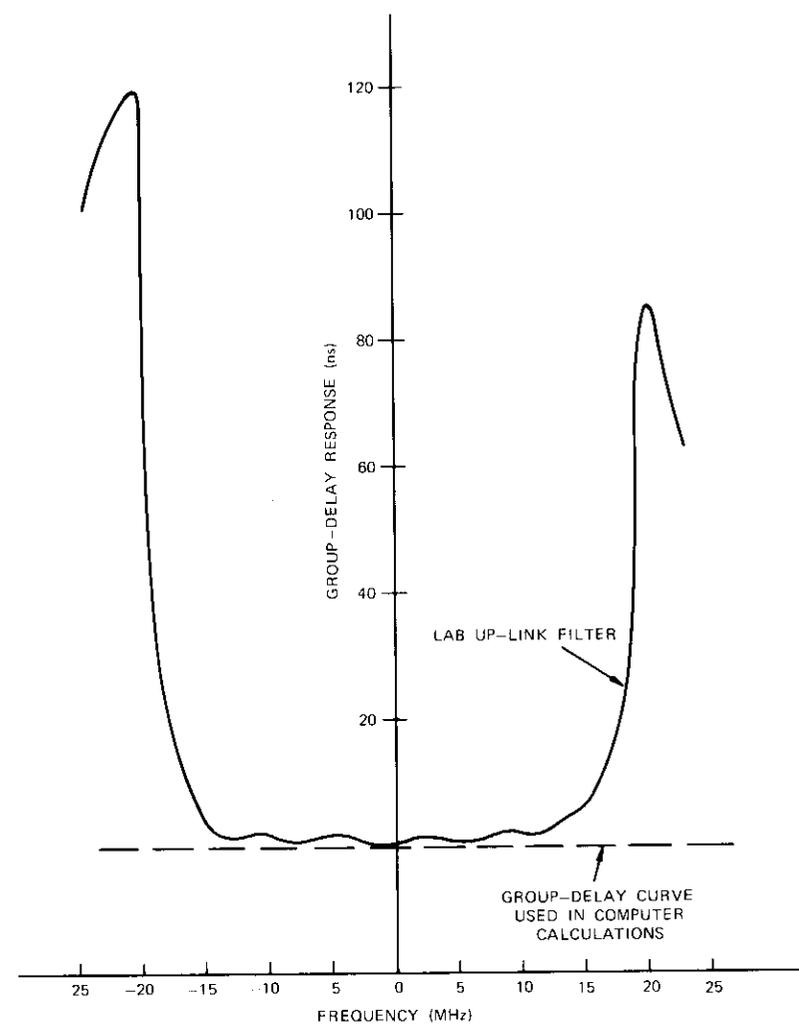


Figure A-2. Group-Delay Response for Linear Filtering Preceding the TWTA



Carrie Devieux, Jr., received a B.S.E.E. from the University of Montreal, an M.S.E.E. from McGill University, and a Ph.D. from the Polytechnic Institute of Brooklyn.

Prior to joining COMSAT Laboratories in 1972, he worked on the design of meteorological satellites at RCA and communications and weather satellite systems at Computer Sciences Corporation in Falls Church, Virginia. As a member of the technical staff in the Communications Systems Analysis Department of the Transmission Systems Laboratory, he is engaged in the investigation of the transmission performance of future satellite systems with multiple frequency reuses operating in the TDMA mode.

Dr. Devieux is a member of Sigma Xi and an active member of the Information Theory Group of IEEE.

Index: echo suppression, echo cancellation, voice communications, subjective quality assessment, international telephone circuits, telephone subscriber call-back interviewing

Results and analysis of worldwide echo canceller field trial

H. G. SUYDERHOUD, S. J. CAMPANELLA, AND M. ONUFRY

(Manuscript received April 23, 1975)

Abstract

Four digital adaptive echo cancellers using impulse response convolutional processing have undergone extensive field tests in the INTELSAT network. This paper provides a description of the echo cancellers involved in the test, followed by a detailed account of the 4-phase trial, which included seven member countries of INTELSAT.

Each phase of the trial is described in terms of echo canceller performance based on objective results obtained from measured performance parameters, and subjective results obtained from call-back interviews. Results and analysis show that echo cancellation techniques can be successfully applied to satellite telephone communications. In many cases echo cancellers have demonstrated a significant improvement in quality relative to echo suppressors.

Introduction

In the early sixties echo cancellation emerged as a most promising although complicated technique for overcoming problems with existing

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

echo control caused by echo suppressors which have inherent performance limitations [1]-[3]. In 1968, COMSAT Laboratories under INTELSAT sponsorship began the development of experimental echo cancellers for voice telephone circuits. The problems with existing echo control techniques on such circuits had been aired extensively and were known to be amplified on circuits with delay, such as satellite circuits [4]-[7].

At the end of 1972, four experimental echo cancellers became available for field testing on customer circuits of the INTELSAT network. Two of the echo cancellers were designed and built by Nippon Electric Company (NEC) of Japan as a result of an INTELSAT contract granted in 1970. Two more echo cancellers were designed and built by COMSAT Laboratories under corporate sponsorship. Also in 1972, INTELSAT member administrations were invited to participate in field tests of the echo cancellers. Seven countries, including the U.S., agreed to participate in these tests and were organized by COMSAT in conjunction with the U.K., France, West Germany, Brazil, Japan, Australia, and AT&T and the Hawaiian Telephone Company in the U.S.

During each test phase, member countries, usually operating in pairs, were requested to make satellite circuits available for installation of the echo cancellers, one at each end of the circuit. Performance was evaluated in two ways:

- a. by observing the rate at which cancellers caused echo to converge to a minimum, the effects of time-variant echo paths when they existed, the effects of double talk, and the echo return loss enhancement;
- b. by conducting organized call-back interviews in which customers who had just completed a call over the circuits with echo cancellers were asked to assess the call through a sequence of standard questions often used in telephone quality ratings [5].

As a baseline for comparing customer assessments, the echo cancellers were randomly interchanged with echo suppressors on circuits known only to the test coordinator of each administration. In one phase of the program, a cable circuit equipped with echo suppressors was included as a third test condition and results were also compared.

This paper will briefly describe the designs of the two echo cancellers and subsequently present the results and analysis of the entire field trial effort.

Echo cancellation vs echo suppression

It is important to stress the essential differences between echo cancellation and echo suppression techniques in telephone communications. An international telephone circuit between international switching centers (CTs) can be represented by two separate channels, one in each direction of transmission. At each end, beyond the CT in the national network, this mode of 4-wire transmission is converted to 2-wire transmission via a hybrid coil circuit before reaching the telephone. This conversion circuit is the source of the echo, since a fraction of the signal traveling on the A-to-B path returns on the B-to-A path (see Figure 1).

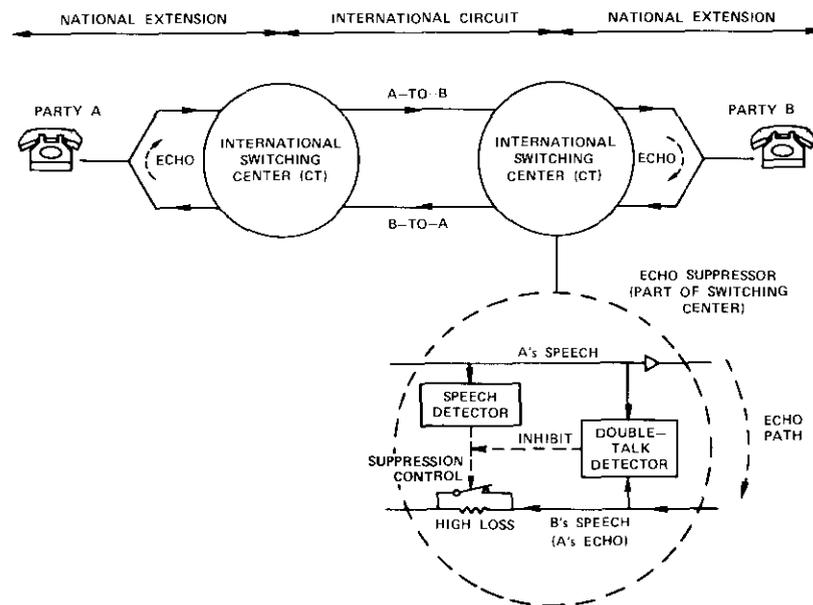


Figure 1. *International Telephone Circuit Including the Echo Suppressor*

An echo suppressor in each CT in the 4-wire circuit employs a speech-operated switched loss circuit to prevent B-to-A transmission when A talks.* However, if B talks at the same time and B's speech power becomes

* For the echo suppressor near A, interchange A and B in the discussion.

equal to or greater than that of A, the switched loss circuit is removed as a result of comparison of the two signals in a differential circuit. Upon removal of the loss, A's echo is permitted to return and become mixed or interspersed with B's speech. This situation may also result in severe chopping of syllables when B's speech power is less than that of A, and loud echo bursts can occur after the loss is removed.

On the other hand, an echo canceller is designed to prevent A's echo while simultaneously allowing B's speech, but at the expense of increased complexity. A replica of the echo is created in the canceller by adaptively modeling the transmission characteristics of the echo path. The replica of the echo is subtracted from the real echo in a difference circuit which is inserted in the B-to-A path (see Figure 2). The same difference circuit

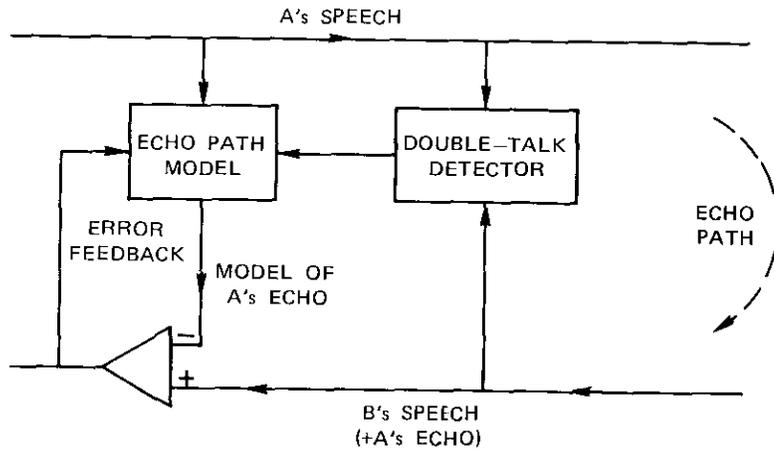


Figure 2. Echo Canceller Principle

allows B's speech to flow without interruption; hence the conflict existing in an echo suppressor is entirely eliminated [1].

Echo canceller designs

Two designs were implemented for the trial; one resulted from an INTELSAT contract with NEC (Japan), and one resulted from an in-house development at COMSAT Laboratories. Both are adaptive designs using time-domain digital convolution processing [8]-[9]. The primary differ-

ence between the designs, described in the following, is the method used to detect double talk. Upon detecting double talk, the COMSAT echo canceller disables the self-adaptive loop; both cancellers "freeze" the most recent echo path impulse response model, which determines the amount of echo return loss enhancement.

Figure 3 is a block diagram of the COMSAT echo canceller. It consists of two storage registers, X and H. The former stores and moves a 32.8-ms window of digitized samples of the speech signals received from the talker at the distant end. The second register, H, builds up and stores a replica of the echo path impulse response on command from the adaptive update circuit.

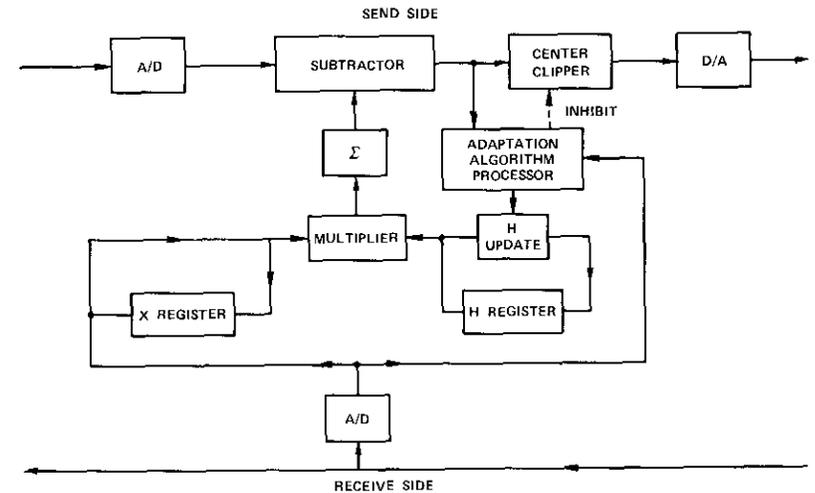


Figure 3. Block Diagram of the COMSAT Echo Canceller

During each sampling interval, the contents of the X and H registers are successively multiplied and accumulated to produce a sample of the estimated echo by convolution. This sample is subsequently subtracted from the real echo sample appearing at the send-in side of the echo canceller. This constitutes the cancellation process.

The H register update adder receives a command for each of the stored values from the circuit labeled "adaptive update," which is more fully described in Reference 9. These commands are inhibited at the onset of double talk, which is detected and integrated as part of the updating process. Thus, in the COMSAT echo canceller the impulse response model

During phases I, II, and III, populations of users were interviewed in each of the participating countries.* During phase IV, only users in Australia and Japan were interviewed.

Interviewing procedures were similar but not identical in each country. Answers to the following basic questions, asked during call-back interviews, provided raw data for the comparative analysis:

- a. Did you or the party you spoke to have any difficulty talking or hearing over the connection?
- b. What was the nature of the difficulty?
- c. Which of the following four categories best describes your assessment of the call? 1. excellent, 2. good, 3. fair, or 4. poor.

Question *a* elicited just a simple yes or no answer and resulted in a single percentage point called "percent difficulty" (%D) for each test condition. Question *b*, asked only if the answer to question *a* was affirmative, was used to obtain the customers' assessment of their difficulties, which were later classified in nine groups to be discussed in the next section. Question *c* resulted in a histogram of the results (in percentages) scored in each of the four quality assessment grades. A numerical average, weighted over the four grades and called the mean opinion score (MOS), was also computed from these data.

In addition to subjective assessment by customer call-back interviews, certain measurements and observations of the echo canceller's performance were made. Some of these observations yielded pertinent data on echo path temporal stability, whereas others revealed the extent to which the cancellation principle as implemented was capable of providing the required kind of echo protection.

The data resulting from the call-back interviews were gathered by each administration or by an independently designated agency, as in the case of the U.S. The raw data were coded and telexed to COMSAT Laboratories for statistical analysis. Several administrations performed their own analysis. No basic difference between the results obtained from these analyses and those obtained by COMSAT Laboratories has been observed.

Field trial results

Upon arrival in each location, the echo cancellers were checked out and installed on the test circuit. Proper operation was then verified by

* Strictly speaking, the West German customers were not interviewed, but the quality comparison was accomplished by service observation.

observing several critical performance criteria. The two most important criteria were speed of convergence and echo return loss enhancement. Both items determined in general how well the echo canceller performed and in particular how well the performance was maintained on time-variant echo paths. Furthermore, during several days at the start of each phase, the initial call-backs were checked to verify that the procedures and interview routines were correctly executed.

Speed of convergence and impulse response observations

Convergence to an echo return loss enhancement value greater than 15 dB was attained within the first three words of a conversation. In most cases the ultimate echo return loss enhancement measured about 25 dB, corresponding to a fully built up impulse response. Of course, somewhat lower values were found in circuits with high initial echo return losses. It was also generally observed that, with the center clipper,* even a first utterance such as "hello" was sufficient to prevent echo from being returned to the speaker.

Photographs of typically encountered echo path impulse responses were taken in several countries; examples are presented in Figure 5. Note that the impulse responses are shown in time reverse order since they were stored in this order in the H register memory to accommodate convolution processing.

Impulse response *a* in Figure 5 was obtained on a circuit between London and Leicester in the U.K., with an end delay of approximately 4 ms. Impulse response *b* was obtained on a very short echo path in Frankfurt, West Germany, during a test conversation with Rio de Janeiro, Brazil. The large ripple was caused by a low-frequency "hum" on the line; this phenomenon did not affect the ability of the canceller to maintain good cancellation.

Impulse response *c*, recorded in France, is an example of multiple reflections in the echo path. This occurred frequently in France because the echo canceller in the Paris CT was installed on a circuit which was switched on a 2-wire basis into the French network. Thus the echo path included a second 4- to 2-wire conversion point which could be far removed for calls well outside Paris. This situation also existed in Brazil and West Germany. Impulse response *d*, recorded in Hawaii, a 4-wire switch point, showed an end delay of about 3.6 ms.

* Center clipping is the process of zeroing a signal when its instantaneous value drops below a given threshold.

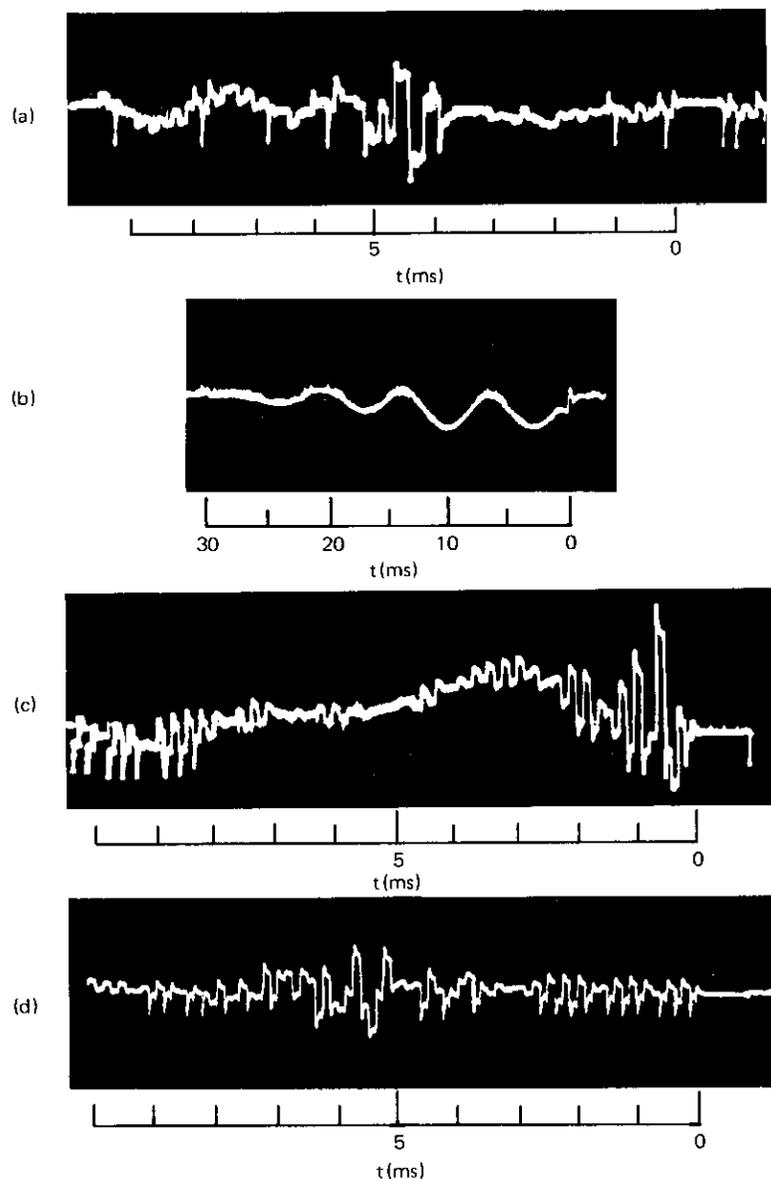


Figure 5. Example of Impulse Responses Encountered in the Networks

Time-variant echo paths

It was observed that two basically different phenomena caused time variance in the echo path. The first phenomenon was the speech level dependence of syllabically companded carrier systems, observed on N-type carrier systems in the U.S. terrestrial extension. This type of time variance results primarily in magnitude rather than phase changes of the impulse response. The second phenomenon, offsets in modulators and demodulators of carrier systems, caused a cyclic change (phase roll) of the impulse response, affecting both its amplitude and phase at any given point [10].

Neither of these two types of time-variant echo paths caused noticeable degradation in cancellation enhancement. This was established by observing the "send-out" side of the canceller for any possible spurts of echo when speech was present on the "receive-in" side of the canceller.

Although no noticeable spurts of echo were measured for the cases of phase roll observed, it cannot be concluded that phase roll could not be a problem elsewhere in the world where rates higher than those observed (about 0.5 Hz) might occur. Also it should be noted that observations were possible on only a small percentage of the total population of calls and troublesome cases may simply have gone unobserved.

Subjective assessment by call-back

It is generally agreed that telephone quality assessment by customer interviewing requires sample sizes of at least 100 to yield minimally reliable results, since fluctuations of small samples can be rather large in this kind of statistical process.

The method of call-back interviewing followed the general guidelines of the C.C.I.T.T. [11]. A trained interviewer asked each customer who completed a call over the circuit under test the three basic questions discussed in the preceding section. Six statistically homogeneous populations were sampled, with sample sizes generally well in excess of 100 for each test condition of the trial. In three cases anomalous results necessitated a separate analysis of the data involved.

Analysis of answers to the three basic questions is summarized in Figure 6 for the six homogeneous populations. For each case, the percentage difficulty, %D, and the percentages of customers rating the quality of the calls as excellent, good, fair, or poor were computed. A weighted mean of these four categories, the mean opinion score (MOS), is shown at the right-hand side of Figure 6.

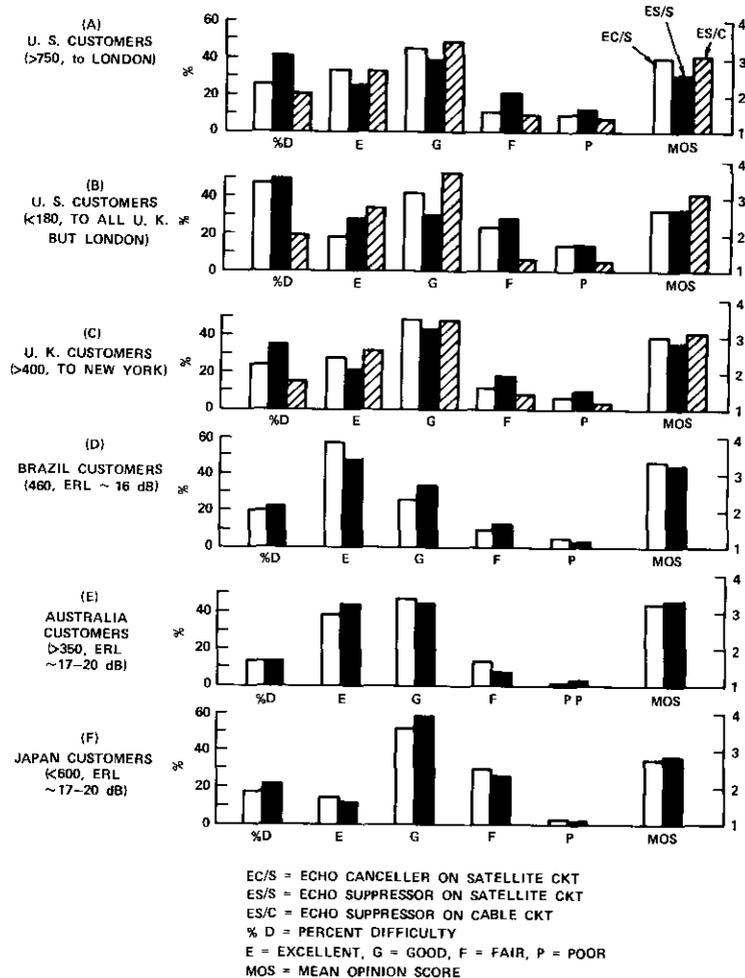


Figure 6. Distribution of Customer Interview Responses for Six Populations

The results for cases A, B, and C were all obtained during phase I of the trial, between the U.K. and the U.S. (New York and surrounding area), including results obtained with cable circuits equipped with echo suppressors. About 80 percent of the U.S. customers sampled talked to London customers (case A), and the circuit equipped with an echo canceller performed significantly better. This circuit represented a condition

in which the circuits were close to the preferred reference equivalent (i.e., low loss) and exhibited a relatively low echo return loss, causing noticeable echo. The data show a significant improvement obtained by using echo cancellers.

Case B in Figure 6 shows results for U.S. customers talking to U.K. customers, excluding the London area. It was generally surmised that higher losses prevailed for these customers and that the advantage of the echo canceller with respect to the echo suppressor was decreased or negated.

Case C in Figure 6 shows results of conversations between U.K. customers residing in the London area and New York area parties. The data indicate an improvement with echo cancellers bordering on statistical significance.

In cases D-F of Figure 6 the echo return losses for most of the calls were well above average, i.e., greater than 17 dB, resulting in a low echo level. Consequently, differences between cancellers and suppressors were not statistically significant.

Three special cases warrant further discussion. The first case is that of calls from West German customers to customers in Brazil. The Bundespost of West Germany preferred the use of service observation rather than call-back interviews to obtain basic data for analysis. Because the data set and analysis are not directly comparable with those presented in Figure 6, they are presented separately in Table 1. The ratings in this case were made by trained service observers. These ratings also resulted in percentages of excellent, good, fair, and poor, and an MOS value.

TABLE 1. RESULTS OF RATINGS BY SERVICE OBSERVERS OF CALLS BY WEST GERMAN CUSTOMERS

	Satellite Circuit Equipped with Echo Canceller	Satellite Circuit Equipped with Echo Suppressor
Number of Calls	325	324
Mean Opinion Score	3.41	3.40
Percent Excellent	59.7	57.1
Percent Good	25.9	28.7
Percent Fair	9.9	11.4
Percent Poor	4.6	2.8

A limited sample of the echo return loss in Brazil was also measured since its value had a direct effect on perception of echo effects by German customers. Since the echo return loss measured 18 dB or better most of the time, it could be expected that a comparison of cancellers and suppressors would reveal little difference. This was substantiated by the data presented in Table 1.

The second special case consisted of results obtained from U.K. customers primarily residing outside the London area. In this case, the sample sizes were almost all far below the desired minimum of 100, and, in addition, poorer transmission conditions prevailed for the customers in this area. Hence, the results were mixed and consequently insufficiently clear for interpretation.

The third special case occurred during phase II of the trial, conducted between France and Brazil. Special analysis of the results was necessary because the data obtained on the echo canceller circuit showed a severe time trend. This was created by an unresolved condition of intermittently high loss occurring in the circuit, as evidenced by a 1-to-1 correspondence between complaints of low volume and poor ratings during the interviews. On the other hand, the echo suppressor circuit did not experience this condition and showed no time trend at all. Those results were thus treated as usual.

The canceller results for Brazil and France are shown in Figures 7 and 8, respectively. They represent the course of the trial after approximately every 10th calendar day of data collection. As shown in Figures 7 and 8, it was possible to fit a statistically significant second-order regression curve through the data points, confirming the exhibited time trend.

Because of the time trend, comparison of the canceller and suppressor data did not result in a definitive conclusion. However, it may be seen from Figures 7 and 8 that such a comparison would have ranged from "significantly worse" in the beginning to "significantly better" at the end. Further, it was known that the trouble was not caused by the canceller operation itself but by a lossy connection elsewhere. Apparently, when this condition cleared, the performance of the canceller equipped circuit could be represented by the last portion of the curves, indicating superior performance.

The results from the circuits equipped with echo suppressors for both Brazilian and French customers did not exhibit a time trend and are presented in Table 2.

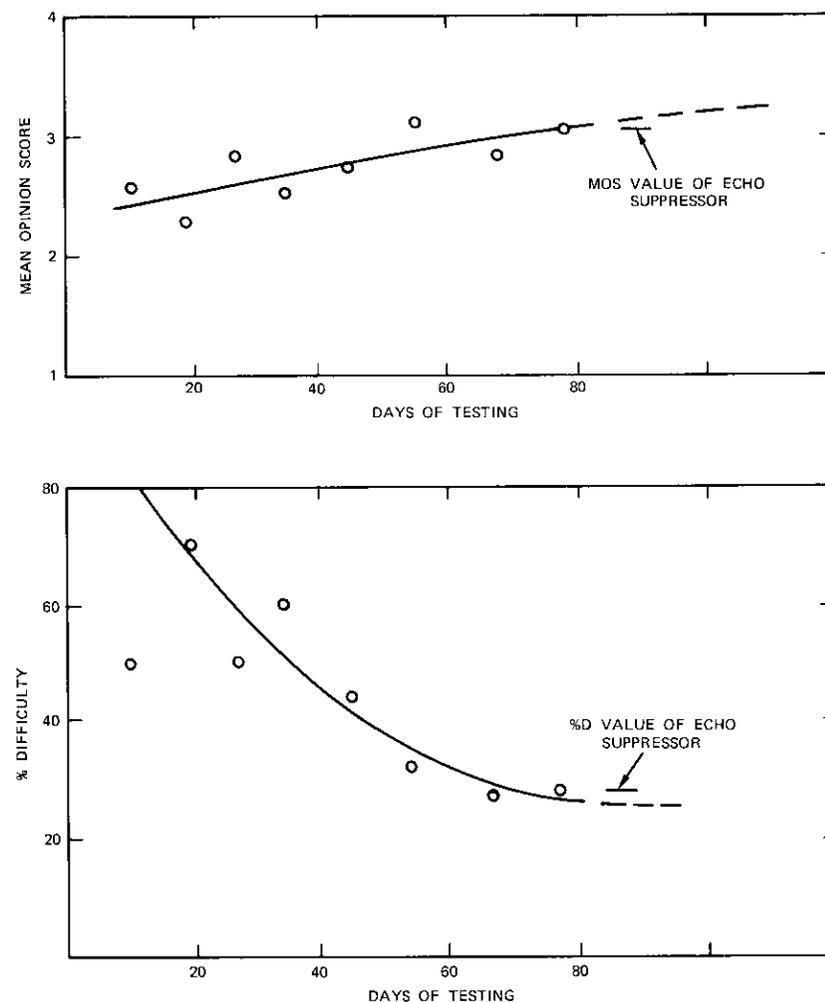


Figure 7. Brazilian Customer Interview Results of Echo Canceller Circuit [mean opinion score (MOS) and percentage difficulty (%D) as functions of time]

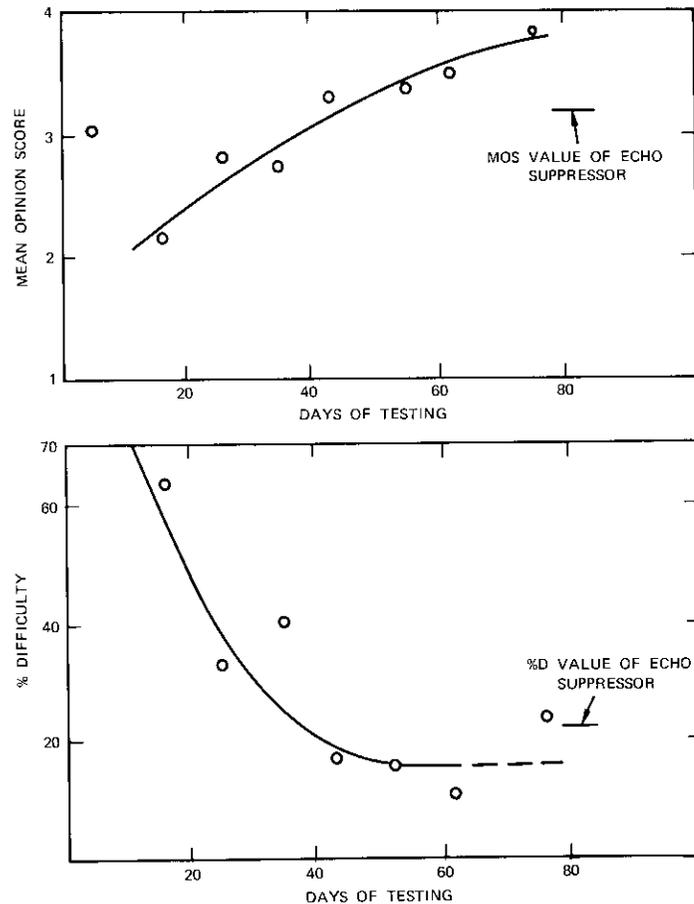


Figure 8. French Customer Interview Results of Echo Canceller Circuit [mean opinion score (MOS) and percentage difficulty (%D) as functions of time]

TABLE 2. RESULTS OF BRAZILIAN AND FRENCH CUSTOMER INTERVIEWS ON ECHO SUPPRESSOR SATELLITE CIRCUITS

Result (%)	Brazil	France
Excellent	37.6	40.3
Good	39.5	42.5
Fair	14.4	14.8
Poor	8.5	2.4
Difficulty	28.4	24.3
Low Volume	16.0	10.0
MOS	3.06	3.19

Nature of customer difficulties

There is a widespread range of opinions on the value of customer reported difficulties for finding the true causes of these difficulties. Some researchers believe that customer reported difficulties bear little relationship to engineering shortcomings of a circuit. However, the authors have found that these reports are sometimes quite helpful in the search for solutions of certain echo control related problems encountered during this and other field trials.

The difficulties reported by the customers during the interviews were generally categorized in eight specific and one general area of complaint. Some typical results of the percentages of complaints in these categories are shown in Figure 9. It may be seen that low volume and noise were generally the predominant causes of complaint.

Conclusions and future outlook

The field trial demonstrated that echo cancellers significantly improve the performance of single-hop satellite circuits when echo is the principal cause of difficulty. This is the case of circuits characterized by low loss, low noise, and poor (low) echo return loss. It was also found that conditions in the currently used terrestrial terminations frequently are such that echo is not the primary cause of difficulty; its effects are overshadowed by high loss and noise in tests involving a large percentage of the population. In this case, the performance of echo cancellers does not differ much from that of echo suppressors.

With this in mind, the overall results of the echo canceller field trial can be summarized as follows. Objective observations of the echo cancellers have proved that the designs worked very well at all times, exhibiting no drift problems. The convergence algorithms implemented in the two (NEC and COMSAT) designs are very robust and have been able to cope with all echo path conditions in which they have been operating. This includes nonlinearities of the kind exhibited in syllabically companded carrier systems, and time variance at a rate of about 0.5 Hz.

Subjective evaluations by means of call-back interviews have shown that, when circuits with low loss and less than average echo return loss are equipped with echo cancellers, they exhibit a significant improvement in quality relative to similar circuits equipped with echo suppressors. The call-back interview method has also demonstrated that the quality improvement achieved by using cancellers diminishes when echo return loss is well above average, i.e., >16 dB, and the circuits have relatively

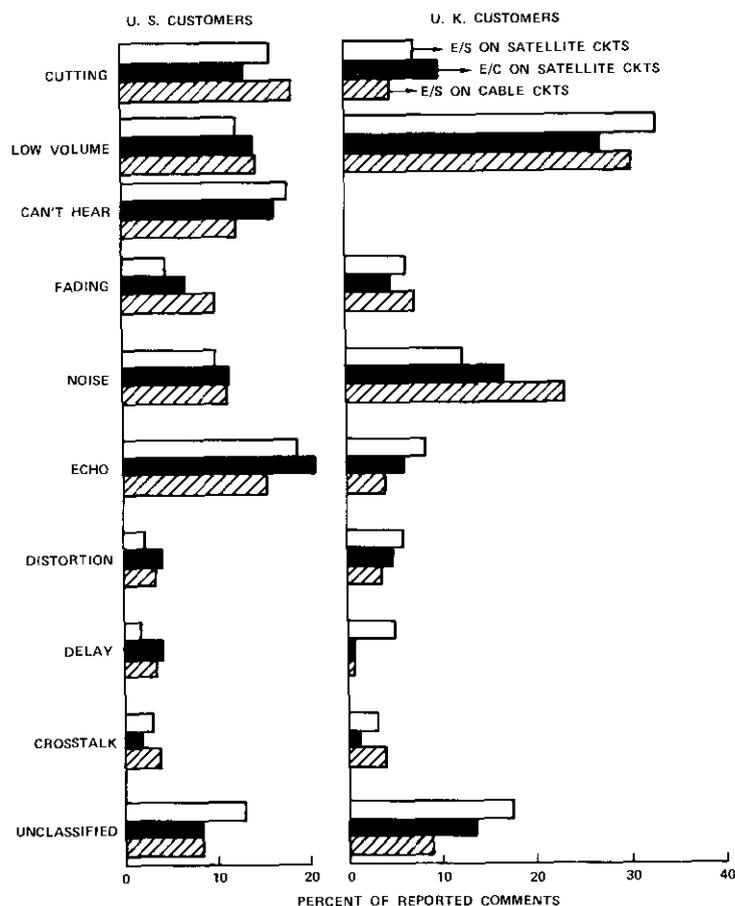


Figure 9. Typical Examples of Reported Difficulties

low loss. When high loss and/or high noise mask the signal, the results of subjective assessments are no longer based on problems with echo control but rather depend on low volume and high noise.

The C.C.I.T.T. recommends that international circuits operate at lower loss and noise levels than those generally provided at present. Once these objectives are met, echo will become an increasing cause of user difficulty since the masking effects of loss and noise will be removed while

the echo level will become higher. Thus the problems of echo control will increase as the C.C.I.T.T. objectives are approached. It is therefore expected that the need for echo cancellers will significantly increase as the telephone administrations of the world upgrade their facilities to meet the C.C.I.T.T. overall reference equivalent objectives [12].*

Although the tests did not include double-hop circuits, it should be expected that in this case the effects of echo would be more aggravating and would cause greater difficulty. Thus, echo cancellers would be more effective on such circuits.

Echo cancellers of present design have proven to be successful devices. Although the present designs exceed the limits of economical viability, it should be remembered that the devices tested are experimental prototypes. Techniques for significant simplification have been developed and are being integrated into future designs to provide greater cost effectiveness.

One approach which is also being considered for improved cost effectiveness is an echo control system which utilizes a sophisticated main processor in conjunction with simplified channel units. The main processor is used for obtaining and maintaining the echo path models, while the simpler channel units perform the actual cancellation. This approach may be considered to be a modification of the echo suppressor pool [13], in which echo suppressors are assigned to circuits as they are required.

Acknowledgment

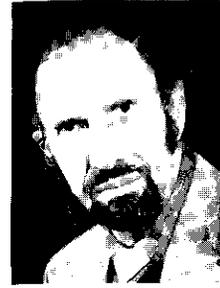
The authors wish to thank the Administrations of Australia, Brazil, France, Japan, West Germany, United Kingdom, and AT&T and Hawaiian Telephone Companies in the United States for their cooperation and participation in the field trial. A special note of gratitude is due to the personnel of each of the Administrations who were responsible for maintaining the equipment, collecting the data, and administering the test program at each location.

References

- [1] M. M. Sondhi, "An Adaptive Echo Canceller," *Bell System Technical Journal*, Vol. 46, No. 3, March 1967, pp. 497-511.
- [2] A. Miura et al., "The Blockless Echo Suppressor," *Trans. IEEE*, Japan 51-A, No. 7, 1968, p. 259.

* Overall reference equivalent (ORE) is a measure of end-to-end loss between telephone handsets.

- [3] R. B. Zmood, "Some Aspects of the Development of a Self-Adaptive Hybrid," P.M.G. Research Laboratories, Report No. 6150, August 1966.
- [4] J. W. Emling and D. Mitchell, "The Effects of Time Delay and Echoes on Telephone Conversations," *Bell System Technical Journal*, Vol. 42, No. 6, November 1963, pp. 2869-2891.
- [5] G. K. Helder, "Customer Evaluation of Telephone Circuits with Delay," *Bell System Technical Journal*, Vol. 45, No. 9, p. 1157.
- [6] R. D. Riesz and E. T. Klemmer, "Subjective Evaluation of Delay and Echo Suppressors in Telephone Communications," *Bell System Technical Journal*, Vol. 42, No. 6, November 1963, pp. 2919-2941.
- [7] R. G. Gould and G. K. Helder, "Transmission Delay and Echo Suppression," *IEEE Spectrum*, Vol. 7, No. 4, April 1970, pp. 47-54.
- [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, and M. Onufry, "Performance of a Digital Adaptive Echo Canceller in a Simulated Satellite Circuit Environment," *Progress in Astronautics and Aeronautics*, Vol. 33, 1974, pp. 455-477.
- [10] H. Brueggeman et al., "Experiments on Adaptive Echo Cancellation," *Australian Telecommunications Research*, Vol. 4, No. 1, May 1970, pp. 14-15.
- [11] International Telegraph and Telephone Consultative Committee (C.C.I.T.T.), Annex 1 to Question 2/XII, *White Book, IV Plenary Assembly, Mar del Plata*, September 23-October 25, 1968, pp. 2-6 Geneva: The International Telecommunications Union, 1969.
- [12] International Telegraph and Telephone Consultative Committee (C.C.I.T.T.), Rec. P.11 and P.21, *Green Book, V Plenary Assembly*, Vol. V, 1972, pp. 5-6, 24-28.
- [13] The International Telegraph and Telephone Consultative Committee (C.C.I.T.T.), Rec. G.131, "Stability and Echo," *Green Book, V Plenary Assembly*, Vol. III, 1972, p. 54.



H. G. Suyderhoud was born in Borneo, Indonesia. He received the Ingenieur degree from Delft Technological University in the Netherlands in 1955, and the M.S. in Applied Mathematical Statistics from Rutgers in 1966. He is presently Manager of the Signal Processing Department of the Communications Processing Laboratory at COMSAT Laboratories.

For more than 11 years prior to joining COMSAT, he worked for Bell Telephone Laboratories in systems engineering of exchange area transmission, broadband carrier terminals, and picturephone. He is a member of IEEE and the American Statistical Association.

S. J. Campanella received his B.E.E. from the Catholic University of America, his M.S.E.E. from the University of Maryland, and his Ph.D. in E.E. from Catholic University. He is presently Director of the Communications Processing Laboratory of COMSAT Laboratories. In this position he is responsible for technical orientation of the efforts pursued in the Signal Processing and Digital Control and Modulation Techniques Departments. He has performed research, development, and engineering in voice and visual signal processing and in digital techniques for telephone and data communications via satellite links.

Dr. Campanella is also recognized for his work in echo control and digital speech interpolation. He is a member of Sigma Xi, Phi Eta Sigma, IEEE, AIAA, and the Acoustical Society of America. He is also a member of the board of directors of EASCON and the editorial board of the COMSAT Technical Review.





Michael Onufry obtained the B.S. degree from Pennsylvania State University in 1962 and did graduate work at both Northeastern University and George Washington University. He is presently Assistant Manager of the Signal Processing Department in the Communications Processing Laboratory of COMSAT Laboratories. His work includes computer simulation, design, and simulation of digital echo cancellers and a variety of speech processing devices. He also

assists in planning and implementing field trials for the echo canceller and other equipment for the INTELSAT network. Moreover, he is responsible for conducting experiments to assess subjectively speech processing devices and to study the effects of delay and echo on telephone conversations, and for intelligibility tests on new speech processing systems. He is a member of IEEE.

Index: pulse code modulation, digital simulation, digital television, communications systems

Study and simulation of a variable length code DPCM for the luminance signal in DITEC

I. DINSTEIN

(Manuscript received March 13, 1975)

Abstract

Some quantizing and coding algorithms for differential pulse code modulation (DPCM) will be analyzed and evaluated in terms of data statistics generated on the digital television communication system (DITEC) developed at COMSAT Laboratories. A DPCM encoder for monochrome television signals using entropy coding and selective prediction is proposed. Simulations of this encoder show that, at a transmission rate of 4.5 bits/picture element (pel), there is no perceptible degradation between the original and the reconstructed pictures. A transmission buffer of 4,000 bits has been found adequate to accommodate the digitized signal.

Introduction

The objective of this study is to develop an encoder for the luminance signal in the DITEC (*digital television communication*) system developed at COMSAT Laboratories [1], [2]. At present the NTSC composite color television signal is separated into its luminance (Y) and chrominance (I and Q) components by using a comb filter. The Y component is sampled at a rate of 6.018 MHz and digitized with 7 bits/sample. The encoder uses a combination of differential pulse code modulation (DPCM) and pulse

code modulation (PCM). When the absolute value of the difference between two consecutive samples is less than or equal to 16, the difference is digitally quantized and transmitted. When the absolute value of that difference is greater than 16, the four most significant bits of the later sample are transmitted. In both cases, a 5-bit code is required, 4 bits for either the DPCM or the PCM value and the extra bit to indicate to the receiver which of the two has been transmitted. The I and Q components are sampled at rates of 1.77 and 0.668 MHz, respectively; digitized with 6 bits/sample; and coded in a similar manner using 4 bits/sample. These sampling rates and coding yield a bit rate of about 29 Mbps, and a rate-7/8 error correction code brings the transmission bit rate up to about 33 Mbps.

Differential PCM has been widely used in the transmission of video signals [3]–[5]. In a DPCM coder, the brightness level of each picture element (pel) is estimated from a linear combination of some of its neighboring pels, and the quantized difference between the real value and the estimation is transmitted. When the estimation is optimized, the differences are independent and therefore the redundancy due to the high correlation between adjacent pels is eliminated from the transmission. For practical picture material, DPCM yields a reduction of about 2 bits/pel as compared to PCM at the same picture quality.

Various algorithms have been proposed for the design of differential quantizers. Subjective evaluation of edge busyness [6] has indicated that human observers can tolerate a decrease in the signal-to-noise ratio (S/N) of about 12 dB in picture regions where the differential value is more than 12 percent of the maximum possible difference value. This means that, for a picture digitized with 7 bits/sample, edges with differential values greater than 16 may be represented with coarser quantization, e.g., 5 bits/sample, without degrading the picture quality as evaluated by a human observer. Hence, an efficient differential quantizer does not treat small and large differences in the same manner.

Division of the range of differential values into two regions is a common practice. For example, in his sliding scale operation, Brown [7] used fine and coarse differential values. Limb [8] defined high- and low-detail categories and experimented with adaptive coding in which a sequence of m picture elements was assumed to fall into the high- or low-detail category if more or less than N of the picture elements had differential values exceeding a specified threshold, respectively.

The distribution of the differences between the actual values of pels and the corresponding estimations is close to a Laplacian distribution, which has the shape of a narrow bell centered at zero. The use of fixed

length codes to code the elements of such a distribution yields a bit/pel rate which is higher than the entropy of the elements because the distribution of the differences is nonuniform. Entropy coding can be used to further reduce the required bit rate by assigning short codes to highly probable differences and longer codes to differences with low probabilities. A DPCM encoder with entropy coding may yield an S/N which is 5.6 dB higher than that of an equivalent DPCM encoder without entropy coding [9]. The implementation of an entropy coder for a 4-bit DPCM system has been reported [10].

This paper describes an encoder which employs DPCM and entropy coding. It also includes simulation results for the transmission of commercial-quality television pictures with no perceptible degradation of the transmitted picture due to quantizing and/or coding. In addition, it discusses the use of run length codes for sequences of pels with differential values of zeros, the transmission of only nonzero DPCM values, and the transmission of codes to represent distinct DPCM values within each line.

Prediction within the DPCM loop of the proposed encoder is performed selectively. For each segment of pels (a specified portion of a line), both horizontal and vertical predictions are performed; the one that yields the smaller number of bits per pel is used for the transmission.

Computer programs to simulate the proposed encoder were implemented on a PDP 11/10 computer interfaced to the DITEC system and on an IBM 360/65 computer to process data digitized on a flying spot scanner. The DITEC-PDP processing showed that a transmission bit rate of 4.5 bits/pel with a transmission buffer 4,000 bits long was adequate for the Y component. Processing on the scanner-IBM computer enabled subjective evaluation of the reconstructed pictures. There was no perceptible degradation between the original pictures and the reconstructed ones.

Differential pulse code modulation

The basic principle of DPCM is that the quantized difference between an estimation of each sample and its actual value is transmitted rather than the sample value itself. Use of linear prediction in a feedback loop yields bandwidth reduction for the transmission of redundant signals.

Figure 1 is a block diagram of a general DPCM system. The current sample, S_i , is linearly predicted by

$$P_i = \sum_{j=1}^N \alpha_j R_j$$

where R_j is a previously reconstructed sample obtained by using the delay D_j , and α_j is a weighting coefficient for R_j . The difference, d_i , between the current sample and its prediction is quantized and a quantizing error, q_i , is introduced. The quantized difference, $d_i + q_i$, is transmitted to the receiver. Both the transmitter and the receiver reconstruct the current sample by adding the quantized difference to the respective prediction:

$$R_i = P_i + d_i + q_i$$

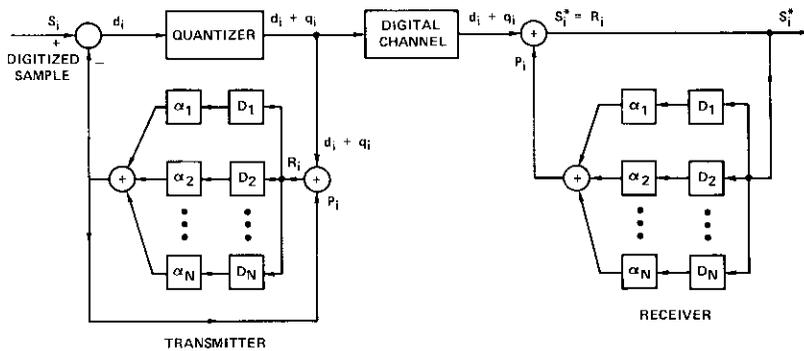


Figure 1. Block Diagram of DPCM System

Use of reconstructed samples for the prediction eliminates the accumulation of quantizing errors at the receiver.

O'Neal [3] discussed the optimization of the weighting coefficients and the use of various combinations of previous elements as prediction elements, and estimated the video signal-to-quantizing noise ratio of DPCM.

Evaluation of some quantizing and coding algorithms

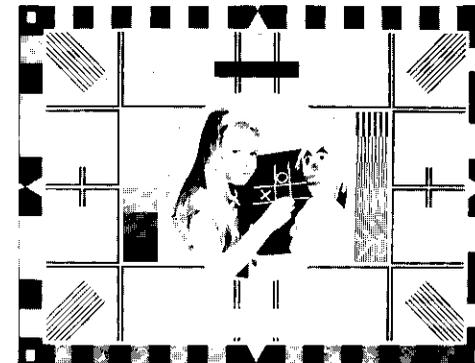
The following analysis has been performed to evaluate some possible quantizing and coding algorithms. The algorithms are introduced and discussed in terms of the computed statistics. The three slides used in this study are shown in Figure 2. Only the luminance component of the composite television signal has been analyzed since it accounts for at least 70 percent of the bits to be transmitted. Furthermore, it has been assumed that algorithms developed for the luminance component will be suitable for the chrominance components as well.



ZELDA



BEACH (ZELDA & FRIENDS)



BBC, TEST

Figure 2. Slides Used in the Preliminary Study

The processed luminance component in the DITEC system was sampled at a rate of 6.018 MHz, which is 765 times one-half the line frequency, to yield 318 samples during the active line period. The samples of adjacent lines of a field were interleaved, as shown in Figure 3, because the sampling rate is an odd multiple of half the line frequency. The processing for this section was performed on a PDP 11/10 computer interfaced to the DITEC system.

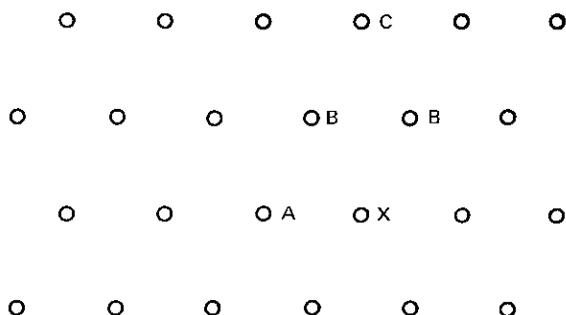


Figure 3. *Interleaved Samples*

Digitized data were transferred from the DITEC system to the PDP computer disk pack, one line at a time. Once stored on the disk, the data were processed and some relevant statistics were printed out. Only three lines were stored in the computer core at any particular time. The results were compiled into tables and are summarized and discussed in the following sections.

Prediction schemes

A good prediction scheme for DPCM is one that yields small differential values. As the accuracy of the predictor increases, the differences become smaller. The performance of any prediction scheme depends on the picture material. A scheme using the nearest vertical pel as a predictor is excellent for the color bar test pattern, for example, but is not as good for a picture with horizontal stripes. For a collection of various practical picture material, there is no preferred direction to be used for the prediction. However, for a small region of a certain picture, a preferred direction is likely to exist.

Based on this principle, the following prediction schemes have been used:

a. The nearest horizontal pel is used as a predictor (horizontal prediction, or DF1).

b. Each line is partitioned into 10 segments. For each segment, the differential values are determined twice, using both horizontal and vertical predictions. The prediction selected is the one which yields the smaller average difference for the considered segment. (This prediction is referred to as "best of two," or DF2.)

c. Similar to the previous scheme, each line is partitioned into 10 segments. For each segment, the differential values are determined three times, using horizontal, vertical, and diagonal predictions. As in scheme b, the prediction selected is the one which yields the smallest average difference. (This prediction is referred to as "best of three," or DF3.)

These three schemes have been applied to the simulations; their performance is discussed in the following subsections.

Transmission of nonzero differences only

It is well known that the probability of zero differences is generally relatively high. This is the rationale for transmitting only the nonzero values. However, there must be a way of informing the receiver which pels have zero difference values and which have nonzero values.

Consider the following algorithm: Let N be the number of pels per TV scan line, $P(0)$ the probability of zero differences, and k the number of bits per sample in a fixed length code DPCM. The number of nonzero differences for each line is $[1 - P(0)] N$. The number of bits required for transmitting the nonzero differences in each line is $[1 - P(0)] Nk$, but an extra bit per pel is required to indicate to the receiver which pels have zero differences and which have nonzero differences. Hence, the number of bits per line required to transmit nonzero differences only is

$$[1 - P(0)] Nk + N$$

The number of bits per line required for regular DPCM is Nk , and the number of bits per line that may be saved by transmitting nonzero differences only is

$$\begin{aligned} Nk - \{[1 - P(0)] Nk + N\} &= Nk - Nk + P(0) Nk - N \\ &= [P(0) k - 1] N \end{aligned} \quad (1)$$

Obviously, equation (1) should yield positive values. Since N is positive, the expression is positive if

$$P(0) k - 1 \geq 0 \tag{2}$$

and since k is positive,

$$P(0) \geq \frac{1}{k} \tag{3}$$

Transmission of nonzero differences only is beneficial when inequality (3) holds. Table 1 lists the probabilities of zero differences in a sample of lines from the pictures considered in this study (see Figure 2). As might be expected, quiet scenes such as Zelda have higher probabilities of zeros

TABLE 1. PROBABILITY OF ZEROS

Line	Zelda			Beach			BBC Test		
	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3
1	0.294	0.325	0.325	0.133	0.133	0.133	0.293	0.293	0.293
12	0.196	0.204	0.212	0.218	0.211	0.219	0.223	0.219	0.223
23	0.192	0.263	0.235	0.187	0.199	0.203	0.187	0.207	0.207
34	0.153	0.231	0.224	0.113	0.160	0.168	0.230	0.234	0.254
45	0.145	0.208	0.188	0.047	0.078	0.070	0.309	0.297	0.332
56	0.149	0.263	0.231	0.102	0.117	0.125	0.379	0.402	0.453
67	0.118	0.165	0.192	0.082	0.125	0.121	0.281	0.301	0.301
78	0.102	0.133	0.200	0.090	0.125	0.121	0.258	0.285	0.285
89	0.137	0.153	0.157	0.129	0.184	0.187	0.195	0.184	0.195
100	0.169	0.208	0.243	0.129	0.156	0.152	0.219	0.227	0.246
111	0.176	0.212	0.188	0.086	0.141	0.129	0.172	0.187	0.246
122	0.271	0.298	0.325	0.125	0.187	0.176	0.168	0.176	0.207
133	0.267	0.298	0.278	0.117	0.160	0.145	0.211	0.219	0.250
144	0.271	0.337	0.353	0.117	0.125	0.137	0.207	0.242	0.289
155	0.267	0.294	0.314	0.121	0.164	0.113	0.297	0.320	0.406
166	0.278	0.278	0.267	0.117	0.137	0.133	0.250	0.258	0.266
177	0.282	0.314		0.086	0.094	0.102	0.242	0.262	0.320
185	0.204	0.231		0.105	0.141	0.133	0.266	0.266	0.305
199	0.267	0.294	0.263	0.133	0.133	0.141	0.336	0.375	0.359
210	0.176	0.247	0.220	0.141	0.141	0.152	0.187	0.184	0.211
221	0.188	0.227	0.231	0.164	0.168	0.164	0.199	0.199	0.199
232	0.239	0.275	0.251	0.219	0.238	0.211	0.340	0.328	0.363

than busier scenes such as the beach. In Table 2 the probability that equation (3) will hold for $k = 3, 4,$ and 5 has been estimated. Note that difference image DF2 (best of 2) for Zelda has considerably higher values of $P(0)$ than the corresponding DF1. It might also be concluded that this algorithm would not be appropriate for busy pictures such as the beach.

TABLE 2. PROBABILITIES FOR $P(0) \geq 1/k, k = 3, 4,$ AND 5

$P(0)$	Zelda			Beach			BBC Test		
	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3
≥ 0.20	10/22	19/22	16/20	2/22	2/22	3/22	16/22	17/22	20/22
≥ 0.25	8/22	11/22	7/20	0	0	0	10/22	11/22	11/22
≥ 0.33	0	1/22	1/20	0	0	0	3/22	2/22	4/22

Run length code for zero differences (or zeros and ones)

A run length code is appropriate when there is a considerable probability of sequences of pels having the same difference value [11]. For a string of L pels of value M , the code of M followed by the code of L replaces the string of L codes of M which would otherwise be required. This yields a savings of about $(L - 2)k$ bits, where k is the number of bits per pel. Computation shows that the only values to be considered for run length codes are 0 and ± 1 . Table 3 indicates the number of zeros in strings of three or more, and Table 4 indicates the number of zeros and/or plus or minus ones in such strings. These numbers should yield some bit rate reduction and hence further experimentation with this algorithm is desirable.

Transmission of distinct values

When the number of distinct values in a line is less than or equal to 16, a table of distinct values can be transmitted (coded with six bits per value). In this case, a pointer to the table (coded with four bits) is required for each pel. This procedure yields a very low quantizing noise.

As can be seen in Table 5, the requirement for the number of distinct values to be less than or equal to 16 is very strict. Some modification of this scheme may be considered. It can be assumed that differential values in the range of $(-3, +3)$ are always present; therefore, the number of

TABLE 3. NUMBER OF SEQUENCES OF ZEROS (LENGTH OF THREE OR MORE)

Line	Zelda			Beach			BBC Test		
	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3
1	5	6	6	3	3	3	7	7	7
12	5	5	5	6	6	5	4	3	3
23	3	4	5	3	4	2	3	4	2
34	1	5	5	1	2	3	8	7	9
45	2	5	2	0	0	0	8	8	6
56	1	7	6	2	2	3	13	12	16
67	1	2	3	1	1	1	5	6	5
78	1	1	3	0	0	0	8	6	5
89	1	3	3	1	4	4	5	3	3
100	1	4	4	0	1	1	4	6	7
111	3	4	3	2	5	3	3	2	5
122	5	9	8	3	4	4	3	3	3
135	6	8	5	1	2	2	5	3	6
144	7	8	9	1	1	1	5	6	8
155	6	8	9	1	2	0	11	11	14
166	5	5	4	2	2	2	5	6	6
177	9	8	10	0	0	0	3	3	7
185	3	5	4	1	3	2	5	5	7
199	8	10	8	0	0	0	10	12	10
210	3	7	6	0	0	0	6	3	3
221	2	3	4	3	4	3	4	4	4
232	5	6	6	3	3	4	11	9	7

distinct values to be handled in this way is $16 + 7 = 23$. Another possible modification is one in which the range of differential values is partitioned into two regions, fine and coarse, and only the coarse region is treated in this manner.

Variable length codes, fine-coarse partition

The distribution of differential values of the pictorial material is close to Laplacian. This means that, as the absolute value of the difference becomes smaller, its probability increases. This is one good reason for partitioning the range of differences into regions of small (fine) and large

TABLE 4. NUMBER OF SEQUENCES OF ZEROS AND PLUS OR MINUS ONES (LENGTH OF THREE OR MORE)

Line	Zelda			Beach			BBC Test		
	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3
1	13	14	15	16	16	16	28	28	28
12	24	23	24	19	20	20	29	29	29
23	22	29	26	15	14	14	22	19	21
34	24	24	21	9	12	14	24	27	24
45	18	23	24	14	16	14	28	27	28
56	22	22	21	9	14	14	31	31	26
67	15	20	18	9	15	17	33	34	36
78	13	11	18	15	16	16	25	29	26
89	20	22	17	15	20	19	21	25	25
100	20	20	22	14	18	17	20	20	25
111	16	18	18	12	14	15	18	23	24
122	27	32	29	14	18	18	22	22	27
133	26	29	29	17	17	18	26	23	27
144	29	30	31	18	19	18	20	23	18
155	27	29	31	11	13	14	31	31	27
166	27	27	25	15	16	15	27	26	24
177	21	19	24	9	12	11	35	31	30
185	21	24	26	15	18	16	31	31	32
199	27	26	25	19	20	21	26	21	27
210	24	31	28	20	21	22	15	11	11
221	23	25	26	28	28	28	27	27	27
232	20	25	25	23	26	24	32	31	25

(coarse) differences, and assigning shorter codes to the fine region and longer codes to the coarse region. Such code assignment decreases the average number of bits per pel. In general, assume that the range of differences is partitioned into N regions, and let $P(n)$ be the probability of the n th region and $B(n)$ be the number of bits in codes assigned to that region. Then the average number of bits per pel is given by

$$B_{ave} = \sum_{n=1}^N P(n) B(n) \quad .$$

TABLE 5. NUMBER OF DISTINCT VALUES PER LINE

Line	Zelda			Beach			BBC Test		
	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3
1	32	46	46	35	35	35	41	41	41
12	16	21	17	38	31	35	48	47	46
23	17	14	14	46	41	42	52	48	54
34	18	17	25	44	42	42	45	44	48
45	19	20	24	44	41	41	49	47	36
56	20	20	22	48	48	53	44	40	26
67	19	24	29	42	39	41	46	44	32
78	20	25	22	48	44	45	52	41	28
89	22	25	29	45	36	36	50	45	35
100	20	23	21	43	36	38	55	50	35
111	20	24	21	49	47	49	50	47	38
122	23	18	21	47	38	37	53	57	54
133	17	15	15	43	38	39	47	40	38
144	16	15	14	46	49	49	48	45	52
155	15	13	13	46	49	45	47	37	25
166	17	17	19	42	43	44	35	33	32
177	19	17		51	47	46	41	36	26
185	21	23		45	46	46	58	55	49
199	21	22	20	43	39	38	39	43	39
210	30	31	34	39	36	49	70	61	49
221	25	28	29	29	29	29	42	42	42
232	24	25	27	22	18	19	38	34	20

To minimize B_{ave} , the following must be satisfied: for every two regions, if $P(i) \leq P(j)$, then $B(i) \leq B(j)$ [12]. If this is not the case, then interchanging $B(i)$ and $B(j)$ should yield a decrease in B_{ave} , which contradicts its minimality. Hence, short codes should be assigned to more frequent differences and longer codes to less frequent ones.

Differential values in the fine and coarse regions can be quantized differently. The human eye tolerates more noise on brightness boundaries of a picture than in regions of uniform brightness. Therefore, more quantizing noise may be allowed in the coarse region, hence reducing the average number of bits per pel.

Figure 4, which shows B_{ave} for the case of $N = 2$, gives the probability in percentage. For example, if 3 bits/pel are assigned to fine differences, 7 bits/pel to coarse differences, and the average bits/pel should be less than 4, then at least 75 percent of the pels should have fine values. For an

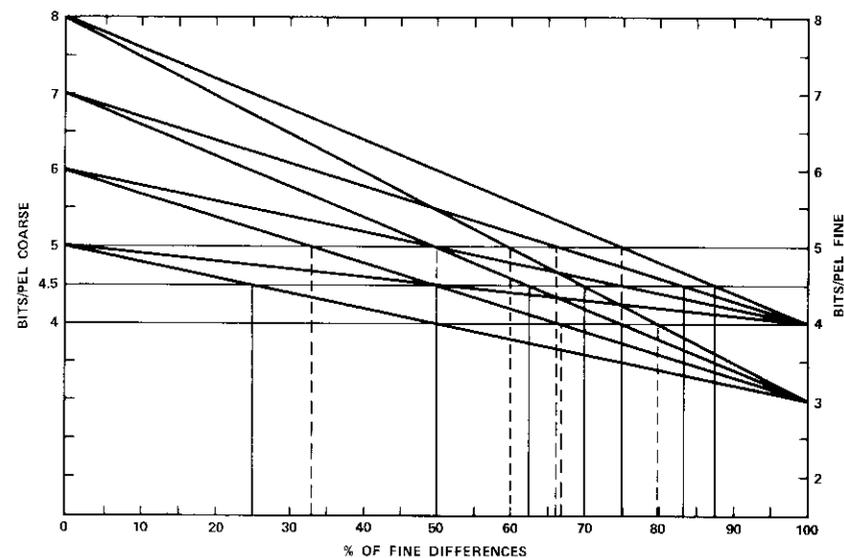


Figure 4. Various Distributions and Variable Code Lengths for Fine and Coarse Regions

average of 4.5 bits/pel, at least 62.5 percent of the differences should be fine.

The differential range can be adaptively partitioned into fine and coarse regions for each line according to the distribution of the differences within the lines. The following steps should be applied to each line:

- The smallest threshold value T is calculated so that the probability of difference values being within the range $(-T, +T)$ is at least 0.65.
- The minimum and maximum differential values are calculated.
- Based on the values of T , the minimum and maximum of the differences, a look-up table to be used for this line and code lengths for fine and coarse regions are selected. These selections should be based on a prescribed rule to be constructed upon completion of this study.

Table 6 lists values of T for a sample of lines from the slides described herein. Even in the busy beach scene, the highest entry is 7, which means that assigning 4 bits/pel in the fine region (in this case) is appropriate. For many other lines, 3 bits/pel in the fine region may yield satisfactory results.

TABLE 6. THRESHOLD VALUES OF T FOR WHICH THE PROBABILITY OF ($|\text{DIFFERENCE}| \leq T$) ≥ 0.65

Line	Zelda			Beach			BBC Test		
	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3
1	5	5	5	5	5	5	2	2	2
12	2	2	2	4	3	3	2	2	2
23	2	2	2	5	4	4	6	5	5
34	2	2	2	7	7	6	3	2	2
45	3	2	3	6	5	5	1	1	1
56	3	2	2	7	6	6	1	1	1
67	3	3	3	6	5	6	2	2	1
78	3	3	3	5	5	5	3	2	1
89	3	3	3	5	4	4	4	2	2
100	3	2	2	4	4	5	4	3	2
111	3	3	3	7	7	7	5	4	2
122	2	1	1	5	4	5	5	4	2
135	2	1	1	5	4	5	3	2	2
144	1	1	1	5	5	4	4	4	3
155	2	1	1	7	6	5	2	2	1
166	2	2	2	5	5	5	2	2	2
177	2	1		6	6	5	2	1	1
185	3	2	2	6	5	6	2	2	2
199	2	2	2	4	3	3	2	1	1
210	2	2	2	4	3	4	15	9	4
221	3	2	2	4	3	3	2	2	2
232	3	2	3	2	2	2	1	1	1

In conjunction with this algorithm, the use of a code similar to the sliding scale or some extension of it is desirable [7].

Table 7 contains the ranges of differences for the sampled lines. It shows that using the ranges for adaptive quantizing [13] will reduce the quantizing noise considerably. It also shows that the use of difference image DF2 (best of 2) results in smaller ranges than the use of DF1 (one predictor). The advantage of DF2 is also demonstrated by the increased probability of small differences (Table 8).

It is interesting to note that the changes in both the ranges and the probability of small values are not reflected in the entropy values shown in Table 9. This is because entropy is independent of the scale of the measurements.

TABLE 7. COMPARISON OF RANGES

Line	Zelda			Beach			BBC Test		
	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3
1	77	80	73	110	110	110	63	63	63
12	113	130	113	122	110	111	78	78	78
23	119	108	108	104	45	58	80	74	75
34	132	132	132	68	68	68	131	131	131
45	97	86	97	79	43	56	91	84	91
56	108	57	59	92	57	108	88	76	73
67	140	141	145	69	45	51	103	78	68
78	121	86	37	136	96	96	118	77	59
89	121	80	121	162	48	44	90	73	63
100	114	49	40	136	34	86	34	78	49
111	154	124	154	113	75	76	92	89	88
122	117	79	79	110	67	50	107	81	125
133	157	119	119	78	50	78	102	89	85
144	136	96	102	83	85	87	105	105	95
155	140	100	100	103	86	56	98	82	63
166	159	159	159	101	60	64	138	138	138
177	135	63	135	112	73	73	86	65	58
185	119	100		99	82	108	91	91	86
199	116	66	74	129	129	128	104	104	88
210	139	81	127	101	85	92	96	79	63
221	102	118	118	105	105	105	124	124	124
232	107	60	107	130	117	130	54	54	68

Variable length code DPCM encoder simulation

The performance of a DPCM is determined by three factors: the prediction, the companding law, and the code assignment. Prediction in the proposed encoder is performed selectively. For each segment of pels (a specified portion of a line), both horizontal and vertical predictions are performed; the one that yields the smaller number of bits per pel is used for the transmission. Two companding laws are used, one for quiet and one for busy regions of the picture. Construction of the companding laws is based on subjective evaluation of statistics of the pictures used in the simulations as described in the previous section. The encoder assigns 3 bits to small differences, 6 bits to medium differences, and 10 bits (6 bits prefix + 4 bits PCM) to pels with large differences.

TABLE 8. COMPARISON OF PROBABILITIES OF $|\text{DIFFERENCE}| \leq 1$

Line	Zelda			Beach			BBC Test		
	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3
1	0.420	0.451	0.447	0.289	0.289	0.289	0.598	0.598	0.598
12	0.443	0.459	0.455	0.418	0.488	0.480	0.562	0.574	0.574
23	0.502	0.584	0.569	0.344	0.426	0.391	0.406	0.426	0.477
34	0.435	0.514	0.510	0.246	0.305	0.301	0.508	0.520	0.543
45	0.416	0.533	0.510	0.211	0.285	0.273	0.652	0.672	0.727
56	0.404	0.537	0.533	0.238	0.301	0.293	0.660	0.703	0.785
67	0.376	0.467	0.451	0.223	0.277	0.258	0.613	0.641	0.668
78	0.325	0.361	0.443	0.285	0.348	0.332	0.555	0.629	0.734
89	0.341	0.420	0.431	0.270	0.375	0.371	0.457	0.527	0.566
100	0.416	0.518	0.518	0.297	0.359	0.344	0.465	0.523	0.559
111	0.451	0.471	0.467	0.195	0.301	0.305	0.383	0.473	0.602
122	0.627	0.675	0.675	0.266	0.355	0.340	0.430	0.504	0.582
133	0.627	0.782	0.675	0.277	0.352	0.348	0.480	0.543	0.590
144	0.659	0.741	0.714	0.320	0.336	0.355	0.375	0.430	0.477
155	0.612	0.671	0.690	0.258	0.332	0.332	0.586	0.641	0.746
166	0.604	0.604	0.612	0.293	0.340	0.355	0.578	0.582	0.613
177	0.588	0.682		0.270	0.305	0.297	0.617	0.652	0.703
185	0.486	0.588		0.254	0.348	0.344	0.578	0.582	0.648
199	0.553	0.608	0.561	0.367	0.381	0.406	0.648	0.668	0.691
210	0.506	0.584	0.549	0.387	0.422	0.406	0.344	0.410	0.437
221	0.471	0.549	0.529	0.449	0.457	0.449	0.484	0.484	0.484
236	0.478	0.541	0.502	0.512	0.539	0.523	0.664	0.687	0.832

Selective prediction

The encoder contains two DPCM loops, one for horizontal prediction and one for vertical prediction. (Horizontal prediction means that the previous nearest horizontal pel is used as the predictor. Similarly, the previous nearest vertical pel is used as the predictor in vertical prediction.) The prediction is selected as follows: Each line is partitioned into 10 segments. Both horizontal and vertical predictions are applied to each segment, and the prediction used is the one that yields the smaller number of bits per pel. One extra bit is transmitted for each segment to indicate to the receiver the prediction which has been used.

Some parameters of the simulation results are listed in Table 10. Selective prediction, which enables the use of a smaller transmission buffer, yields a lower average number of bits per pel and improves the quantization by

TABLE 9. ENTROPY OF DIFFERENTIAL IMAGES

Line	Zelda			Beach			BBC Test		
	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3	1 Pre-dictor	Best of 2	Best of 3
1	3.88	4.15	4.15	4.51	4.52	4.52	3.70	3.7	3.70
12	3.33	3.45	3.34	4.20	3.86	4.03	3.97	3.95	4.01
23	3.23	3.02	3.02	4.65	4.35	4.41	4.50	4.47	4.66
34	3.39	3.22	3.51	4.87	4.72	4.70	3.99	3.96	4.05
45	3.52	3.38	3.60	4.82	4.65	4.70	3.53	3.103	3.64
56	3.53	3.36	3.49	4.97	4.85	4.91	3.24	2.66	3.47
67	3.71	3.69	3.80	4.82	4.67	4.71	3.68	3.28	3.83
78	3.68	3.81	3.62	4.73	4.58	4.63	3.71	2.93	4.12
89	3.74	3.61	3.79	4.67	4.27	4.29	4.05	3.70	4.37
100	3.56	3.49	3.46	4.56	4.27	4.38	4.17	3.68	4.47
111	3.69	3.73	3.69	5.01	4.88	4.89	4.32	3.68	4.61
122	3.11	2.89	2.96	4.75	4.37	4.45	4.38	3.93	4.54
133	3.08	2.88	2.90	4.59	4.42	4.51	3.89	3.70	4.13
144	2.88	2.64	2.70	4.58	4.57	4.52	4.22	4.17	4.38
155	3.03	2.86	2.80	4.88	4.70	4.61	3.51	2.88	3.85
166	3.05	3.05	3.05	4.62	4.57	4.54	3.72	3.59	3.74
177	3.19	2.99		4.81	4.68	4.71	3.55	3.10	3.76
185	3.56	3.38	3.47	4.80	4.56	4.62	4.06	3.65	4.10
199	3.39	3.27	3.35	4.33	4.23	4.15	3.48	3.29	3.52
210	3.75	3.55	3.72	4.30	4.17	4.43	4.88	4.44	5.22
221	3.69	3.53	3.60	3.82	3.81	3.82	3.94	3.94	3.94
232	3.58	3.58	3.65	3.43	3.34	3.35	3.28	2.42	3.43

decreasing the number of lines that are considered to be busy regions. (A line is considered to be busy if, when it is processed, the transmission buffer contains more than 3,180 bits.)

Variable length code assignments

Based on subjective evaluation of the distribution of differences in three pictures in the present study, the range of differences has been partitioned into three regions, small, medium, and large, and codes of 3, 6, and 10 bits, respectively, have been assigned to these regions. Small, medium, and large difference ranges are defined as $\pm(0, 3)$, $\pm(4, 31)$, and $\pm(32, 127)$, respectively, for quiet regions and $\pm(0, 6)$, $\pm(7, 41)$, and $\pm(42, 127)$, respectively, for busy regions of a picture. The regions and the codes are listed in Tables 11a and 11b.

TABLE 10. SIMULATION RESULTS: DITEC-PDP

Slide	Transmission Bit Rate (bits/pel)	Mode Threshold	Prediction	Average Bits/Pel (field)	Maximum Bits/Pel (line)	Maximum No. of Bits in Buffer	No. of Lines in Busy Mode
Zelda	4.0	3,180	Horizontal	3.747	4.47	1,323	0
Zelda	4.0	3,180	Horizontal/ Vertical	3.640	4.47	221	0
Beach	4.0	3,180	Horizontal	3.960	4.67	5,329	161
Beach	4.0	3,180	Horizontal/ Vertical	3.954	4.81	3,662	133
Beach	4.5	3,180	Horizontal	4.298	4.90	3,222	12
Beach	4.5	3,180	Horizontal/ Vertical	4.257	4.82	1,933	0
Net	4.0	3,180	Horizontal	4.168	5.39	18,580	237
Net	4.0	3,180	Horizontal/ Vertical	3.996	5.39	4,588	180
Net	4.5	3,180	Horizontal	4.450	5.39	3,395	85
Net	4.5	3,180	Horizontal/ Vertical	4.365	5.39	3,250	4

TABLE 11a. LOOK-UP TABLE, NORMAL MODE

Difference Range	Quantized Value	Code	Maximum Error	No. of Bits/Code
0	0	0 0 0	0	3
1	1	S ^a 0 1	0	3
2, 3	2	S 1 0	-1	3
4, 6	5	S 1 1 0 0 0	±1	6
7, 9	8	S 1 1 0 0 1	±1	6
10, 12	11	S 1 1 0 1 0	±1	6
13, 15	14	S 1 1 0 1 1	±1	6
16, 20	18	S 1 1 1 0 0	±2	6
21, 25	23	S 1 1 1 0 1	±2	6
26, 30	28	S 1 1 1 1 0	±2	6
31, 127	4 MSB of PCM + 4	1 1 1 1 1 1 X X X X ^b	±4	10

^a S = 0 for positive differences, 1 for negative differences.

^b XXXX are the four most significant bits of the corresponding PCM code.

TABLE 11b. LOOK-UP TABLE, BUSY MODE

Difference Range	Quantized Value	Code	Maximum Error	No. of Bits/Code
0	0	0 0 0	±0	3
1, 3	2	S ^a 0 1	±1	3
4, 6	5	S 1 0	±1	3
7, 11	9	S 1 1 0 0 0	±2	6
12, 16	14	S 1 1 0 0 1	±2	6
17, 21	19	S 1 1 0 1 0	±2	6
22, 26	24	S 1 1 0 1 1	±2	6
27, 31	29	S 1 1 1 0 0	±2	6
32, 36	34	S 1 1 1 0 1	±2	6
37, 41	39	S 1 1 1 1 0	±2	6
42, 127	4 MSB of PCM + 4	1 1 1 1 1 1 X X X X ^b	±4	10

^a S = 0 for positive differences, 1 for negative differences.

^b XXXX are the four most significant bits of the corresponding PCM code.

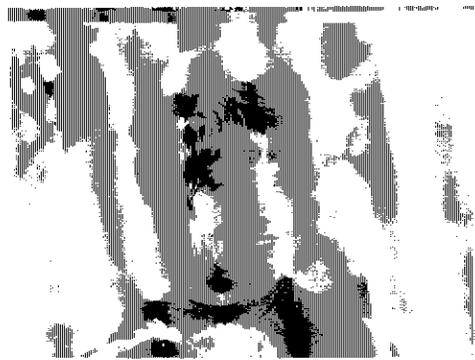
Computer simulations

Programs to simulate the proposed encoder have been implemented on a PDP 11/10 and an IBM 360/65 computer. Processing on the PDP computer, which is interfaced to the DITEC system, makes it possible to study the required bit rate and transmission buffer for data generated in the DITEC system. At present there is no way to display the reconstructed pictures on the DITEC monitor. To provide for subjective evaluation of the encoder, the pictures processed by the PDP are digitized on the flying spot scanner and processed by an IBM computer; the reconstructed pictures are then produced on the flying spot scanner.

Simulation of the DITEC-PDP

In the DITEC system, the NTSC composite color signal is decoded into its luminance (*Y*) and chrominance (*I* and *Q*) components. The luminance component is sampled at a rate of 6.018 MHz and digitized with 7 bits/sample. Three pictures are processed (Figure 5) using the parameters listed in Table 10. The companding laws and code assignments are given in Tables 11a and 11b.

The transition between normal and busy modes depends on a threshold specified by the user. The codes are variable in length; therefore, a buffer is required to permit a constant transmission bit rate. The user specifies



RECONSTRUCTED DPCM
TRANSMISSION RATE: 3.0 BITS/PEL

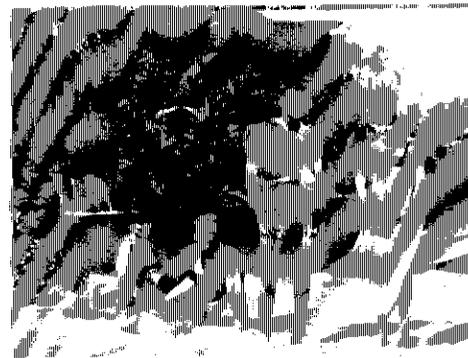


ORIGINAL 7-BIT PCM

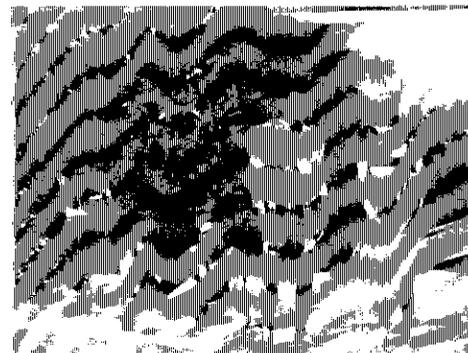


RECONSTRUCTED DPCM
TRANSMISSION RATE: 4.0 BITS/PEL

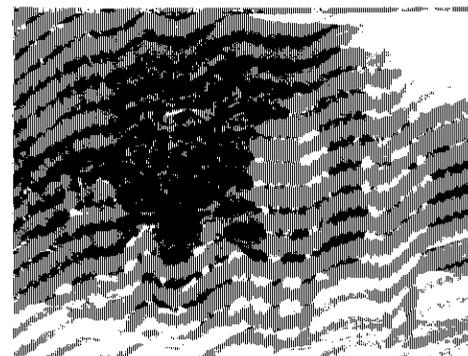
Figure 5a. *Zelda*



RECONSTRUCTED DPCM
TRANSMISSION RATE: 3.5 BITS/PEL

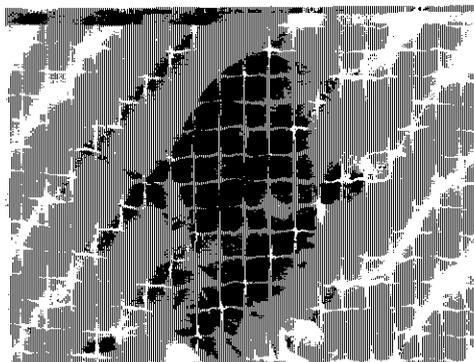


ORIGINAL 7-BIT PCM



RECONSTRUCTED DPCM
TRANSMISSION RATE: 4.0 BITS/PEL

Figure 5b. *Beach*



RECONSTRUCTED DPCM
TRANSMISSION RATE: 3.5 BITS/PEL



ORIGINAL 7-BIT PCM



RECONSTRUCTED DPCM
TRANSMISSION RATE: 4.0 BITS/PEL

Figure 5c. *Net*

the transmission bit rate (4.0 and 4.5 in the present experiments) as well as a mode threshold. The mode is normal as long as the number of bits in the buffer is less than the specified threshold (3,180 bits in these experiments). The encoder switches to the busy mode when the number of bits in the buffer is more than the specified threshold and switches back to the normal mode when the number of bits in the buffer is reduced below the specified threshold. Transitions between modes occur only between the lines.

Each scan line is partitioned into a number of segments specified by the user (10 in the experiments described herein). The coding for each segment is performed twice, based on horizontal and vertical prediction. The prediction that yields codes with a smaller average number of bits per pel is used.

The objective of the simulation was to study bit rate and buffer length requirements. Three images were processed, *Zelda*, *beach*, and *net* (Figure 5). Simulation results are summarized in Table 10. The average bit per pel entry refers to the average for the whole field, while the maximum bit per pel entry refers to the maximum of the average number of bits per pel for each line. Two significant results should be noted:

- a. A transmission bit rate of 4.0 bits/pel may require quite long buffering for busy images, e.g., 18,580 bits for the *net* scene when horizontal prediction is used. A transmission bit rate of 4.5 bits/pel is preferable.
- b. The advantage of the horizontal/vertical prediction is obvious. It reduces the length of the required buffer and also decreases the number of lines in the busy mode, hence reducing quantization noise (due to finer quantization of the difference values).

Simulation on the Scanner and IBM Computer

The encoder discussed in the previous subsection was implemented on an IBM 360/65 computer. The same three slides (*Zelda*, *beach*, and *net*) were digitized on a flying spot scanner into 512×512 seven-bit samples and recorded on magnetic tape. The digitized slides were processed on the computer and the reconstructed pictures were recorded on a magnetic tape which was played back on the scanner to produce photographic prints of the reconstructed images. The objective of this processing was to enable subjective evaluation of the reconstructed pictures.

The original and the reconstructed pictures are shown in Figure 5. There is no perceptible degradation between the original and the recon-

structed pictures, even when almost all of the picture is processed in the busy mode. (A white segment in the left-hand side of a line indicates that the next line is processed in the busy mode.)

Some input and output parameters of this simulation are summarized in Table 12. The scanner's resolution is higher than that of DITEC;

TABLE 12. INPUT AND OUTPUT PARAMETERS FOR THE SCANNER-IBM SIMULATION

Slide	Transmission Bit Rate (bits/pel)	Average Bits/Pel (frame)	Maximum Bits/Pel (line)	No. of Busy Lines	Required Buffer Size (lines)
Zelda	3.0	3.146	3.53	501	2
Zelda	4.0	3.647	4.19	0	2
Beach	3.5	3.508	4.39	355	2
Beach	4.0	3.946	4.46	21	2
Net	3.5	3.551	4.00	497	2
Net	4.0	3.934	4.71	13	2

therefore, the samples are more closely correlated, the differences are smaller, and the average bit rate is smaller. Here again, the average bits per pel refers to the average over the whole frame, and the maximum bits per pel refers to the maximum among the line average bits per pel.

Conclusions

Some algorithms for efficient encoding of monochrome television DPCM have been studied and an encoder using selective prediction and entropy coding has been defined and simulated. The encoder provides for the transmission of 37 distinct values, 21 DPCM and 16 PCM. The combination of selective prediction and entropy coding enables the transmission of those 37 distinct values with a bit rate of 4.5 bits per pel. Nonlinear quantization applied to the differential values yields reconstructed pictures of very good quality. At a transmission rate of 4.5 bits per pel, there is no perceptible degradation between the original and the reconstructed pictures. A transmission buffer of 4,000 bits has been found adequate to accommodate the digitized signal.

Acknowledgment

The author would like to thank several of his colleagues and in particular R. K. Garlow for many helpful discussions, H. Mueller for keeping the

DITEC PDP interface running, and B. Merrihew for skillful operation of the scanner.

References

- [1] L. S. Golding, "DITEC—A Digital Television Communications System for Satellite Links," *2nd International Conference on Digital Satellite Communications*, Paris, France, November 28–30, 1972.
- [2] L. S. Golding, and R. K. Garlow, "Frequency Interleaved Sampling of a Color Television Signal," *IEEE Transactions on Communications*, COM-19, No. 6, December 1971, pp. 972–979.
- [3] J. B. O'Neal, Jr., "Predictive Quantizing Systems (DPCM) for the Transmission of Television Signals," *Bell System Technical Journal*, May-June, 1966, pp. 699–721.
- [4] C. C. Cutler, "Differential Quantizing of Communication Signals," Patent No. 2,605,361, July 29, 1952 (applied for June 29, 1950).
- [5] P. Kaul and L. S. Golding, "A DPCM Codec Using Edge Coding and Line Replacement," *1972 National Telecommunications Conference Record*, December 1972, pp. 27 E-1 to 27 E-6.
- [6] J. C. Candy and R. H. Bosworth, "Methods for Designing Differential Quantizers Based on Subjective Evaluation of Edge Busyness," *Bell System Technical Journal*, September 1972, pp. 1495–1516.
- [7] E. F. Brown, "A Sliding Scale Direct Feedback PCM Coder for Television," *Bell System Technical Journal*, May-June 1969, pp. 1537–1553.
- [8] J. O. Limb, "Adaptive Encoding of Picture Signals," *Picture Bandwidth Compression*, edited by T. S. Huang and O. J. Tretiak, Gordon and Breach, 1972, pp. 341–382.
- [9] J. B. O'Neal, Jr., "Entropy Coding in Speech and Television DPCM Systems," *IEEE Transactions on Information Theory*, IT-7, November 1971, pp. 758–761.
- [10] S. K. Goyal, and J. B. O'Neal, Jr., "Entropy Coded Differential Pulse Code Modulation Systems for Television," *Proceedings of the National Telecommunications Conference*, San Diego, California, December 2–4, 1974, pp. 72–76.
- [11] H. Meyr, H. G. Rosdolsky, and T. S. Huang, "Optimum Run Length Codes," *IEEE Transactions on Communications*, COM-22, No. 6, June 1974, pp. 826–835.
- [12] D. A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes," *Proceedings of IRE*, September 1952, pp. 1098–1101.
- [13] L. S. Golding and P. M. Schultheiss, "Study of an Adaptive Quantizer," *Proc. of IEEE*, Vol. 55, No. 3, March 1967, pp. 293–297.



Its'hak Dinstein received a B.S.E.E. from the Technion, Israel, in 1965, 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 and is

presently a member of the technical staff in the Image Processing Department of the Communications Processing Laboratory. He is a member of IEEE.

Index: Gunn effect, microwave oscillators, Maxwell's equations, models

Cylindrical 3-dimensional model of a Gunn oscillator

V. E. RIGINOS AND I. G. ATOHOUN

(Manuscript received March 10, 1975)

Abstract

This paper presents a 3-dimensional model of an encapsulated X-band Gunn oscillator operating in a coaxial microwave cavity. The physical behavior of the active device, package, and cavity are described by Maxwell's equations and by the velocity-field characteristic of the semiconductor material. The equations have been solved numerically, in cylindrical coordinates, by means of finite differences. All of the boundaries, as well as the surface charges on the semiconductor, have been considered. By means of the model, the effects of the surface charges, the period of oscillation (including domain dissolution and renucleation processes and their effects on the conduction current waveform), and the RF voltage within the cavity have been analyzed.

It is well known from previous experimental results that Gunn oscillators produce several frequency components above the fundamental, that there are strong cavity-device and package-device interactions, and finally that under certain operating conditions high-field domains are trapped at the device anode. All of these phenomena are quantitatively predicted, for the first time, by using the 3-dimensional model described herein.

This paper is based on work which began at the Stevens Institute of Technology and has continued at COMSAT Laboratories. It was partially supported by the International Telecommunications Satellite Organization (INTELSAT). Views expressed in this paper are not necessarily those of INTELSAT.

Introduction

As in the case of other transit-time devices, the difficulties inherent in the analysis of Gunn-effect devices render it difficult to predict the exact electrical behavior of the device, particularly when it is encapsulated or in a waveguide cavity. The need for more detailed information concerning the prediction of the electrical behavior of semiconductor oscillators necessitates the development of a highly refined model of these microwave power sources.

This paper is an attempt to add to the available knowledge of the theory of operation of transit-time devices in general and Gunn-effect devices in particular. In microwave applications, Gunn-effect devices usually operate in electromagnetic cavities. Under such operating conditions, cavity-controlled oscillations coupled with microwave power are generally obtained. The degree to which microwave power is generated from these devices depends upon the electric characteristics of both the semiconductor and the cavity. If proper electrical characteristics are desired, a thorough understanding of the interactions between the device and its cavity is therefore essential.

Such electrical characteristics can be determined from a careful study of a unified model of the device and the resonant cavity. The model can be generated from an analysis of the device in the appropriate coordinate system. The coordinate system is, on the one hand, related to the physical structure of the device and, on the other hand, to the geometry of the cavity and/or the device package.

For most applications in the X-band frequency range, the devices are usually cylindrical, encapsulated in a package consisting of a hollow dielectric cylinder with metallic end caps, and operating within a microwave cavity. Thus, due to the geometry of both device and package, a cylindrical coordinate system has been used in this model. A coaxial cavity has been chosen to simplify the calculations.

A method for numerically determining the effects of an electromagnetic cavity on microwave oscillations in Gunn-effect devices is also presented. The effects of surface charges and the dielectric package on the electrical performance of Gunn devices are investigated by including these charges and dielectric interfaces as part of the boundary conditions imposed on the device.

This paper is intended to present a general technique for 3-dimensional analysis of transit-time oscillators, to show the effects of finite boundaries on the electrical behavior of Gunn-effect devices, and, finally, to test the validity of the 1-dimensional approach.

Previous work

Since Gunn's [1] discovery of microwave power generation from a DC-biased bulk gallium arsenide (GaAs) diode, various mathematical and experimental studies, as well as computer calculations of this effect, have been presented. These studies can generally be divided into two categories: those concerning the analysis of the dynamics of Gunn-effect devices, and those primarily concerned with the analysis of the device cavity-mode operation in terms of lumped parameter cavity models.

Butcher et al. [2] presented a 1-dimensional analysis of stable domain propagation. By means of a 1-dimensional finite difference technique, Thim [3] studied the effects of random doping fluctuations in bulk *n*-type GaAs semiconductors. Shoji [4] extended Butcher's analysis to the study of 2-dimensional domain dynamics. Copeland [5] developed the small perturbation theory of the growth of 2- and 3-dimensional space charge. The lateral spreading velocity of a dipole domain and a 2-dimensional space-charge growth were investigated by Suzuki et al. [6].

Various other authors [7]–[14] investigated the effects of circuit properties on Gunn-effect oscillators and presented diverse lumped models of Gunn devices operating in electromagnetic cavities. The small signal equivalent circuit of Gunn-effect devices was derived by Hobson [15], who assumed stable domain propagation. Robrock [16]–[18] studied the effects of device parameters on high-field domain dynamics and extended his analysis to include the derivation of the device lumped models and the study of domain absorption at the anode. Recently, Gunshor [19] introduced a new lumped circuit representation which encompasses both small and large signal operation of Gunn-effect devices. In addition, many authors [20]–[24] have reported the effects of a dielectric coat or load and the presence of transverse fields on stable domain formation in Gunn-effect devices.

Device physics

The Gunn effect is explained by the Ridley-Watkins-Hilsum theory [25], [26]. According to this theory, when a suitable semiconductor sample is subjected to a high-electric field, some of the conduction electrons are transferred from a low-energy, high-mobility minimum of the energy band to the high-energy, low-mobility minima of the band. This electron transfer reduces the overall electron mobility of the sample and gives rise to a negative differential resistivity (NDR) region.

When the material is biased in the NDR region, the electric field, due to the cathode discontinuity, breaks up into distinct high- and low-field regions within the sample [25]. The most commonly observed field configuration is a high-field domain associated with a moving space charge dipole. This space charge, which nucleates at the cathode discontinuity, travels along the length of the sample at approximately the electron drift velocity and is absorbed at the anode. As the old domain is being absorbed, a new domain is nucleated near the cathode and the cycle is repeated, giving rise to the familiar Gunn oscillations.

The electronic transport within III-V semiconductors which exhibit the Gunn effect (such as GaAs and InP) can be described by means of Maxwell's equations and the average electron drift velocity versus electric field characteristics. The velocity-field characteristic for a given material can be calculated either by solving the Boltzmann equation or by using Monte Carlo methods. However, this calculation is a formidable task since it involves the solution of n -coupled nonlinear integrodifferential equations, where n is the number of valleys in the energy band [27]. The difficulty is compounded by the lack of precise knowledge concerning several coefficients of the relevant scattering mechanisms [28]. The characteristic for GaAs, the most widely studied material, has been calculated in this fashion by Butcher and Fawcett [29], Heinle [30], Broadman, Fawcett, Ruch, and others [31]-[35]. It has been accurately measured for fields up to 14 kV/cm by Ruch and Kino [36] and indirectly determined for higher fields [37], [38]. Figure 1 summarizes these results.

Based on the velocity-field relationship, a macroscopic description of the charge and field distribution within the encapsulated semiconductor device and the microwave cavity is given by the Maxwell equations

$$\vec{\nabla} \times \vec{H} = \vec{J}_c + \frac{\partial \vec{D}}{\partial t} \quad (1)$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (2)$$

$$\vec{\nabla} \cdot \vec{B} = 0 \quad (3)$$

$$\vec{\nabla} \cdot \vec{D} = \rho \quad (4)$$

The charge density within an n -type semiconductor is given by

$$\rho = -e[n(E) - N] \quad (5)$$

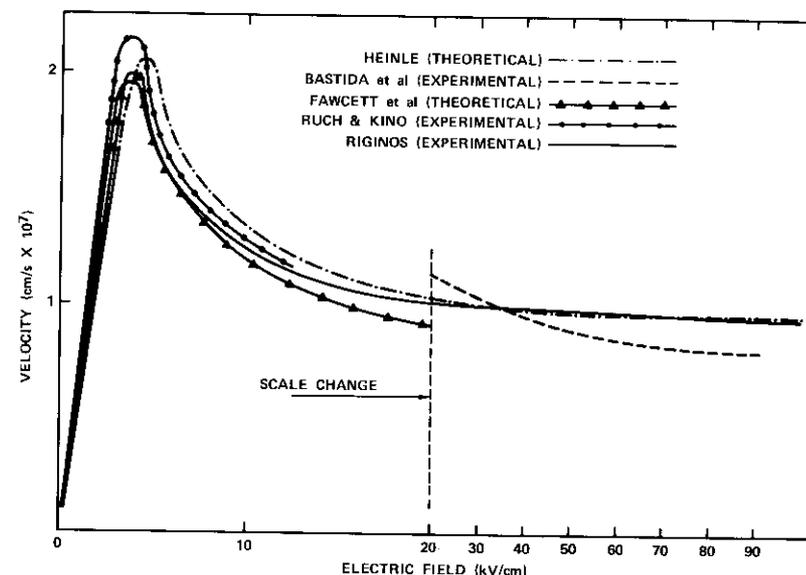


Figure 1. Velocity-Field Characteristic of GaAs

where $n(E)$ is the field-dependent carrier concentration, e is the electronic charge, and N is the doping density.

The Maxwell equations can be solved by means of a vector potential \vec{A} and a scalar potential Φ such that

$$\vec{E} = -\vec{\nabla}\Phi - \frac{\partial \vec{A}}{\partial t} \quad (6)$$

subject to the Lorentz condition [39]

$$\vec{\nabla} \cdot \vec{A} + \mu\epsilon \frac{\partial \Phi}{\partial t} = 0 \quad (7)$$

where μ is the absolute magnetic permeability and ϵ the dielectric permittivity of the material. For an isotropic medium where

$$\vec{D} = \epsilon \vec{E}$$

the combination of equations (4)-(7) results in

$$\nabla^2 \Phi - \mu \epsilon \frac{\partial^2 \Phi}{\partial t^2} = \frac{e}{\epsilon} [n(E) - N] \quad (8)$$

Equation (1) can be transformed into the current continuity equation

$$\vec{\nabla} \cdot \vec{J}_c + \frac{\partial \rho}{\partial t} = 0 \quad (9)$$

where \vec{J}_c is the conduction current. For the semiconductor material under consideration, the conduction current is given by

$$\vec{J}_c = -e \{ n(E) \vec{V}(E) - \vec{\nabla} [D(E) n(E)] \} \quad (10)$$

where $D(E)$ is the diffusion coefficient and $\vec{V}(E)$ is the average electron drift velocity, which has a magnitude given by the velocity field characteristic and points in the same direction as the electric field. Combination of equations (5), (9), and (10) results in

$$\frac{\partial n(E)}{\partial t} = \vec{\nabla} \cdot \{ \vec{\nabla} [D(E) n(E)] - n(E) \vec{V}(E) \} \quad (11)$$

Simultaneous solution of equations (8) and (11) under the appropriate boundary conditions determines the scalar potential and the carrier concentration distributions. The boundary conditions for the electric field are

$$\hat{n} \times (\vec{E}_2 - \vec{E}_1) = 0 \quad (12)$$

$$\hat{n} \cdot (\epsilon_2 \vec{E}_2 - \epsilon_1 \vec{E}_1) = \sigma \quad (13)$$

where \hat{n} is a unit vector pointing from medium 1 to medium 2, and σ is the surface charge. Similarly, for the conduction current the boundary condition is given by [40]

$$\hat{n} \cdot \vec{J}_c = - \frac{\partial \sigma}{\partial t} \quad (14)$$

Mathematical model

Figure 2 is a photomicrograph of the cross section of a Varian Associates VSX-9201 Gunn diode and its package. Figure 3, on which

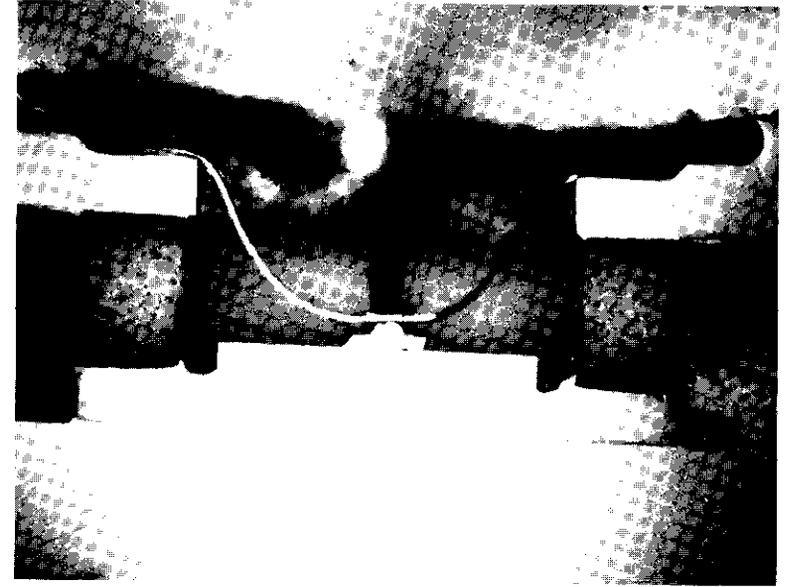


Figure 2. Photomicrograph of the Packaged Gunn Diode

the model is based, is a schematic representation of this device and package geometry within a coaxial microwave cavity. In this figure region I is the semiconductor active region, region II is the package, region III the cavity, and regions IV and V are the anode and cathode electrodes, respectively.

Both device and cavity possess circular symmetry and, although the package is not symmetric due to the bond wire, it has also been assumed to possess circular symmetry to conserve computer storage. Hence, there is no ϕ dependence. The specific equations applicable to these various regions are discussed below. In region I equations (8) and (11) become

$$\nabla^2 \Phi_I(r, z, t) - \mu_I \epsilon_I \frac{\partial^2 \Phi_I}{\partial t^2} = \frac{e}{\epsilon_I} [n(r, z, t) - N(z)] \quad (15)$$

and

$$\frac{\partial n(r, z, t)}{\partial t} = \vec{\nabla} \cdot [D \vec{\nabla} n(r, z, t) - n(r, z, t) \vec{V}(r, z, t)] \quad (16)$$

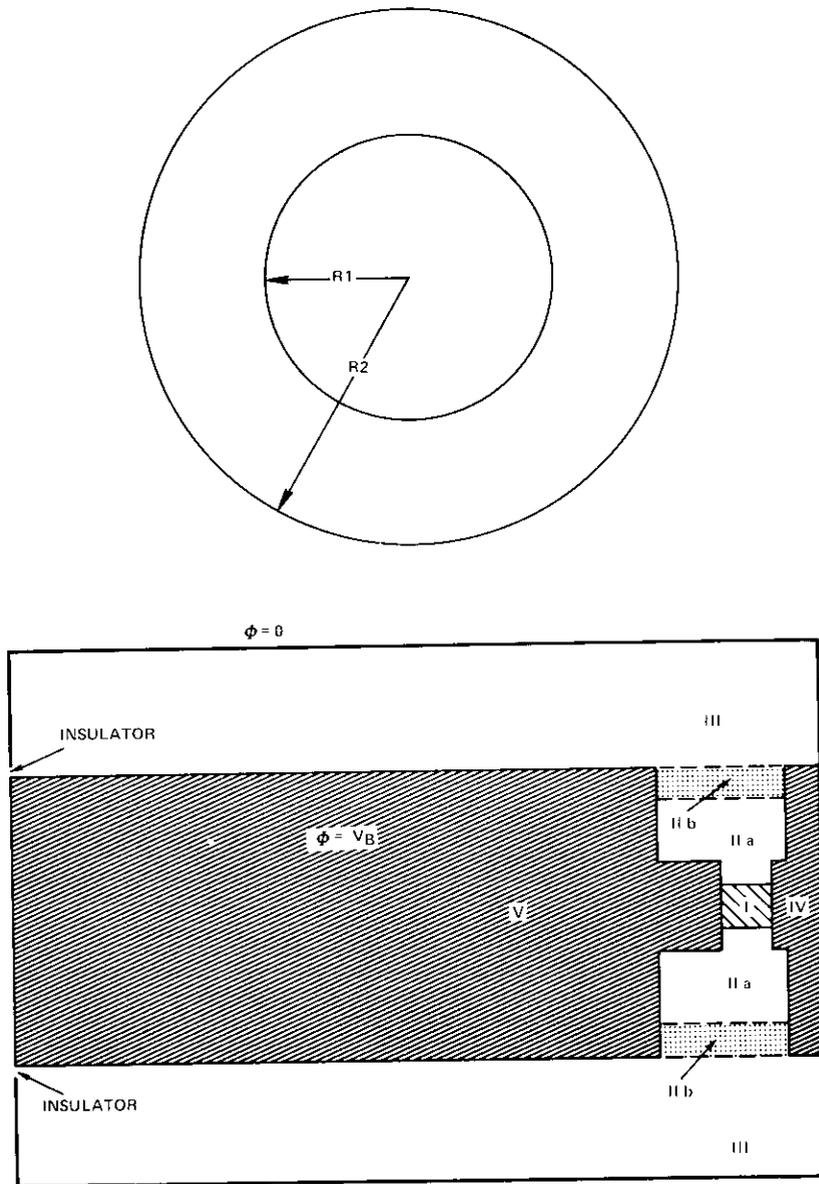


Figure 3. Encapsulated Gunn-Effect Device Operating in a Coaxial Microwave Cavity

where D is the diffusion coefficient, which for this model has been assumed to be field independent.

Regions II and III are charge free; hence equations (8) and (11) reduce to

$$\nabla^2 \Phi_\alpha(r, z, t) - \mu_\alpha \epsilon_\alpha \frac{\partial^2 \Phi_\alpha}{\partial t^2} = 0, \quad \alpha = \text{II, III} \quad (17)$$

It should be noted that the package (region II) consists of air and a ceramic wall which are designated as regions IIa and IIb, respectively.

If it is assumed that the electrodes consist of perfect conductors, equations (12) and (13) become

$$ER_r(r, 0, t) = ER_r(r, L, t) = 0 \quad (18)$$

$$EZ_z(r, 0, t) = \sigma(r, 0, t) \quad (19)$$

$$EZ_z(r, L, t) = -\sigma(r, L, t) \quad (20)$$

where L is the length of the semiconductor device, and ER and EZ are the radial and axial components of the \vec{E} -field. From equations (12)–(14),

$$\epsilon_{II} ER_{II}(R, z, t) - \epsilon_I ER_I(R, z, t) = \sigma(R, z, t) \quad (21)$$

$$EZ_I(R, z, t) = EZ_{II}(R, z, t) \quad (22)$$

and

$$JR_C(R, z, t) = \frac{\partial \sigma(R, z, t)}{\partial t} \quad (23)$$

where R is the radius of the semiconductor and JR_C is the radial component of the conduction current density which is given by equation (10).

At the interface between regions IIa and IIb, as well as that between IIb and III, the surface charge is zero. Also, the potential at the cavity walls and at the anode (region IV) is zero, while at the cathode (region V) it is equal to the applied bias voltage.

Results

Equations (15)–(23) were solved numerically for the semiconductor device and its package. Regions I and II were represented by a grid of

points, the equations were replaced by their 5-point finite difference form, and the ensuing linear equations were solved by using the Jacobi method.

In region I (the semiconductor), both equations (15) and (16) are applicable. Equation (15) is numerically unconditionally stable [41], while equation (16), which is a parabolic equation, is only conditionally stable. The stability conditions [3]

$$\left(\frac{\Delta t}{G}\right) V \ll 1 \tag{24}$$

$$\frac{D\Delta t}{G^2} \ll 1 \tag{25}$$

together with the condition for dipole domain formation [12],

$$n_0 L > 10^{12} \text{ cm}^{-2} \tag{26}$$

where n_0 and L are the doping density and device length, respectively, specify the doping density, grid size, and time step for an oscillator operating at a given frequency.

The package (region II) is about 50 times larger than the device. To conserve computer storage, an expanding grid was utilized rather than the fixed grid size necessary for the device. Even with an expanding grid in the microwave cavity (region III), which is about three orders of magnitude larger than the package, an enormous amount of computer storage would be required. Since an analytic solution of equation (17) for the cavity can be found, a Green's function technique was used. Details of the solution of equations (15)–(23) in all three regions are presented in the appendix.

The modeled device was $9.6 \mu\text{m}$ long and its radius was $18 \mu\text{m}$. The doping concentration was 3×10^{15} . The package consisted of a hollow dielectric cylinder with a height of $544 \mu\text{m}$ and inner and outer radii of 680 and $952 \mu\text{m}$, respectively (Figure 3). The microwave cavity was 1.55 cm long with a radius of 0.2 cm. From inequalities (24) and (25), the grid size and the time step were selected as $0.3 \mu\text{m}$ and 0.1 ps.

The time evolution of the device potential is shown in Figure 4. The domain dissolution process (at $t > 50$ ps), which is characterized by a reduction of the domain voltage with a concomitant increase in the outside domain electric field, can be seen in Figure 5. As the outside potential

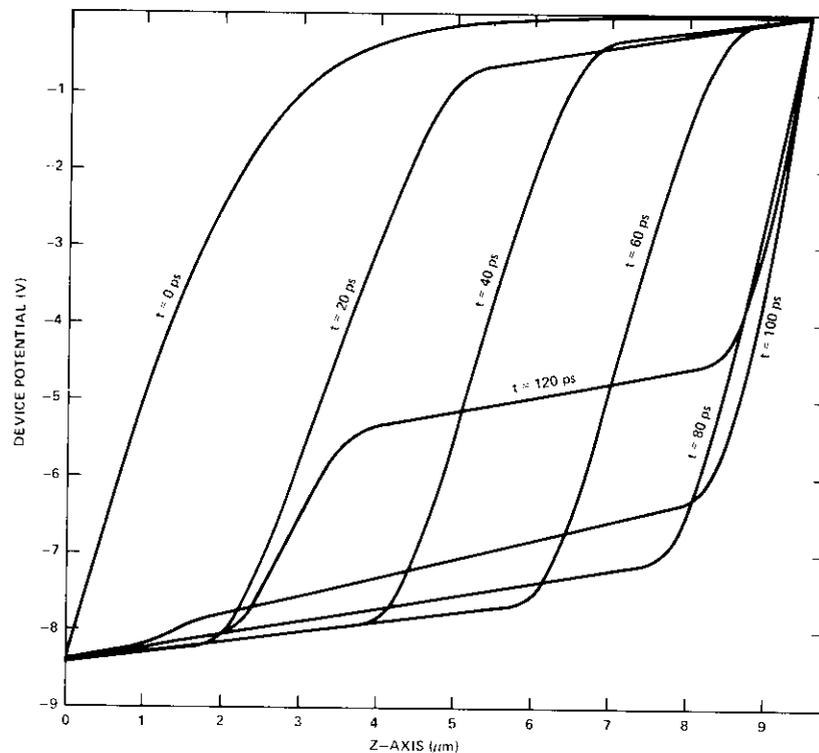


Figure 4. Time Evolution of the Device Potential

increases, the domain potential defined by

$$V_d = V_{bias} - E_0 L$$

where L is device length and E_0 the outside field, decreases to maintain the bias voltage constant. With the increase in the outside potential, the electric field increases as shown in Figure 5. The presence of two different domains concurrently can be seen in Figures 4 and 5. At $t > 50$ ps the domain near the cathode is initiated during the dissolution process of that at the anode. As the potential continues to rise, the peak field of the new domain increases; when its value exceeds the Gunn threshold, the new domain begins to propagate toward the anode. At $t = 136$ ps, the old domain will be completely dissolved while the new one will have reached its steady-state value and will have traveled $3.9 \mu\text{m}$ away from the cathode.

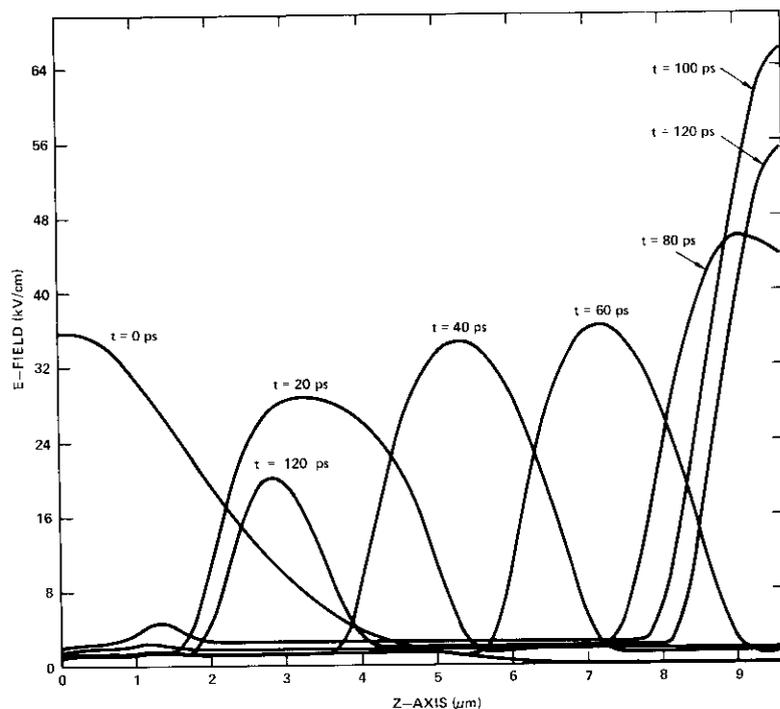


Figure 5. Time Evolution of the Domain Peak Electric Field at the Center Axis

The current can be evaluated by integrating the conduction current density [equation (10)] over the device cross section. The conduction current waveform is shown in Figure 6. The period of this waveform is 106 ps, which is the sum of the nucleation, transit, and dissolution times. Employing the usual approximation of the period,

$$T = \frac{L}{v_s}$$

where $v_s = 10^7$ cm/s is the saturation velocity, would result in a predicted period of 96 ps, which would be in error by 10 percent. In addition, two distinct oscillations of the waveform can be seen. These are associated with the dissolution and nucleation of the dipole domain. The frequency components of the current waveform which are higher than the fundamental are a direct result of this dipole dissolution and renucleation.

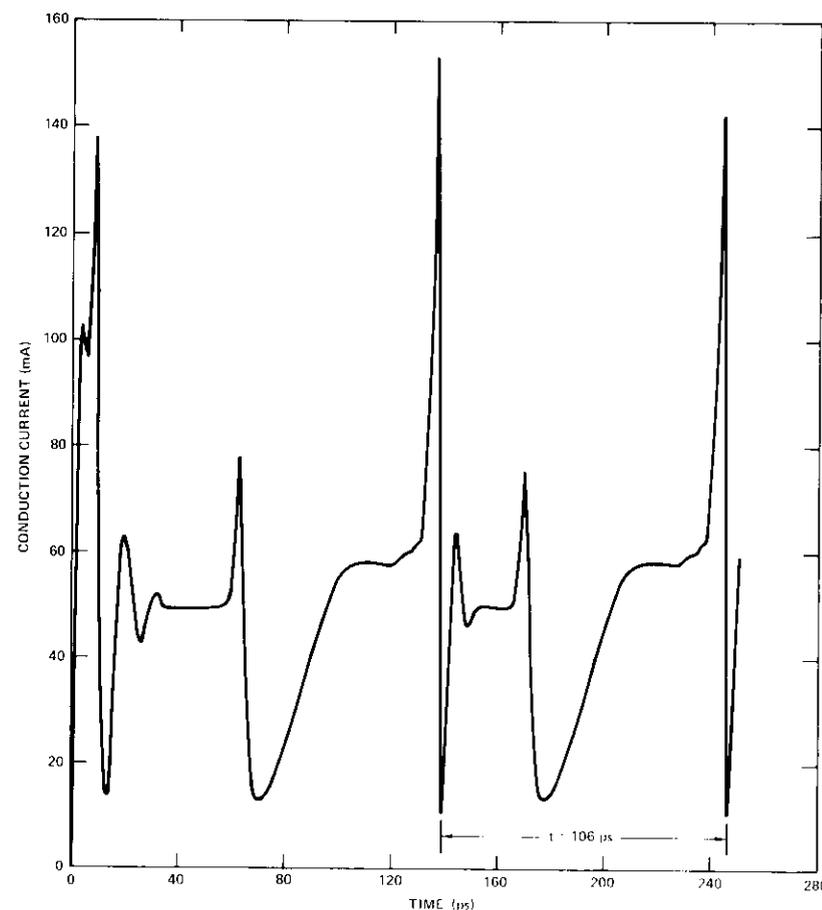


Figure 6. Conduction Current Waveform

Due to the geometry of the device, the electric field consists of two components, radial and axial. At the cylindrical surface there exist surface charges which oppose the flow of current outside the device. Figure 7 shows the net surface charge density at $t = 50$ ps. These charges cause a local rise of the radial field which, in turn, raises the overall electric field at the cylindrical surface boundary. The magnitude of the electric field distribution at $t = 50$ ps is shown in Figure 8. As can be seen from the figure, the magnitude of the field rises abruptly at the surface while the waveform curves toward the anode. This increase amounts to

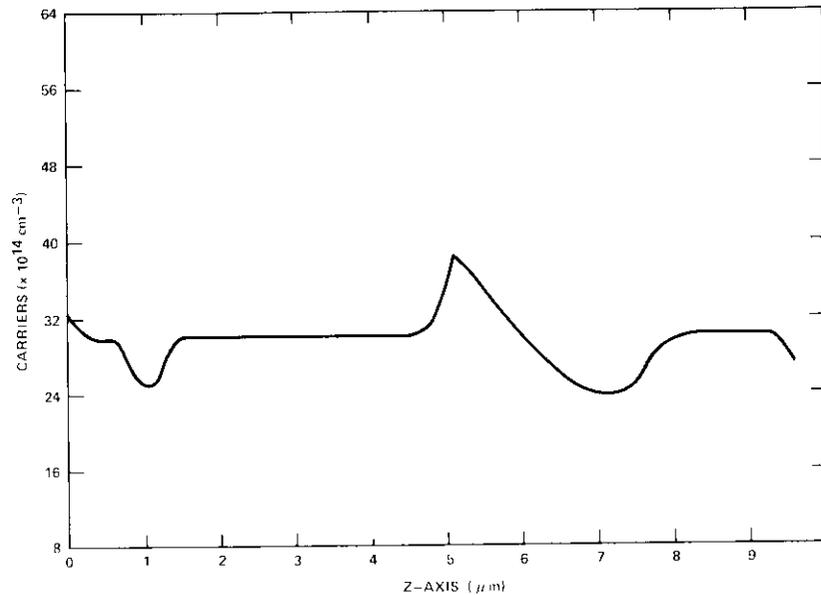


Figure 7. Carrier Density

about 9 percent as compared to the magnitude of the electric field at the central axis.

Figure 9 shows the potential distribution inside the microwave cavity at 120 ps. The value of the maximum RF potential within the cavity is shown as a function of time in Figure 10. The point of maximum potential is located at $L/2$, where L is the length of the cavity, and at three device lengths from the cavity wall. The model has been used to quantitatively investigate cavity-device and capsule-device interactions, as well as domain trapping. These results are presented elsewhere.

Conclusions

A 3-dimensional model of a cylindrical X-band Gunn oscillator has been developed. Incorporation of the device surface charge into the boundary conditions imposed on the device increases the radial component of the domain electric field at the surface boundary and hence affects the magnitude of the field along the interface. A comparison of the magnitudes of the peak electric field along the center axis and along the surface interface indicates that this increase is about 9 percent. Consequently, a sta-

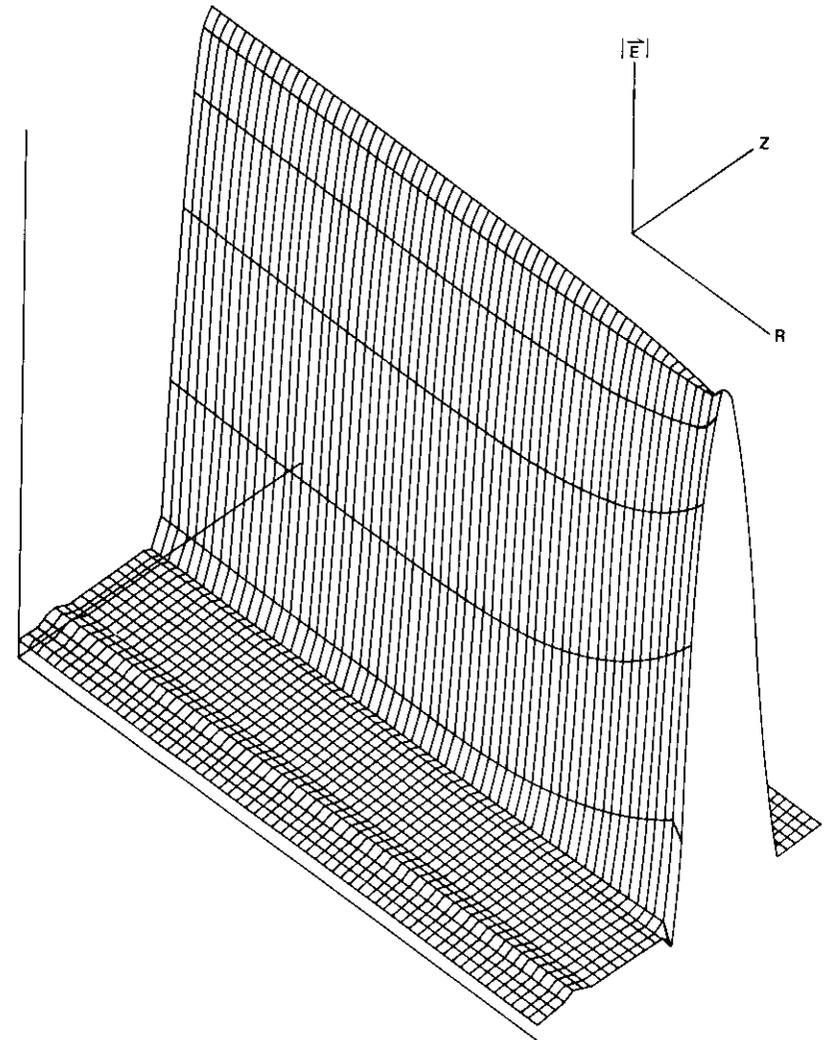


Figure 8. Domain Profile

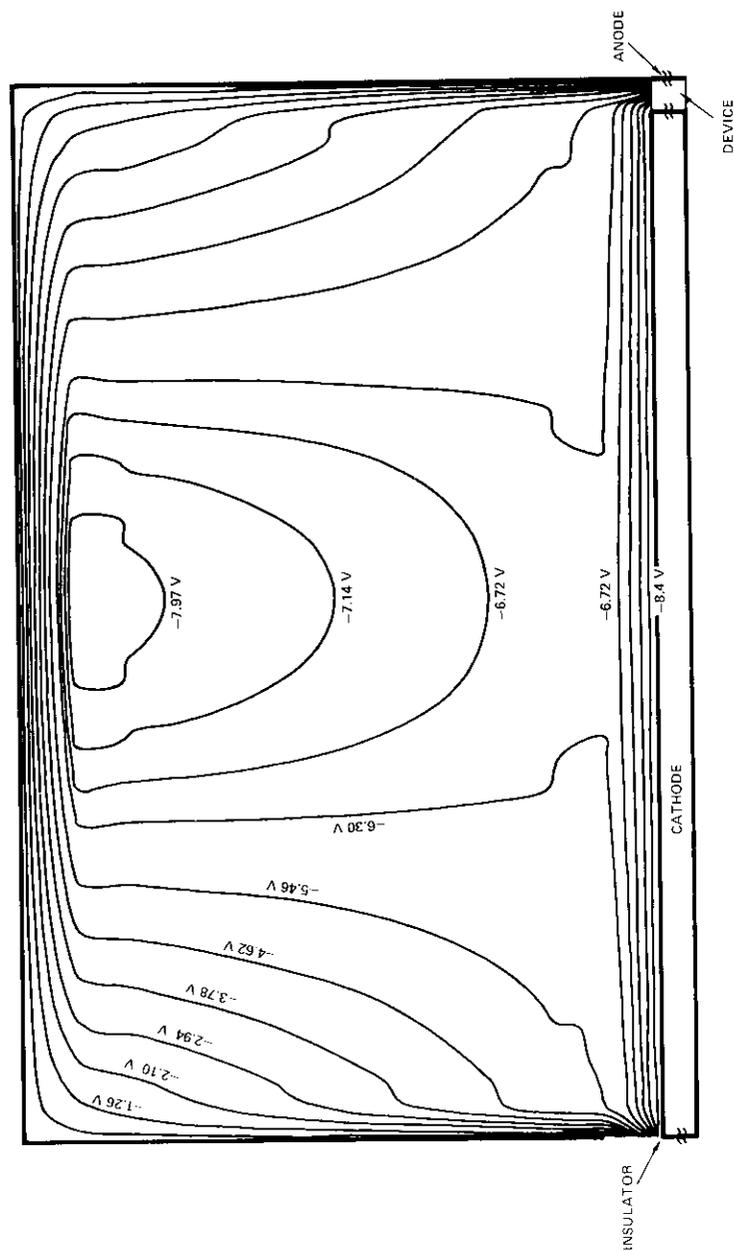


Figure 9. Potential Distribution in the Cavity

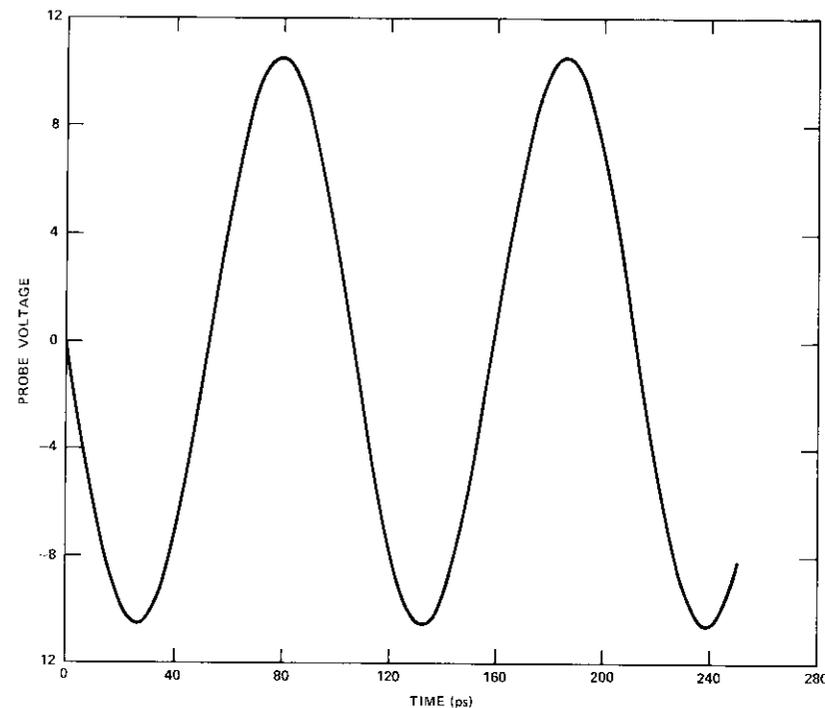


Figure 10. RF Voltage at a Grid Point Close to the Cavity Wall

tionary high-field domain can result at the anode, hence preventing Gunn oscillations [22]. For Gunn oscillations to prevail, a sufficiently large doping fluctuation is required. In the present investigation it has been found that a 20-percent doping notch is sufficient.

This increase (9 percent) in the domain peak field cannot be predicted from a 1-dimensional formulation which assumes an infinite periodic structure. In such a formulation Neumann boundary conditions are imposed at the device surface. The increase in the radial field is caused by carrier diffusion from the device center toward the boundary. This effect cannot be predicted by using Neumann conditions.

The model also predicted a reduction of about 23.5 percent in the radial electric field when the device package was included in the boundary conditions imposed on the device. Such a reduction indicated that the device package perturbs the electric field.

The exact device current waveform has also been predicted by the model.

The waveform has high-frequency components and a period of oscillation of 106 ps for a device length of 9.6 μm . This period consists of both domain nucleation and dissolution times, as well as domain transit time. Knowledge of the exact period of oscillation is extremely important in the design of the electromagnetic cavity in which the device is to be operated since it will permit the designer to avoid using a tuner which is usually necessary.

Acknowledgment

The authors wish to thank Dr. P. P. Bohn of the Stevens Institute of Technology and Mr. F. Assal of COMSAT Laboratories for helpful technical suggestions. They also wish to thank Mr. T. D. Kirkendall of COMSAT Laboratories for the preparation of Figure 2, and the staff of the COMSAT Computer Center for their help in programming.

References

- [1] J. B. Gunn, "Microwave Oscillations of Current in III-V Semiconductor," *Solid State Communications*, Vol. 1, September 1963, pp. 88-91.
- [2] P. N. Butcher, W. Fawcett, and N. R. Ogg, "Effect of the Field-Dependent Diffusion on Stable Domain Propagation in the Gunn Effect," *British Journal of Applied Physics*, Vol. 18, June 1967, pp. 755-759.
- [3] H. W. Thim, "Computer Study of Bulk GaAs Devices with Random One-Dimensional Doping Fluctuations," *Journal of Applied Physics*, Vol. 39, July 1968, pp. 3897-3904.
- [4] M. Shoji, "Two-Dimensional Gunn-Domain Dynamics," *IEEE Transactions on Electron Devices*, ED-16, No. 9, September 1969, pp. 748-758.
- [5] J. A. Copeland, "Growth of Two- and Three-Dimensional Space-Charge from Negative Differential Resistivity," *Journal of Applied Physics*, Vol. 39, October 1968, pp. 5103-5105.
- [6] N. Suzuki, H. Yanai, and T. Ikoma, "Simple Analysis and Computer Simulation on Lateral Spreading Space Charge in Bulk GaAs," *IEEE Transactions on Electron Devices*, ED-19, No. 3, March 1972, pp. 364-375.
- [7] K. R. Freeman and G. S. Hobson, "A Survey of CW and Pulsed Gunn Oscillators by Computer Simulation," *IEEE Transactions on Electron Devices*, ED-20, No. 10, October 1973, pp. 891-903.
- [8] Y. Ito et al., "Experimental and Computer Simulation Analysis of a Gunn Diode," *IEEE Transactions on Microwave Theory and Techniques*, MTT-19, No. 12, December 1971, pp. 900-905.
- [9] J. A. Copeland, "Theoretical Study of a Gunn Diode in a Resonant Circuit," *IEEE Transactions on Electron Devices*, ED-14, No. 2, February 1967, pp. 55-58.
- [10] K. V. Kuznetsov, M. E. Levinstein, and M. S. Shur, "The Investigation of the Gunn Diode Operation in a Resonator using a Computer Model," *Solid-State Electronics*, Vol. 14, March 1971, pp. 207-220.
- [11] P. N. Butcher and W. Fawcett, "Stable Domain Propagation in the Gunn Effect," *British Journal of Applied Physics*, Vol. 17, November 1966, pp. 1425-1432.
- [12] H. Kroemer, "Nonlinear Space-Charge Domain Dynamics in a Semiconductor with Negative Differential Mobility," *IEEE Transactions on Electron Devices*, ED-13, No. 1, January 1966, pp. 27-40.
- [13] D. E. McCumber and A. G. Chynoweth, "Theory of Negative Conductance Amplification and of Gunn Instabilities in 'Two Valley' Semiconductors," *IEEE Transactions on Electron Devices*, ED-13, No. 1, January 1966, pp. 4-21.
- [14] R. B. Robrock, II, "Analysis and Simulation of Domain Propagation in Nonuniformly Doped Bulk GaAs," *IEEE Transactions on Electron Devices*, ED-16, No. 7, July 1969, pp. 647-653.
- [15] G. S. Hobson, "The Equivalent Circuit of a Gunn Effect Device," *International Proceedings on Microwave and Optical Generation and Amplification*, Cambridge, England, September 1966, pp. 314-318.
- [16] R. B. Robrock, II, "Effects of Device Parameters on Domain Dynamics in Bulk GaAs," *Proc. IEEE*, Vol. 58, May 1970, pp. 804-805.
- [17] R. B. Robrock, II, "A Lumped Model for Characterizing Single and Multiple Domain Propagation in Bulk GaAs," *IEEE Transactions on Electron Devices*, ED-17, No. 2, February 1970, pp. 93-102.
- [18] R. B. Robrock, II, "Extension of the Lumped Bulk Device Model to Incorporate the Process of Domain Dissolution," *IEEE Transactions on Electron Devices*, ED-17, No. 2, February 1970, pp. 103-107.
- [19] R. L. Gunshor and A. C. Kak, "Lumped-Circuit Representation of Gunn Diodes in Domain Mode," *IEEE Transactions on Electron Devices*, ED-19, No. 6, June 1972, pp. 765-770.
- [20] R. W. H. Engelmann, "On the Transverse Surface Boundary Effect in Gunn Devices," *Proc. IEEE*, Vol. 57, May 1967, pp. 818-819.
- [21] G. S. Kino and P. N. Robson, "The Effect of Small Transverse Dimensions on the Operation of Gunn Devices," *Proc. IEEE*, Vol. 56, November 1968, pp. 2056-2057.
- [22] K. Kumabe, "Suppression of Gunn Oscillations by a Two-Dimensional Effect," *Proc. IEEE*, December 1968, pp. 2172-2173.
- [23] H. L. Hartnagel, "Gunn Instabilities with Surface Loading," *Electronics Letters*, Vol. 5, July 1969, pp. 303-304.
- [24] S. Kataoka, H. Tatenno, and M. Kawashima, "Suppression of High-Field-Domain Mode Oscillations in GaAs by Dielectric Surface Loading," *Electronics Letters*, Vol. 5, February 1969, pp. 48-50.

- [25] B. K. Ridley and T. B. Watkins, "The Possibility of Negative Resistance Effects in Semiconductors," *Proc. Physics Society* [London], Vol. 78, August 1961, pp. 293-304.
- [26] C. Hilsum, "Transferred Electron Amplifiers and Oscillators," *Proc. IRE*, Vol. 50, February 1962, pp. 185-189.
- [27] E. M. Conwell, "High Field Transport in Semiconductors," *Solid State Physics*, Supplement 9, New York: Academic Press, 1967.
- [28] V. Riginos, "Determination of the Velocity-Field Characteristics of Gallium Arsenide from High-Field Domain Measurements," Ph.D. Dissertation, Stevens Institute of Technology, April 1973.
- [29] P. N. Butcher and W. Fawcett, "Calculation of the Velocity-Field Characteristics for Gallium Arsenide," *Physics Letters*, Vol. 21, June 1966, pp. 489-490.
- [30] W. Heinle, "Displaced Maxwellian Calculation of Transport in n-Type GaAs," *Physical Review*, Vol. 178, February 1969, pp. 1319-1325.
- [31] W. Fawcett, A. D. Boardman, and S. Swain, "Monte Carlo Determination of Electron Properties in Gallium Arsenide," *Journal of Physics and Chemistry of Solids*, Vol. 31, 1970, pp. 1963-1990.
- [32] J. G. Ruch and W. Fawcett, "Temperature Dependence of the Transport Properties of Gallium Arsenide Determined by a Monte Carlo Method," *Journal of Applied Physics*, Vol. 41, August 1970, pp. 3843-3849.
- [33] E. M. Conwell and M. O. Vassel, "High-Field Transport in n-Type GaAs," *Physical Review*, Vol. 166, February 1968, pp. 797-821.
- [34] H. Hillbrand, "Monte Carlo Calculation of NDR in Gallium Arsenide," *Physica Status Solidi (A)* [Germany], Vol. 5, May 1971, pp. K113-K116.
- [35] P. E. Bauhahn and W. R. Curtice, "Monte Carlo Calculation of Drift Velocity of Electrons in n-GaAs," *Proc. IEEE*, Vol. 60, September 1972, pp. 1106-1107.
- [36] J. G. Ruch and G. S. Kino, "Transport Properties of GaAs," *Physical Review*, Vol. 174, October 1968, pp. 921-931.
- [37] E. M. Bastida et al., "Indirect Electron Drift Velocity Versus Electric Field Measurements in GaAs," *Applied Physics Letters*, Vol. 18, January 1971, pp. 26-31.
- [38] V. E. Riginos, "Nonsaturating Velocity-Field Characteristic of Gallium Arsenide Experimentally Determined from Domain Measurements," *Journal of Applied Physics*, Vol. 45, July 1974, pp. 2918-2922.
- [39] W. K. H. Panofsky and M. Phillips, *Classical Electricity and Magnetism*, Reading, Mass.: Addison-Wesley, 1962, p. 240.
- [40] I. G. Attohoun, "Investigation of Boundary Effects on Gunn Oscillator Operation," Ph.D. Dissertation, Stevens Institute of Technology, January 1975.
- [41] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, New York: McGraw-Hill, 1953, Vol. 1, p. 706.

Appendix A. Evaluation of Equations (15)-(23)

In cylindrical coordinates the Laplacian is given by

$$\nabla^2\Phi = \frac{\partial^2\Phi}{\partial r^2} + \frac{1}{r} \frac{\partial\Phi}{\partial r} + \frac{\partial^2\Phi}{\partial z^2} + \frac{1}{r^2} \frac{\partial^2\Phi}{\partial \phi^2} \quad (A1)$$

If $\Delta z = \Delta r = G$, the grid size, then

$$r = i \cdot \Delta r = iG$$

$$z = j \cdot \Delta z = jG$$

where i and j are integers such that

$$0 \leq iG \leq R$$

$$0 \leq jG \leq L$$

If the differentials are replaced with finite differences, then equation (A1) becomes

$$\nabla^2\Phi|_{ij} = \frac{1}{G^2} \left\{ \Phi_{i+1j} + \Phi_{i-1j} + \Phi_{ij+1} + \Phi_{ij-1} + \frac{1}{2i} [\Phi_{i+1j} - \Phi_{i-1j}] - 4\Phi_{ij} \right\}, \quad i \neq 0 \quad (A2)$$

When $i = 0$ (zero radius) then the second term of equation (A1) is indeterminate. This difficulty can be overcome by using L'Hospital's rule, where

$$\lim_{r \rightarrow 0} \left(\frac{1}{r} \frac{\partial\Phi}{\partial r} \right) = \frac{\partial^2\Phi}{\partial r^2}$$

so that equation (A1) becomes

$$\nabla^2\Phi|_{0j} = \frac{1}{G^2} [\Phi_{0j+1} + \Phi_{0j-1} + 4\Phi_{1j} - 6\Phi_{0j}] \quad (A3)$$

The gradient and second time derivative are similarly given by

$$\vec{\nabla}\Phi = \frac{\partial\Phi}{\partial r} \hat{i}_r + \frac{\partial\Phi}{\partial z} \hat{i}_z$$

or

$$\vec{\nabla}\Phi|_{ij} = \frac{1}{2G} \{(\Phi_{i+1j} - \Phi_{i-1j}) \hat{i}_r + (\Phi_{ij+1} - \Phi_{ij-1}) \hat{i}_z\} \quad (A4)$$

and

$$\frac{\partial^2\Phi}{\partial t^2}|_{ij} = \frac{1}{(\Delta t)^2} [\Phi_{ij}(t) - 2\Phi_{ij}(t - \Delta t) + \Phi_{ij}(t - 2\Delta t)] \quad (A5)$$

where Δt is the time step.

The potential and carrier concentration in region I are obtained by substituting equations (A1) through (A5) into equations (15) and (16). For $0 < iG < R$,

$$\begin{aligned} \Phi_{ij}(t) = & \frac{1}{4 + [\mu\epsilon G^2/(\Delta t)^2]} \left\{ \Phi_{i+1j} + \Phi_{i-1j} + \Phi_{ij+1} + \Phi_{ij-1} \right. \\ & + \frac{1}{2i} [\Phi_{i+1j} - \Phi_{i-1j}] + \frac{\mu\epsilon G^2}{(\Delta t)^2} [2\Phi_{ij}(t - \Delta t) - \Phi_{ij}(t - 2\Delta t)] \\ & \left. - \frac{eG^2}{\epsilon} [n_{ij} - N_j] \right\} \quad (A6) \end{aligned}$$

and

$$n_{ij}(t) = \frac{A_{ij}}{B_{ij}} \quad (A7)$$

where

$$\begin{aligned} A = & \frac{n_{ij}(t - \Delta t)}{\Delta t} + \frac{1}{G} \left\{ \left[\frac{VR_{ij}}{2} + \frac{D}{G} \left(1 - \frac{1}{2i} \right) \right] n_{i-1j} \right. \\ & + \left[\frac{D}{G} \left(1 + \frac{1}{2i} \right) - \frac{VR_{ij}}{2} \right] n_{i+1j} + \left[\frac{D}{G} + \frac{VZ_{ij}}{2} \right] n_{ij-1} \\ & \left. + \left[\frac{D}{G} - \frac{VZ_{ij}}{2} \right] n_{ij+1} \right\} \end{aligned}$$

$$B_{ij} = \frac{1}{\Delta t} + \frac{1}{G} \left\{ \frac{VR_{ij}}{i} + \frac{1}{2} [VR_{i+1j} - VR_{i-1j} + VZ_{ij+1} - VZ_{ij-1}] + \frac{4D}{G} \right\}$$

$$VZ_{ij} = \frac{V(E_{ij}) EZ_{ij}}{E_{ij}}$$

$$VR_{ij} = \frac{V(E_{ij}) ER_{ij}}{E_{ij}}$$

and D is the diffusion coefficient, E_{ij} the electric field, ER_{ij} its radial component, EZ_{ij} its axial component, and $V(E)$ the velocity-field characteristic. At $i = 0$,

$$\begin{aligned} \Phi_{0j}(t) = & \frac{1}{6 + [\mu\epsilon G^2/(\Delta t)^2]} \left\{ \Phi_{0j+1}(t) + \Phi_{0j-1}(t) + 4\Phi_{1j}(t) \right. \\ & \left. + \frac{\mu\epsilon G^2}{(\Delta t)^2} [2\Phi_{0j}(t - \Delta t) - \Phi_{0j}(t - 2\Delta t)] - \frac{eG^2}{\epsilon} [n_{0j}(t) - N_j] \right\} \quad (A8) \end{aligned}$$

and

$$n_{0j}(t) = \frac{A_{0j}}{B_{0j}}$$

where

$$\begin{aligned} A_{0j} = & \frac{n_{0j}(t - \Delta t)}{\Delta t} + \frac{4Dn_{1j}}{G^2} + \frac{1}{G} \left\{ n_{0j+1}(t) \left[\frac{D}{G} - \frac{VZ_{0j}}{2} \right] \right. \\ & \left. + n_{0j-1}(t) \left[\frac{D}{G} - \frac{VZ_{0j}}{2} \right] \right\} \\ B_{0j} = & \frac{1}{\Delta t} + \frac{1}{G} \left[\frac{1}{2} (VZ_{0j+1} - VZ_{0j-1}) + \frac{6D}{G} \right] . \end{aligned}$$

In region II, where an expanding grid is used, substituting equation (A2) into the potential equation gives

$$\begin{aligned} \Phi_{ij}(t) = & \frac{1}{4 + \mu\epsilon(g_{ij}/\Delta r)^2} \left\{ \Phi_{i+1j}(t) + \Phi_{i-1j}(t) + \Phi_{ij+1}(t) \right. \\ & + \Phi_{ij-1}(t) + \left(\frac{g_{ij}}{2R_{ij}} \right) [\Phi_{i+1j}(t) - \Phi_{i-1j}(t)] \\ & \left. + \mu\epsilon \left(\frac{g_{ij}}{\Delta t} \right)^2 [2\Phi_{ij}(t - \Delta t) - \Phi_{ij}(t - 2\Delta t)] \right\} \quad (A9) \end{aligned}$$

where g_{ij} = variable grid size obtained from the expanded grid

R_{ij} = variable radius (distance from the center of the cavity to grid point i, j)

Δt = time step discussed previously

ϵ = dielectric permittivity of either of the two regions (IIa or IIb)

μ = absolute magnetic permeability of region IIa or IIb.

In region III the potential can be evaluated analytically. If it is assumed that the coaxial cavity oscillates at its fundamental frequency,

$$\omega = 2\pi f = \frac{\pi c}{L}$$

where L is the length of the cavity and $c = 1/\sqrt{\mu\epsilon}$, then the potential can be expressed as

$$\phi(r, z, t) = \phi(r, z) \sin \omega t \quad (\text{A10})$$

and equation (17) becomes

$$\nabla^2 \phi + \frac{\pi^2}{L^2} \phi = 0 \quad (\text{A11})$$

which is the Helmholtz equation.

For the coaxial cavity (Figure 3), where R_1 and R_2 are the inner and outer radii, respectively, the potential must vanish at the circular walls and at $r = R_2$. A general solution of equation (A11) that satisfies these conditions is

$$\phi(r, z) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi z}{L}\right) \cdot \frac{\left[\frac{I_0(\lambda_n r)}{I_0(\lambda_n R_2)} \right] - \left[\frac{K_0(\lambda_n r)}{K_0(\lambda_n R_2)} \right]}{\left[\frac{I_0(\lambda_n R_1)}{I_0(\lambda_n R_2)} \right] - \left[\frac{K_0(\lambda_n R_1)}{K_0(\lambda_n R_2)} \right]} \quad (\text{A12})$$

where I_0 and K_0 are the 0th-order modified Bessel functions, A_n are undetermined coefficients to be found from the boundary at $r = R_1$, and

$$\lambda_n^2 = \frac{\pi^2}{L^2} (n^2 - 1)$$

At $r = R_1$ it is necessary that

$$\phi(R_1, z) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi z}{L}\right) = f(z) \quad (\text{A13})$$

where $f(z)$ is the value of the potential at the center conductor of the cavity. Then the Fourier coefficients of equation (A13) are given by

$$A_n = \frac{2}{L} \int_0^L f(z) \sin\left(\frac{n\pi z}{L}\right) dz \quad (\text{A14})$$

The value of $f(z)$ is equal to the bias voltage for $0 \leq z \leq L_1$, where L_1 is the length of the cathode electrode, and $f(L) = 0$. For $L_1 < z < L$ the value of $f(z)$ is given only at the discrete grid points evaluated from the interface between regions IIb and III, that is, $f(z) = V_m$ at $z = L_1 + m\Delta z$, where m is an index associated with these discrete points and Δz is the grid size. The integral of equation (A14) can then be evaluated:

$$\int_0^L f(z) \sin\left(\frac{n\pi z}{L}\right) dz = \int_0^{L_1} f(z) \sin\left(\frac{n\pi z}{L}\right) dz + \int_{L_1}^L f(z) \sin\left(\frac{n\pi z}{L}\right) dz$$

The first term is given by

$$\int_0^{L_1} V_B \sin\left(\frac{n\pi z}{L}\right) dz = \frac{LV_B}{n\pi} \left(1 - \cos \frac{n\pi L_1}{L}\right) \quad (\text{A15})$$

The integral from L_1 to L can only be evaluated numerically. Using the trapezoidal rule gives

$$\int_{L_1}^L F(z) dz = \frac{\Delta z}{2} \left[F(L_1) + f(L) + 2 \sum_{m=1}^M F(L_1 + m\Delta z) \right]$$

or

$$\int_{L_1}^L f(z) \sin\left(\frac{n\pi z}{L}\right) dz = \frac{V_B \Delta z}{2} \sin\left(\frac{n\pi L_1}{L}\right) + \Delta z \sum_{m=1}^M V_m \sin\left[\frac{n\pi(L_1 + m\Delta z)}{L}\right] \quad (\text{A16})$$

where M is the total number of grid points between L_1 and L . Substituting equations (A12) and (A14)–(A16) into equation (A10) yields

$$\begin{aligned} \Phi(r, z, t) = & \sum_{n=1}^{\infty} \frac{L}{2} \left\{ \frac{LV_B}{n\pi} \left[1 - \cos \left(\frac{n\pi L_1}{L} \right) \right] + \frac{V_B \Delta z}{2} \sin \left(\frac{n\pi L_1}{L} \right) \right. \\ & \left. + \Delta z \sum_{m=1}^M V_m \sin \frac{n\pi(L_2 + m\Delta z)}{L} \right\} \\ & \cdot \left[\frac{I_0(\lambda_n r)}{I_0(\lambda_n R_2)} \right] - \left[\frac{K_0(\lambda_n r)}{K_0(\lambda_n R_2)} \right] \\ & \cdot \left[\frac{I_0(\lambda_n R_1)}{I_0(\lambda_n R_2)} \right] - \left[\frac{K_0(\lambda_n R_1)}{K_0(\lambda_n R_2)} \right] \\ & \cdot \sin \left(\frac{n\pi z}{L} \right) \sin \left(\frac{\pi c}{L} t \right) \end{aligned} \quad (\text{A17})$$

where

$$\lambda_n^2 = \frac{\pi^2}{L^2} (n^2 - 1) \quad (\text{A18})$$

To evaluate the boundary conditions [equations (9)–(23)], the surface charge density at the semiconductor surface must be determined. The surface charge density is defined by

$$\sigma = \lim_{\Delta s \rightarrow 0} \frac{\Delta q}{\Delta s}$$

where Δq and Δs are the charge and surface elements, respectively. Now

$$\frac{\Delta q}{\Delta s} = \frac{\Delta q}{\Delta v} \frac{\Delta v}{\Delta s} = \rho \frac{\Delta v}{\Delta s}$$

and

$$\rho_{ij} = -e[n_{ij} - N_j] \quad .$$

Since

$$\frac{\Delta v}{\Delta s} = \frac{r dr d\phi dz}{r d\phi dz} = dr$$

at $r = R$, and

$$\frac{\Delta v}{\Delta s} = \frac{r dr d\phi dz}{r d\phi dr} = dz$$

at $z = 0$ and $z = L$, the surface charge can be expressed numerically as

$$\begin{aligned} \sigma_i &= -eG[n_{Ri} - N_i] & \text{at } r = R \\ \sigma_i &= -eG[n_{i0} - N_0] & \text{at } z = 0 \end{aligned}$$

and

$$\sigma_i = -eG[n_{iL} - N_L] \quad \text{at } z = L \quad . \quad (\text{A19})$$



Vasilis E. Riginos was born in Athens, Greece. He received B.E., M. Eng., and Ph.D. degrees in electrophysics from the Stevens Institute of Technology, Hoboken, New Jersey, in 1964, 1970, and 1973, respectively.

Dr. Riginos was NDEA Fellow at the Stevens Institute of Technology, where he worked on the transport properties of GaAs. He also helped with the development of the Physical Electronics Laboratory and taught physical electronics for several years. In 1974 he joined COMSAT Laboratories, where he is presently a member of the technical staff of the Transponders Department. He is a member of AAAS, IEEE, Sigma Xi, and the American Physical Society.

Ignace G. Atohou received a B.S.E.E. from Colorado State University (1967); an M.S.E.E. from Université Laval, Quebec, Canada (1969); and a Ph.D. in electrical engineering from Stevens Institute of Technology (1975). From January 1968 to September 1969 he was a research and teaching assistant at the Université Laval and worked on semiconductor integrated antennas. He joined Stevens Institute of Technology in September 1970 as an assistant instructor.



A member of the technical staff of the Transponders Department at COMSAT Laboratories since August 1, 1974, Dr. Atohou has been engaged in the modeling of Gunn-effect devices and the communications characterization of Gunn-effect amplifiers operating in the X-band. He is an associate member of Sigma Xi and a member of IEEE.

Index: phase modulation, beacons, design criteria

28-GHz varactor continuous phase modulator

Y. S. LEE

(Manuscript received February 20, 1975)

Abstract

This paper describes the design and measured performance of the 28-GHz varactor continuous phase modulator/driver system which has been developed for an experimental satellite-borne Beacon transmitter. The first upper and lower sidebands are 7 dB down from the output carrier spectral level when the modulator is operating with a single modulating frequency of either 530 or 265 MHz. The modulator system has good long-term output spectral amplitude level stability, typically ± 0.25 dB over an operating ambient temperature range between -15°C and $+45^{\circ}\text{C}$.

Introduction

This paper describes the design and performance characteristics of a reflection-type varactor phase modulator/driver system which has been developed for the 28-GHz Satellite Beacon Propagation Experiment [1], [2].

A linear continuous phase modulator using varactor diodes is attractive for direct modulation of low-power microwave sources when system simplicity and modulator efficiency are important. Varactor diode phase modulator design techniques [3], [4] have been reported for applications below S-band frequencies. One varactor diode can be used for a 180°

peak-to-peak linear phase modulation with a minimum phase-voltage nonlinearity of about 1.4 percent. A 360° phase modulation can be obtained by using two varactor diodes [3]. Recently, direct centimeter- or millimeter-wave modulation techniques [5]–[7] have been developed for application to modern digital PCM-PSK communications systems. However, application of direct millimeter-wave analog phase modulator techniques has not been fully investigated in the literature.

The Beacon transmitter specification required that the 28.56-GHz carrier be continuously modulated by a signal coherent with the carrier to generate the first-order upper and lower sidebands 7 dB down at ± 264.4 MHz for two satellites and ± 528.8 MHz for the third. This requirement was met by using a 28-GHz carrier directly modulated with a linear peak-to-peak phase deviation of 94° , or 0.82 radian, at the modulating signal frequency rate in a varactor phase modulator. This direct modulator simplified the transmitter system design by eliminating some of the multiplier-induced problems which would be introduced if the modulation were obtained at a subharmonic level with less modulator output phase deviation.

The computer-aided design of the overall packaged varactor diode network emphasized minimization of the modulator insertion loss and residual AM component to produce pure phase modulation. Another important design factor was minimization of the modulator system performance sensitivity to environmental factors such as temperature by using saturated level operation of the modulating signal driver output and temperature-compensated bias regulator circuits. Driver redundancy was also provided for the modulator/driver system.

Design of phase modulator and drivers

28-GHz reflection-type varactor circuit

Figure 1 is a schematic model of the reflection-type varactor phase modulator circuit. The network represented by the $ABCD$ transformation matrix includes the diode junction resistance, package circuit parameters, and coupling circuit elements.

The input reflection coefficient, Γ_{in} , of Figure 1 can be written as a bilinear fractional transformation of the variable junction capacitance. The locus of Γ_{in} is mapped onto a circle in the complex reflection coefficient plane [8] because the diode junction capacitance is defined by

$$C_j(V_m) = C_0 \left(1 + \frac{V_m}{\phi} \right)^{-n}$$

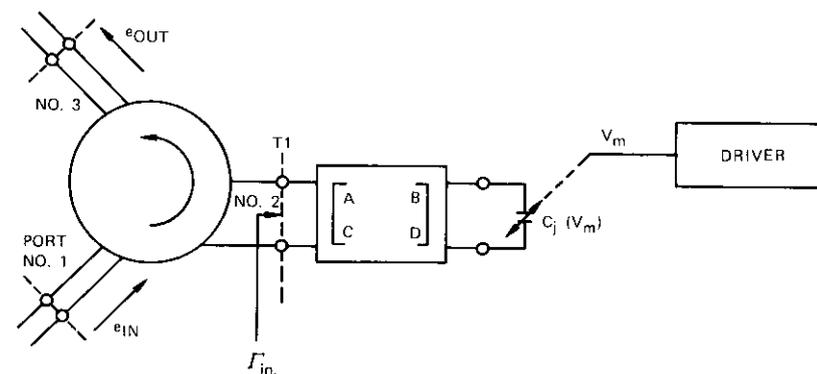


Figure 1. Schematic Diagram of a Reflection-Type Varactor Phase Modulator

where the exponent n depends on the junction doping profile (n equals $1/2$ for an abrupt junction and $1/3$ for a linearly graded junction).

The quasi-static design of the modulator circuit realizes three requirements in the input reflection coefficient plane:

- linear reflection phase angle dependence on modulation voltage across the diode junction,
- minimum power insertion loss by maximizing the radius of the Γ_{in} circle locus,
- minimum residual AM by placing the center of the Γ_{in} circle as close to the origin as possible.

With these three constraints, analytical solution of the circuit parameters is rather difficult when a maximum peak-to-peak phase deviation is specified and the available quality, Q , of the diode is limited. The package parasitic circuit elements of the varactor diode impose restrictions on the realizability of the linear phase modulator, particularly in the millimeter-wave frequency range, and require numerical computer-aided element optimization of the circuit. Figure 2 is a simplified sketch of the 28-GHz phase modulator.

Figure 3 is an equivalent circuit model of the varactor phase modulator shown in Figure 2. The broadband matched impedance transformer permits the reflection phase shift characteristics to be evaluated without restricting the waveguide circuit impedance level. The impedance level can then be selected according to the required phase shift and loss or according to other requirements such as minimum performance sensitivity to en-

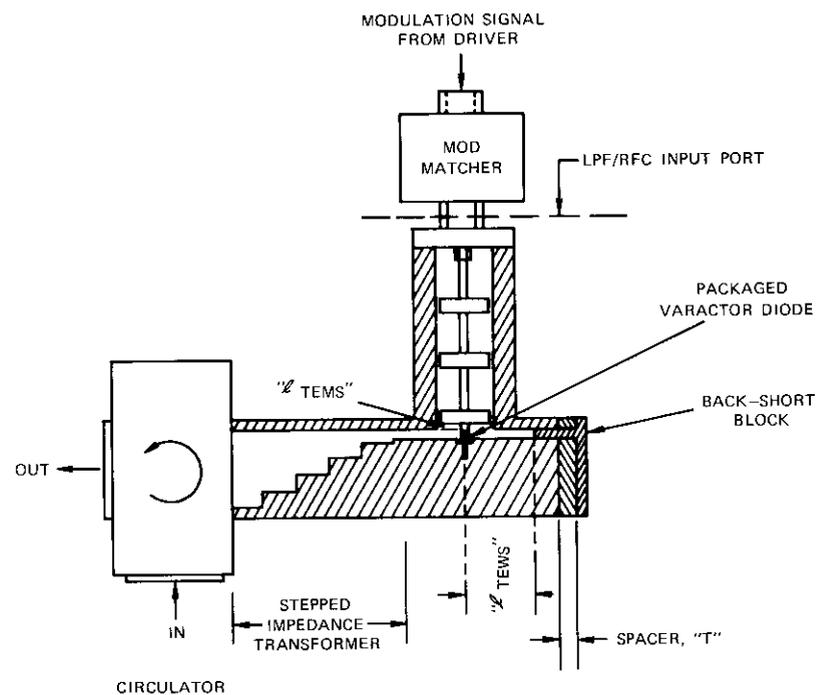


Figure 2. Simplified Cross-sectional Sketch of the 28-GHz Reflector-Type Phase Modulator

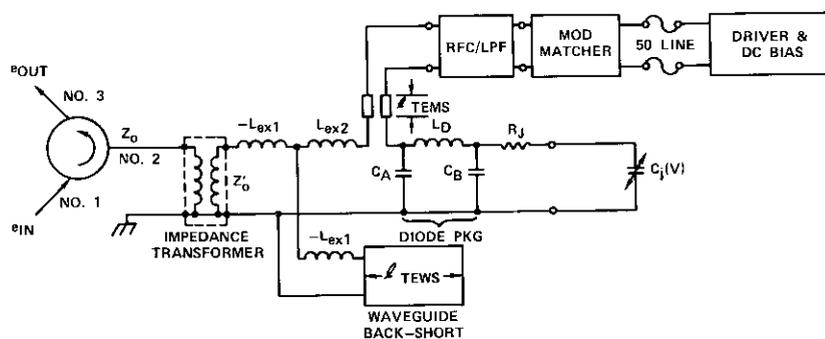


Figure 3. Schematic Circuit Diagram of the 28-GHz Varactor Phase Modulator

environmental temperature. The transformer is a 4-section stepped impedance transformer with residual VSWR of 1.02 over the frequency band of interest (28.0-29.1 GHz).

The package of the varactor diode is modeled by a C-L-C π network, as shown in Figure 3. The reduced height waveguide back-short basically tunes out the package parasitics and the series coaxial line section in front of the LPF. It also provides additional inductance.

Based on the overall circuit model in Figure 3, the magnitude and phase angle of the reflection coefficient, Γ_{in} , as functions of the diode junction voltage have been evaluated by using a microwave circuit analysis program (GCP), with the computed output directly displayed on a Smith chart on the graphic terminal. Typical tuning effects of the reduced height waveguide back-short and the series coaxial line section of Figure 3 are shown in Figures 4a and 4b, respectively. These figures also show the

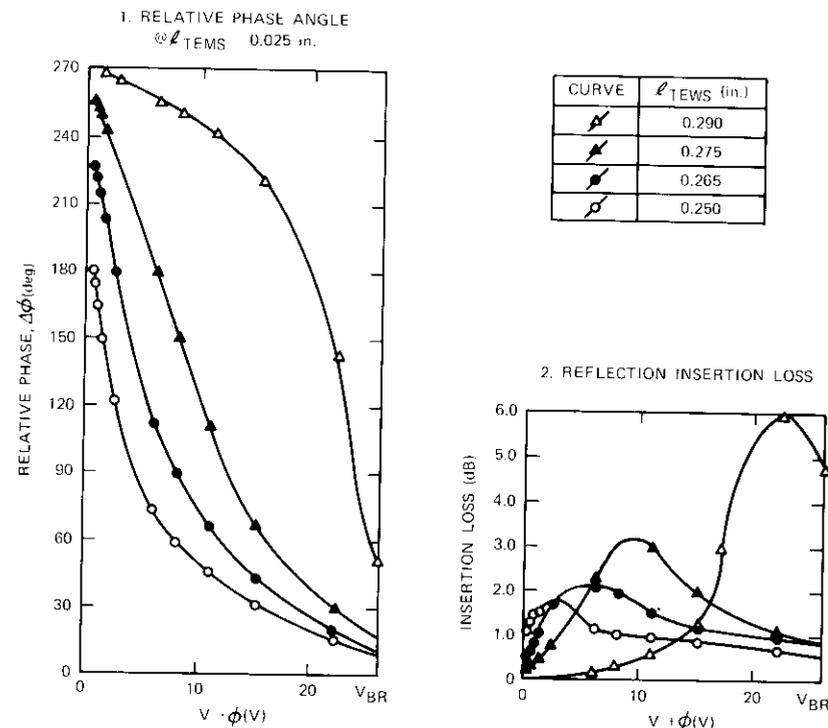


Figure 4a. Circuit Tuning Characteristic of l_{TEWS} in the 28-GHz Phase Modulator

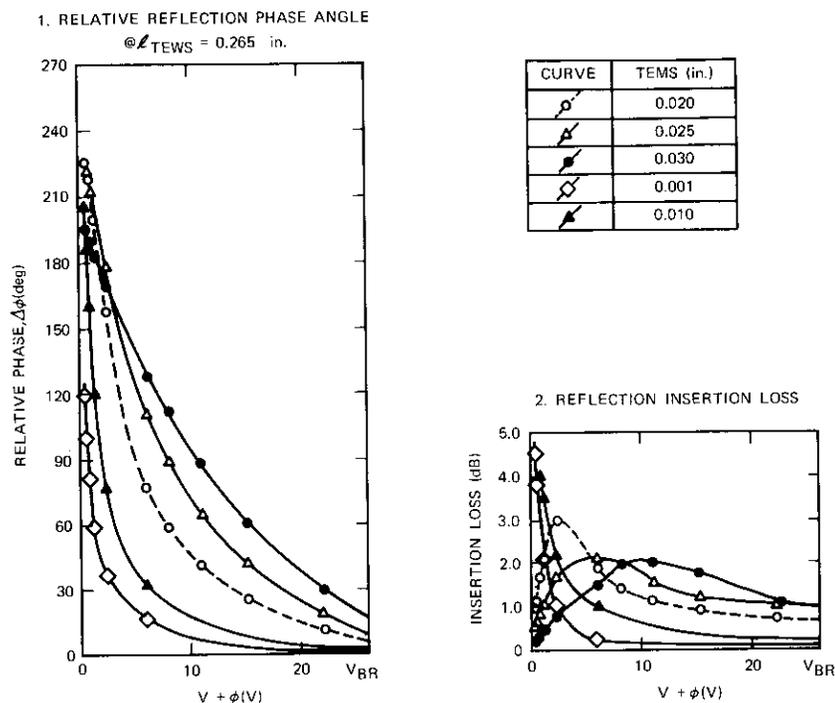


Figure 4b. Tuning Characteristic of l_{TEWS} in the 28-GHz Phase Modulator

dependence of the relative differential reflection phase angle ($\Delta\phi$) and insertion loss on the varactor diode junction voltage, $V + \phi$, where ϕ is the contact potential (≈ 1.2 V) of a high-quality GaAs varactor diode (Alpha type CVE 6810). The varactor diode specifications are listed in Table 1.

TABLE 1. VARACTOR DIODE SPECIFICATIONS

Cutoff Frequency (at $-6V$)	≥ 500 GHz
C_j (at 0 V)	0.4 pF
Capacitance Exponent	$0.42 \leq n \leq 0.5$
Reverse Breakdown Voltage, V_B	≥ -20 VDC
Forward Breakdown Voltage, V_F	≈ 1.2 VDC at 1 mA
Package Type	Min-Dot
Package Capacitance, $C_p = C_A + C_B$	0.11 pF
Lead Inductance, L_d	≈ 0.1 nH

The reflection characteristics of the modulator circuit are sensitive to the dimensional accuracies of the broadband stepped impedance transformer and the diode mount because of the small guide-wavelength ratio, 0.0017 in./deg, at the carrier frequency, 28.6 GHz. The fringing field equivalent lumped elements around the diode mount are significant in the GCP evaluation of the equivalent circuit performance. On the basis of extensive computer-aided evaluation of the quasi-static reflection coefficient of the equivalent circuit model at the carrier frequency as a function of the modulating diode bias voltage, an overall circuit design has been finalized. A phase modulator has been built and experimental results obtained.

Modulation signal circuit

The varactor junction capacitance in the circuit of Figure 3 is driven by a constant voltage pumping at relatively low modulating frequencies. At the modulation frequency the diode junction has a very high capacitive reactance, but the equivalent internal source impedance of the modulating signal circuit looking from the diode junction is very low in comparison. The pump power loss for a fully voltage pumped diode is computed as [9]

$$P_d \approx \omega_m^2 C_0^2 V_M^2 R_s$$

for an abrupt junction having a constant series resistance, R_s , where V_M is the peak pumping voltage amplitude. The power dissipation is about 0.314 mW for a high- Q diode with the following parameters: $V_M \leq V_B/2 = 10$ V; $C_0 = 0.4$ pF; pump frequency, $f_m = 500$ MHz; and $R_s = 2.0$ ohms. Thus the actual power loss in the modulator is negligible.

The input impedance looking into the carrier frequency RF choke [6-section coaxial low-pass filter (LPF) in Figure 2] is almost purely reactive, resulting in a high input VSWR at the modulation signal frequency. The reflected wave at the modulation input port causes instability or degrades the driver output characteristics. The interconnecting 50-ohm line length between the driver output port and the RF choke input port is also critical to the modulator/driver system performance. Therefore, to obtain stable performance from the modulator/driver system, a modulation signal matching circuit (Mod Matcher) has been introduced in front of the RF choke input port, as shown in Figure 2.

The Mod Matcher is required to provide the following simultaneously:

- a. good impedance matching between the RF choke input and the modulator driver output ports (usually 50 ohms nominal);

b. a large voltage transformation ratio, V_2/V_1 , within the modulation frequency band, to supply sufficient modulation voltage across the varactor junction capacitor;

c. additional attenuation to the unwanted harmonics from the driver output.

The modulation signal circuit consists of the Mod Matcher, the coaxial low-pass filter (which provides 150-dB stopband attenuation at the carrier frequency), diode package circuit elements, and varactor junction resistance. It is modeled by a 2-port equivalent network characterized by its transfer parameters, A_M , B_M , C_M , and D_M , as shown in Figure 5.

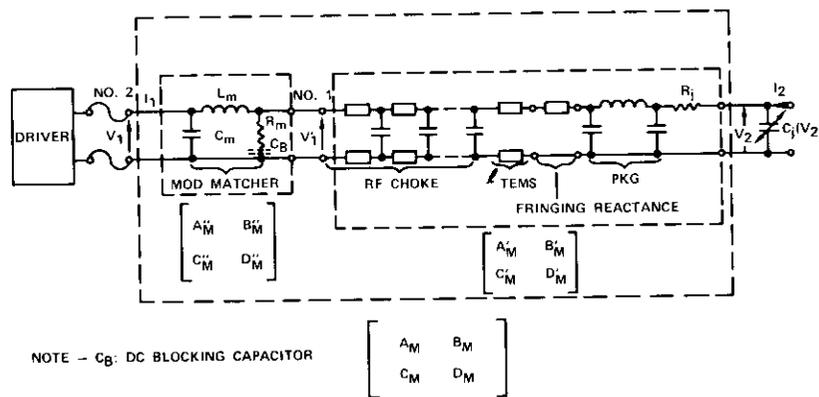


Figure 5. Modulating Signal Circuit Including the Modulation Matching Network

The modulation signal voltage across the junction capacitance is computed as follows:

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A_M & B_M \\ C_M & D_M \end{bmatrix} \begin{bmatrix} 1 & 0 \\ j\omega_m C_j(V_2) & 1 \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} \quad (1)$$

For the terminal condition $I_2 = 0$, the voltage transfer ratio is

$$\frac{V_2}{V_1} = \left\{ A_M + j\omega_m C_0 B_M \left(1 + \frac{V_0 + V_2}{\phi} \right)^{-n} \right\}^{-1} \quad (2)$$

where V_0 is the applied DC bias voltage. When $|\omega_m C_0 B_M| \ll |A_M|$, a

constant voltage pumping at the angular frequency ω_m is maintained at the diode junction:

$$V_1 = A_M V_2 \quad (3)$$

An exact solution of equation (2) for V_2/V_1 is not required for the present circuit because $\beta_0 \equiv |\omega_m C_0 B_M|$ is negligible compared to A_M . For a varactor diode having a typical junction capacitance, $C_0 = 0.3$ pF, it is shown that $\beta_0 = -0.0056 + j0.0009$ and $-0.020 + j0.001$ at $f_m = 265$ and 530 MHz, respectively, for the present circuit, and that equation (3) is valid. The voltage transformation ratio between the RF choke and the diode junction capacitance is $V_2 = 1.140 V_1'$ and $1.83 V_1'$ at 265 and 530 MHz, respectively, where V_1' indicates the modulation signal voltage applied directly to the RF choke input.

The modulation frequency input impedance at port 1 in Figure 5 was modeled as a series R-L-C equivalent circuit in which $R \approx 0$ and $1/\omega_m C > \omega_m L$. The matching network design was therefore simplified with a shunt resistor, R_m , and an L-C-type circuit as shown in Figure 5. The overall modulation signal circuit was subsequently evaluated in detail on the GCP terminal.

The 265-MHz Mod Matcher provided an input residual VSWR less than 1.2 over a 20-MHz bandwidth. The overall modulation signal voltage transformation ratio was computed from equation (3) as follows:

$$|V_2(f_1)| = 1.146 |V_1(f_1)|$$

and

$$|V_2(2f_1)| = 0.358 |V_1(2f_1)|$$

where $f_1 = 265$ MHz. The second harmonic component contained in the driver output was thus attenuated by about 10 dB more than the fundamental 265-MHz component through the overall modulation signal circuit. A similar Mod Matcher circuit designed for the 530-MHz driver had the following characteristics: $|V_2(f_2)| = 1.857 |V_1(f_2)|$, where $f_2 = 530$ MHz, and VSWR < 1.2 over a 17-MHz bandwidth.

Modulator drivers

A stable coherent sinusoidal modulating signal was derived from a crystal oscillator-multiplier chain source which also supplied the 28-GHz carrier power. Drivers were designed to supply the required modulation signal voltages and frequencies.

A temperature-stable driver stage was obtained by using a microwave transistor amplifier or amplifier-doubler with output stage filter circuits operating at saturated output power level. The driver circuits were MIC configuration on Teflon fiberglass ($K_r \approx 2.55$) substrate with lumped elements.

The varactor DC bias voltage was derived from the transistor bias supply through a temperature-stable bias circuit arrangement. The 265-MHz driver was a bipolar $n-p-n$ Si transistor (HP 35826E) amplifier. The output power of the amplifier was typically 16 dBm, with a 0-dBm input at 265 MHz. Its 3-dB gain compression point was at the -4 -dBm input power level.

The amplifier was unconditionally stable with any variation of bias supply voltage (0-19 VDC) and load VSWR. The output power harmonic levels were below 11 dB with 0-dBm input. The driver output power was essentially constant over the input frequency band of 265 ± 30 MHz. The measured output power dependence on ambient temperature was about ± 0.20 dB from -20°C to 50°C .

The 530-MHz modulation signal was obtained with the 265-MHz transistor amplifier followed by a transistor doubler operating in a stable mode [10]. Transistors can be operated as frequency multipliers with power amplification for output frequencies up to about f_T . The loss in power gain of an Xn transistor multiplier operating with an optimum base conduction angle of about $2\pi/n$ compared to that of the class A amplifier of the same device [11] is given by a factor of about $3n - 2$.

The dependence of the transistor bias current stability on the temperature of the doubler stage has been improved by using an FET (2N3822) bias regulator circuit. The 530-MHz driver amplifier-doubler circuit is shown in Figure 6. Typical operating characteristics of the driver stage are listed in Table 2.

The typical output power dependence on ambient temperature from -20°C to $+50^\circ\text{C}$ is shown in Figure 7 for the 530-MHz driver. The output power was measured after the 3-dB hybrid, which was used for redundant connection of two driver stages. The output harmonic content of the driver is plotted in Figure 8a. Figure 8b shows the dependence of output power on input frequency. The amplifier-doubler was stable for any variation of the bias supply from 0 to 19 VDC, input power level up to $+5$ dBm at 265 MHz, and output load VSWR of 4 for all phases.

Driver redundancy provided additional modulator-driver system reliability. A stripline 3-dB hybrid was used for the redundant circuit configuration shown in Figure 9. An identical driver output (modulation

TABLE 2. TYPICAL CHARACTERISTICS OF THE 530-MHz DRIVER

Input	
Frequency	265 MHz
Power	0 dBm
Output	
Frequency	530 MHz
Power	17 dBm
Bandwidth	
1 dB	15 MHz
3 dB	30 MHz
Temperature Effect	
Output Level Variation ($-15^\circ\text{C} \sim +50^\circ\text{C}$)	± 0.3 dB max.
Output Harmonic Levels (referenced to the X2 output frequency spectral level)	
Fundamental Frequency, f_1	< -30 dB
Third Harmonic of f_1	-16 dB
Fourth Harmonic of f_1	-28 dB
Bias Supply	17 VDC \times 38 mA

signal power and varactor DC bias voltage) was obtained with either a V_{cc1} power supply and J_1 input or a V_{cc2} supply and J_2 input which could be switched through the external command signal (not shown in Figure 9) of the Beacon satellite transmitter system.

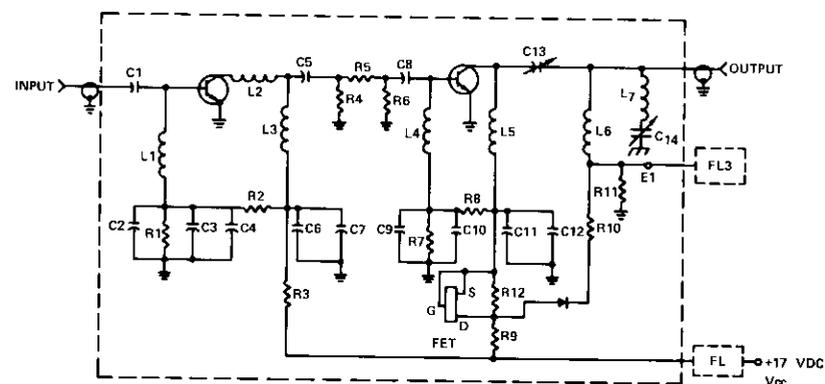


Figure 6. Schematic Diagram of the 530-MHz Driver, Amplifier-Doubler

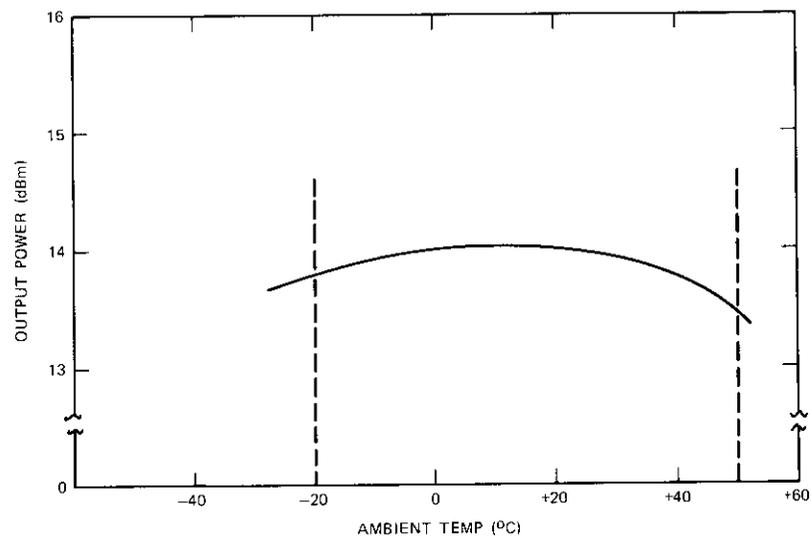


Figure 7. 530-MHz Driver Output Power Dependence on Ambient Temperature

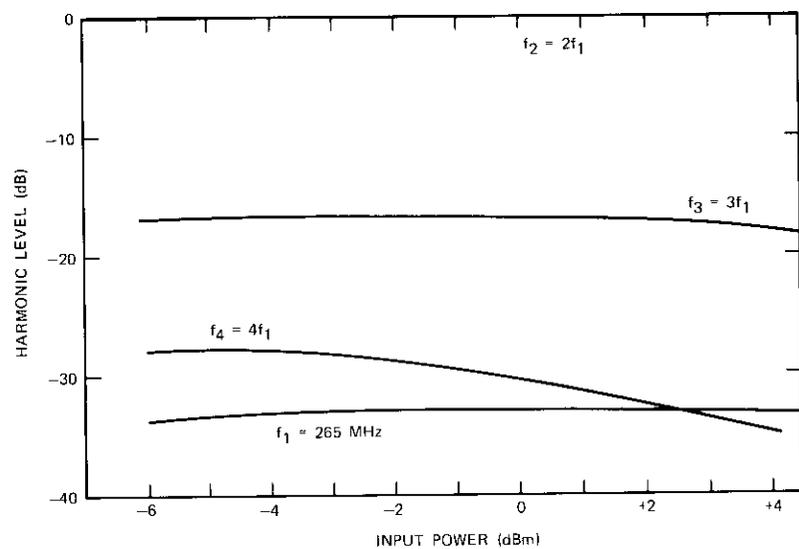


Figure 8a. 530-MHz Driver Output Harmonic Spectral Levels vs 265-MHz Input Power

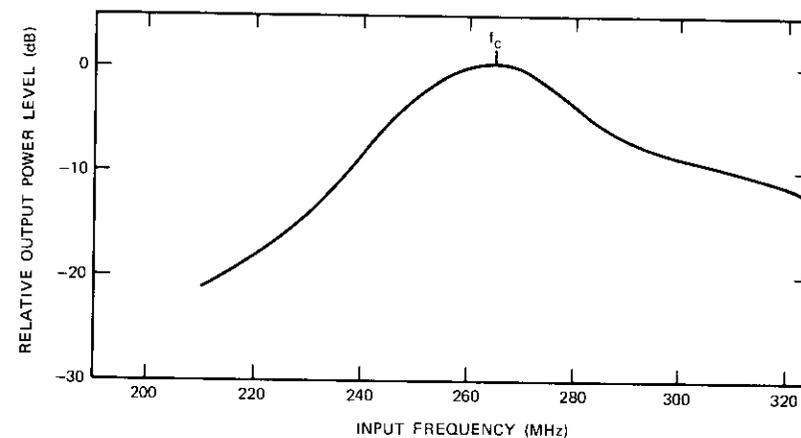


Figure 8b. Output Power vs Input Frequency of the Amplifier-Doubler

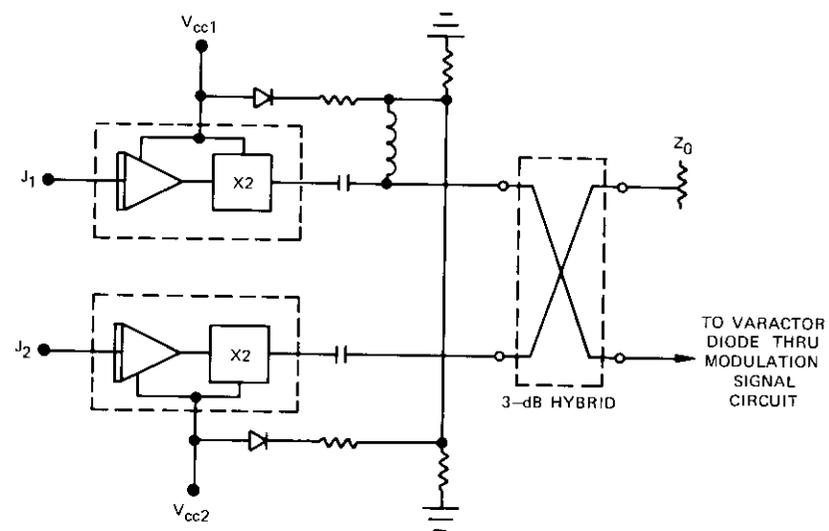


Figure 9. Schematic Circuit Configuration for Driver Redundancy

Measurements and performance

Measurements

The dependence of the first sideband spectral level on the maximum peak-to-peak phase deviation, $\Delta\phi_m$, of the phase modulator is given by

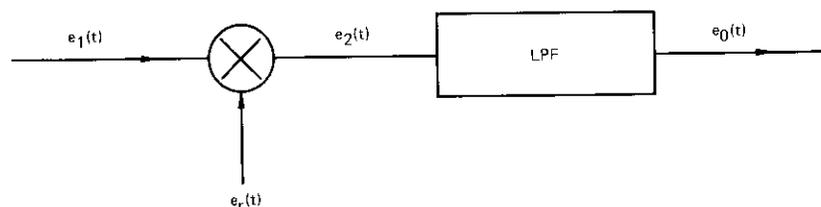
$$\left. \frac{\Delta S_{n=\pm 1}}{\Delta(\Delta\phi_m)} \right|_{\Delta\phi_m = 0.8} = \pm 0.11 \text{ (dB/deg)}$$

The peak-to-peak reflection phase angle must be measured with an accuracy better than 1° , and the modulation frequency is 265 or 530 MHz. Therefore, direct phase measurement of the phase modulator is difficult. The phase measurement limit is about 6° for a measuring system without computer correction.

The reflection phase angle for the maximum phase deviation in the modulator can be indirectly determined in the frequency domain by measuring the modulated output spectral amplitude levels. However, when the actual modulator generates additional residual AM and/or phase-voltage nonlinearity (distortion), the spectral measurements cannot be used for accurate determination of the maximum phase deviation because they do not contain the phase information in each wave component. Therefore, the modulator has been initially tested and experimentally tuned by using time-domain measurements to obtain minimum residual AM and distortion in the modulated output. Spectral measurements have then been employed to characterize the modulator.

TIME-DOMAIN MEASUREMENTS

The basic scheme for detecting the phase modulated signal is shown in Figure 10. The phase modulated noise-free input, $e_1(t)$, and the reference



$$e_r(t) = E_r \cos(\omega_r t + \theta_r)$$

Figure 10. Basic Demodulator Scheme

local oscillator signal, $e_r(t)$ (which is phase coherent with the carrier component for the phase coherent detection), are multiplied to produce the output, $e_2(t)$:

$$e_2(t) = \frac{E_0}{2} \sum_{k=-\infty}^{\infty} J_k(\Delta\phi_m) [\cos\{(\omega_c + \omega_r)t + k\omega_m t + \theta_r\} + \cos\{(\omega_c - \omega_r)t + k\omega_m t - \theta_r\}] \quad (4)$$

where $\omega_c \gg \omega_m$, $E_0 = \eta E_c E_l$, $J_k(\Delta\phi_m)$ is the k th-order Bessel function of the first kind with an argument of $\Delta\phi_m$, and η is the multiplier conversion factor. When $\omega_c = \omega_r$ and the low-pass filter has a cutoff frequency between ω_m and $2\omega_m$, equation (4) gives the demodulated output as follows:

$$e_0(t) = \pm \frac{E_0}{2} J_0(\Delta\phi_m)$$

at $\theta_r = n\pi$, and

$$e_0(t) = \pm E_0 J_1(\Delta\phi_m) \sin \omega_m t$$

at $\theta_r = (n + 1/2)\pi$, where plus signs are for even integers of n and minus signs for odd integers of n .

The detected output has been observed with a sampling oscilloscope in the experimental setup of Figure 11. If the input is purely phase modulated, the amplitude of the detected sinusoidal ω_m component is maximized when the DC value is zero and vice versa as θ_r is varied by $\pi/2$ radians with the phase shifter.

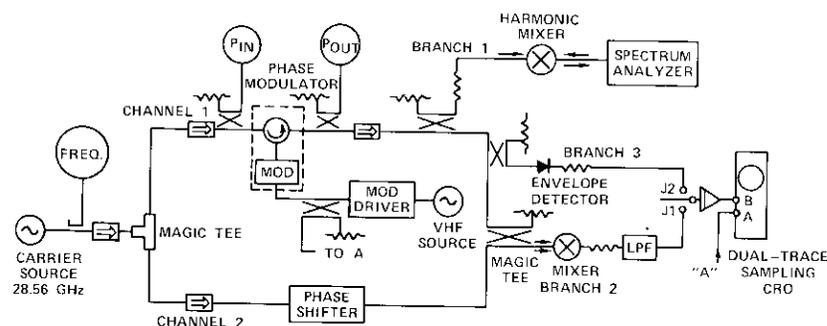


Figure 11. Test Setup for Detection of the 28-GHz Phase Modulator Output

When the PM input, $e_1(t)$, contains residual AM, the low-pass filter output becomes

$$e_0(t) = \frac{E_0}{2} \{ [J_0(\Delta\phi_m) + m\{J_0(\Delta\phi_m) + J_2(\Delta\phi_m)\}] \cos \omega_m t + \cos \theta_r + 2J_1(\Delta\phi_m) \sin \omega_m t \sin \theta_r \} \quad (5)$$

The first term in equation (5) contains the AM-PM combined ω_m component including the second sideband amplitude, $J_2(\Delta\phi_m)$, of the phase modulated output. Accurate measurement of the AM index, m , is difficult because m can be a function of the maximum phase deviation, $\Delta\phi_m$. When $\Delta\phi_m < 1$, $J_2(\Delta\phi_m) < J_0(\Delta\phi_m)$ in equation (5) and an approximate value of m is determined from the detected residual AM and DC levels measured on the sampling oscilloscope.

Figure 12 shows the detected residual AM display (B input) on the dual-trace sampling scope with a 1-ns/division time scale. The lower half of Figure 12 (A input) is the 530-MHz phase modulating driver output waveform. The residual AM index, $m \simeq 0.5$ percent for this measurement.

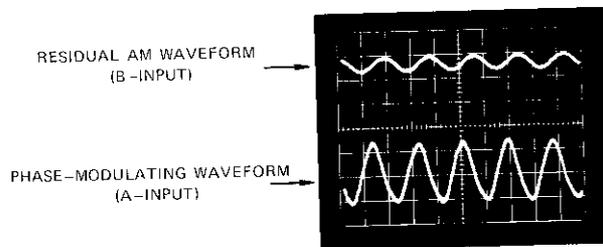


Figure 12. Detected Residual AM and Phase Modulating Waveforms

When $m \simeq 0$, the maximum phase deviation, $\Delta\phi_m$, of the phase modulator is computed from the detected output by shifting θ_r by $\pi/2$ radians for the detected maximum values of DC and demodulated peak-to-peak amplitude of the purely sinusoidal ω_m waveform:

$$\frac{J_1(\Delta\phi_m)}{J_0(\Delta\phi_m)} = \frac{e_0(\omega_m)_{p-p} | \text{ at } \theta_r = \pi/2}{4E_0(\text{DC}) | \text{ at } \theta_r = 0} \quad (6)$$

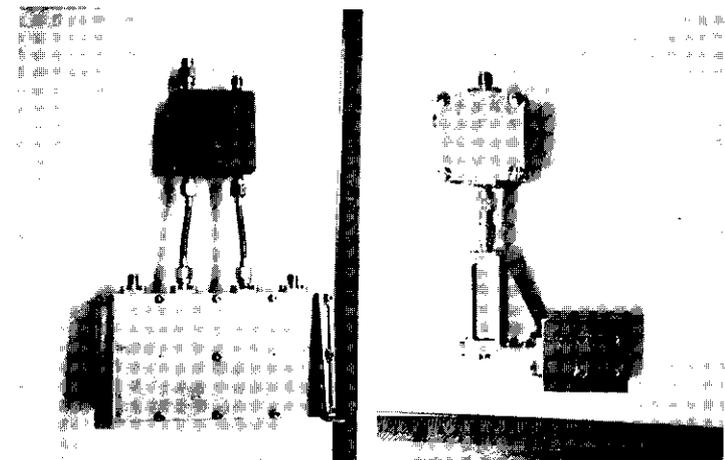
With the detecting elements, including the multiplier, calibrated at the output frequency, ω_m , and at DC, $\Delta\phi_m$ is computed at 94° for the modulated output.

SPECTRAL MEASUREMENTS

The phase modulated output spectral levels were measured after calibration of a high-resolution (IF section) spectrum analyzer and harmonic mixer at each discrete continuous wave frequency input for every sideband frequency component. An equivalent measuring setup was also assembled to obtain increased resolution and accuracy. It consisted of a variable frequency local oscillator with 1,000-Hz square-wave modulation to heterodyne to a fixed IF output (4.0 GHz for this setup) with each spectral component in the phase modulated output. The heterodyned IF output was followed by a narrow-bandpass filter (25-MHz bandwidth) and detected with an SWR meter. The accuracy and resolution of the two measuring systems were comparable, permitting a sideband level accuracy of $\leq \pm 0.2$ dB and a resolution limit of ≤ 0.1 dB.

System performance

Figures 13a and 13b are photographs of the redundant drivers and 3-dB hybrid and the Mod Matcher and reflection-type phase modulator including the circulator, respectively. The GaAs varactor diode was indium solder bonded on a 14-mil reduced height rectangular waveguide with a stepped impedance transformer integrated into it using split-block construction (see Figure 2). Tight mechanical tolerance (± 0.001 in.) was maintained throughout the waveguide sections of the phase modulator.



a. REDUNDANT MOD DRIVER (EM)
(265- OR 530-MHz MODULATION)

b. REFLECTION-TYPE 28-GHz
PHASE MODULATOR

Figure 13. 28-GHz Phase Modulator Assembly

The isolation of the 3-port circulator was >40 dB to avoid error in the modulated output spectral levels (see Appendix A).

The modulator assembly was initially aligned with carrier input power at specified diode bias points without the modulation drive signal. This static test was a convenient preliminary step for the modulator test.

Figure 14 shows the dependence of typical 28.56-GHz modulator experimental performance characteristics, first sideband levels and insertion loss, on the spacer thickness, which determines the length of the reduced waveguide height back-short [$l_{TBWS} = 0.230 + T$ (in.), see Figure 2] at a modulation frequency of either 265 or 530 MHz. The insertion loss includes the 3-port circulator 2-path insertion loss, 0.5 dB.

The overall modulator/driver system generated first upper and lower sideband levels which were equal within ± 0.1 dB, which is the measuring resolution limit of the test setup. The overall phase modulation characteristics showed reasonable agreement with the quasi-static analysis of the phase modulator equivalent circuit as shown in Figures 4a and 4b.

The phase modulation sideband level setting can also be obtained simply by observing the change in carrier output spectral level when the modulating signal is being applied. For the present modulator, the output carrier spectral level drops by 1.5 dB when the modulation signal is turned on.

The modulated first sideband level referenced to the carrier is constant at -7 dB for carrier input power ≤ 13 dBm, and is linear with the modulation signal power. The dependence of the first sideband level on the varactor diode DC bias voltage is -6.6 dB/V at the operating bias point (≈ -1.5 V).

The typical dependence of the system performance characteristics on ambient temperature is shown in Figure 15. Figure 15a shows the dependence of measured first-order sideband levels and insertion loss variations on temperature from -55°C to $+50^\circ\text{C}$ with a 265-MHz driver, and Figure 15b shows the dependence of these parameters on temperature from -20°C to 50°C with a 530-MHz driver. The overall sideband level variation within the acceptance test temperature range (-15°C to $+40^\circ\text{C}$) is about ± 0.25 dB. The varactor diode biasing circuit has been arranged so that the sensitivity of the sideband level to the ambient temperature is minimized. Unconditionally stable performance of the system has been demonstrated through the Beacon specified vibration and thermal vacuum tests.

The typical performance characteristics of the present Beacon phase modulator system are listed in Table 3.

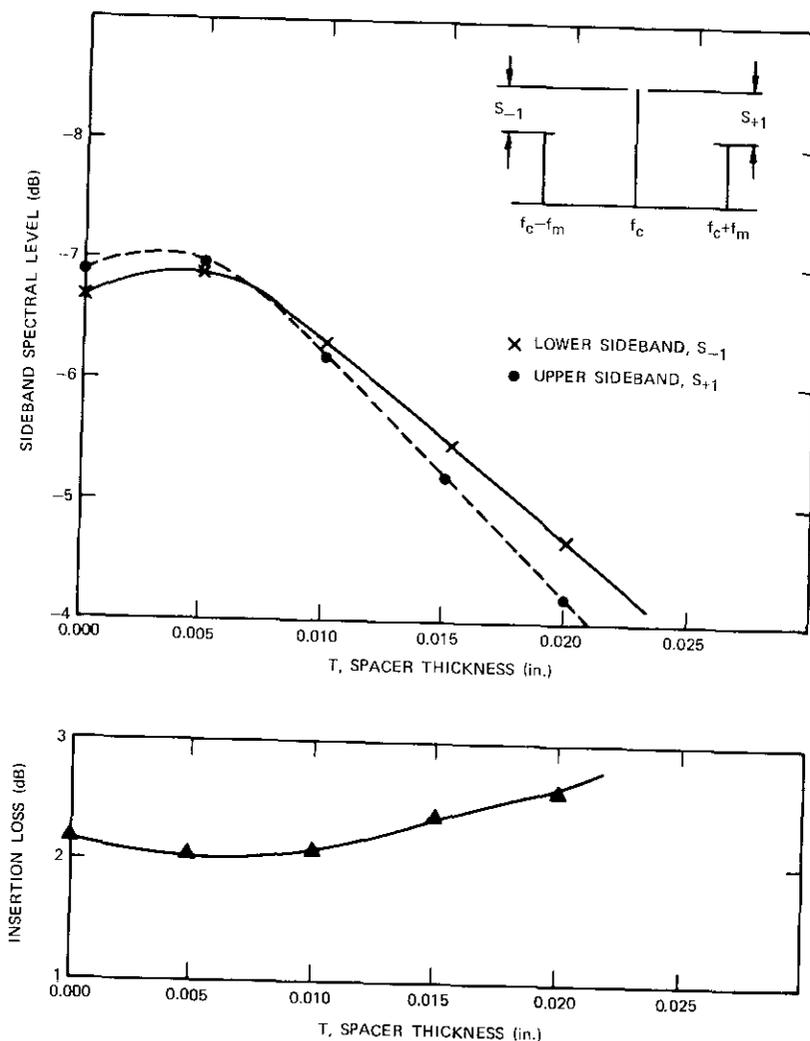


Figure 14. Typical Phase Modulator Characteristic Dependence on the Back-Short Spacer Thickness

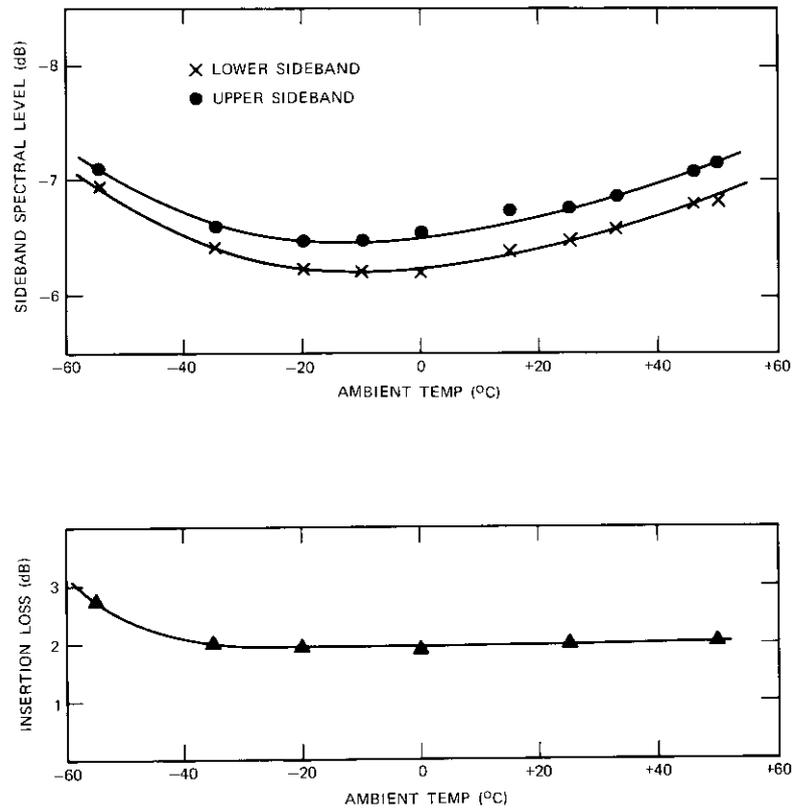


Figure 15a. Typical Performance Dependence on Temperature of the 28-GHz Phase Modulator/265-MHz Driver System

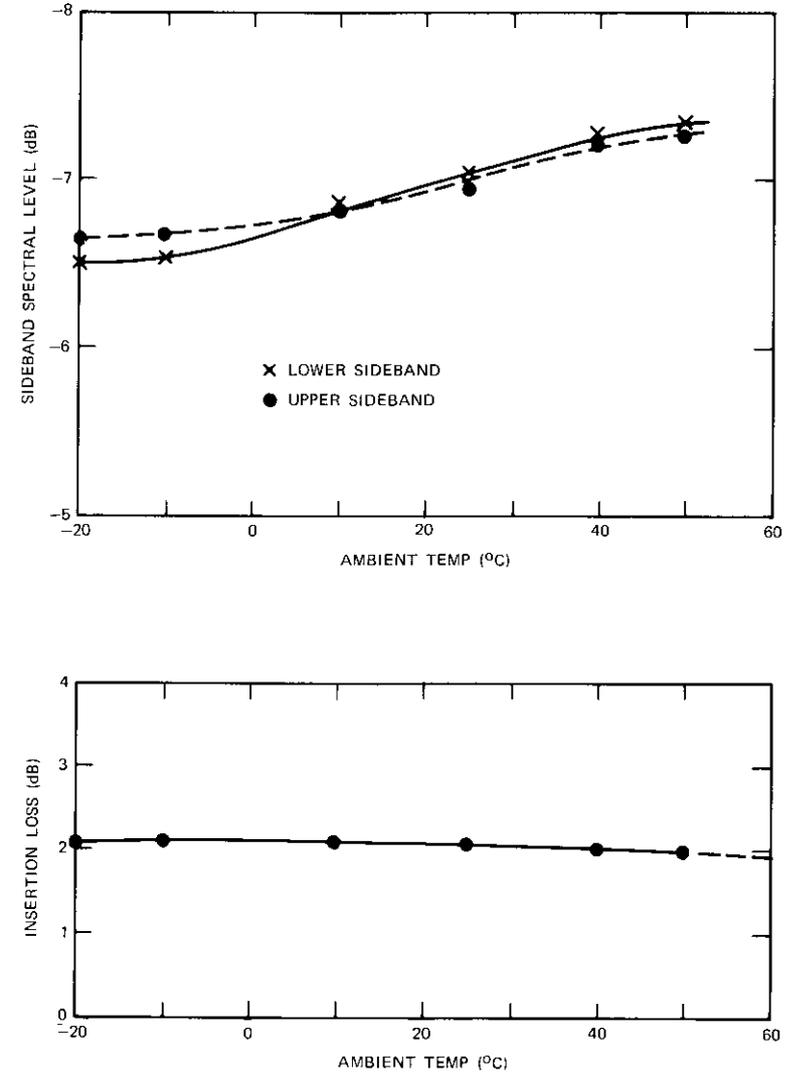


Figure 15b. Typical Performance Dependence on Temperature of the 28-GHz Phase Modulator/530-MHz Driver System

TABLE 3. BEACON PHASE MODULATOR PERFORMANCE

Carrier	
Frequency	28.56 GHz
Input Power	9.5 dBm
Modulation Frequency	
Prototype and F1 and F3	264.4 MHz
Flight Models	
Flight Models F2 and F4	528.8 MHz
Output Characteristics	
Insertion Loss	1.8 dB
Circulator Loss	0.5 dB
Sideband Levels	
First, Upper, and Lower	-7.0 ± 0.1 dB
Second, Upper, and Lower	-20 ± 0.5 dB
Unwanted Harmonic Sidebands	≤ -40 dB
Residual AM Level	0.5%
Environmental Stability	
Temperature Range	-15°C to $+45^{\circ}\text{C}$
First Sideband Level Variation	± 0.25 dB
Insertion Loss Variation	± 0.10 dB
First Sideband	
Modulation Bandwidth (at	≥ 30 MHz
1-dB sideband level variation)	
Modulator Driver	
Configuration	Redundant Inter-connection
Input	0 dBm
Output	≥ 13 dBm
Power Drain	0.680 W (max.)
Weight	
Circulator	24 g
Modulator Subassembly	30 g
Driver Subassembly	290 g
Mod Matcher Subassembly	47 g

Conclusion

The required 7-dB first sideband level in the phase modulator output was obtained with a peak-to-peak linear phase deviation of 94° . Measurements in the time and frequency domains for the characterization of the modulated output agreed with the theoretical analysis. Time-domain measurements made it possible to detect and minimize residual amplitude modulation to about 0.5 percent.

Performance stability of the modulator-driver was obtained with a Mod Matching network and saturated output power level of the driver. Good long-term output spectral stability, typically ± 0.25 dB over the operating temperature range of -15°C to $+45^{\circ}\text{C}$, was obtained, and the system was fully space qualified.

Acknowledgment

The author wishes to thank Mr. W. J. Getsinger for his encouragement throughout the present work and J. Bruno for many careful measurements. Thanks are also due to J. Kaiser for his careful review and helpful comments.

References

- [1] R. D. Briskman, R. F. Latter, and E. E. Muller, "19/28 GHz Radio Propagation Research," *IEEE Transaction on Aerospace and Electronic Systems*, AES-10, No. 4, July 1974, pp. 535-536.
- [2] D. C. Cox, "Design of the Bell Laboratories 18 & 28 GHz Satellite Beacon Propagation Experiment," *ICC-74 Conference Record*, June 17-19, 1974, pp. 27E-1-5.
- [3] R. V. Garver, "360° Varactor Linear Phase Modulator," *IEEE Transactions on Microwave Theory and Techniques*, MTT-17, No. 3, March 1969, pp. 137-147.
- [4] C. S. Kim, "Analytical Study of Direct Modulation Technique," Final Report, Contract AF 19 (629)-4226, AD46453, May 1965.
- [5] W. J. Clemetson et al., "An Experimental MM-Wave Path Length Modulator," *Bell System Technical Journal*, Vol. 50, No. 9, November 1971, pp. 2917-2945.
- [6] H. J. Kuno, D. L. English, and P. S. Pusateri, "Millimeter-Wave Solid-State Exciter-Modulator-Amplifier Module for Gigabit Data-Rate," *1972 International Microwave Symposium Digest*, 72 CHO 612-2 MTT, pp. 195-197.
- [7] Y. Ito et al., "K-band Integrated Microstrip Modulator and Mixer in Waveguide for a High-Speed PCM Radio Repeater," *1972 IEEE International Solid-State Circuits Conference*, February 17, 1972, *Proceedings*, pp. 160-161.
- [8] D. M. Kerns and R. W. Beatty, *Basic Theory of Waveguide Junctions and Introductory Microwave Network Analysis*, New York: Pergamon Press, 1967.
- [9] P. Penfield, Jr., and R. P. Rafuse, *Varactor Applications*, Cambridge, Mass.: M.I.T. Press, 1962.

- [10] H. C. Lee and R. Minton, "Designing Transistor Multipliers," *Microwaves*, Vol. 4, No. 11, November 1965, pp. 18-28.
- [11] R. H. Johnston and A. R. Boothroyd, "High-Frequency Transistor Frequency Multipliers and Power Amplifiers," *IEEE Journal of Solid-State Circuits*, SC-7, No. 1, February 1972, pp. 81-89.

Appendix A. Error with a non-ideal circulator

Error in reflection-type phase modulator design can be caused by the finite isolation of the circulator and/or by the non-ideal impedance transformer in front of the phase modulator circuit, as shown in Figure A-1. Finite isolation of the circulator causes unwanted direct leakage, b_e , from port 1 to port 3. The input residual reflection coefficient of the circulator at port 2 is represented by

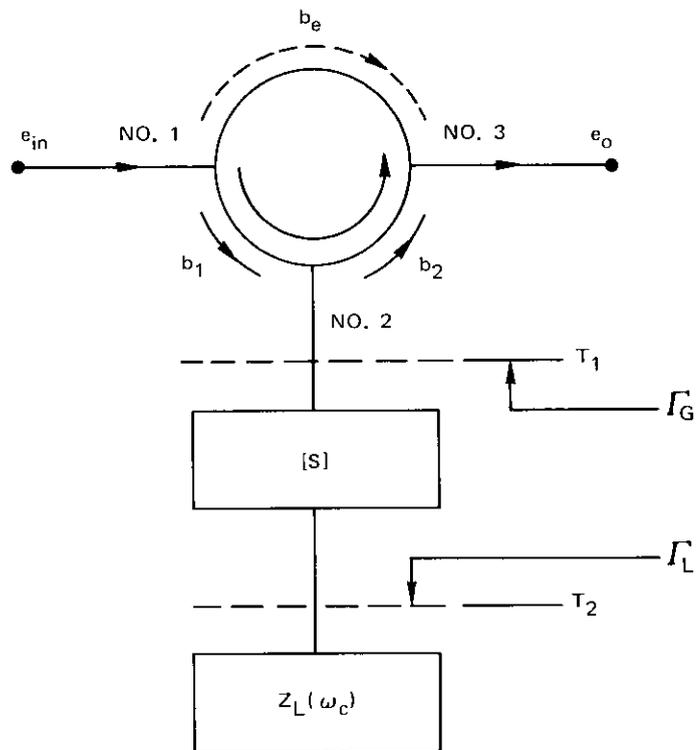


Figure A-1. Non-ideal Circulator and Intermediate Circuit in the Reflection Phase Modulator Circuit

Γ_G . When the circulator and the network are mated at the reference plane T_1 , the reflected signal emerging into port 3 is

$$b_2 = b_1 \left[\frac{(S_{11} - \Gamma_G)(1 - S_{22}\Gamma_L) + S_{12}S_{21}\Gamma_L}{(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L} \right] \quad (A1)$$

and the resulting signal at port 3 is

$$e_o = b_2 + b_e \quad (A2)$$

A well-designed matched impedance transformer satisfies the following conditions:

$$|S_{11}| \ll |\Gamma_L|, \quad |S_{11}| \ll 1 \quad (A3a)$$

$$|S_{22}| \ll 1, \quad |S_{12}| = |S_{21}| \simeq 1.0 \quad (A3b)$$

A further circulator requirement is that $\Gamma_G \simeq 0 + j0$. Equation (A2) is then simplified to yield

$$e_o = b_1\Gamma_L + b_e \quad (A4)$$

and written as follows for the incident carrier and phase modulated wave:

$$e_o(t) = \text{Re} [b_{10} e^{j(\omega_c t + \Delta\phi_m \sin \omega_m t)} + b_{e0} e^{j(\omega_c t + \theta_0')}] \quad (A5)$$

The ideal modulated carrier output spectral density is thus modified so that the output carrier component is added by an error due to the second term in equation (A5). [In general, the overall output spectral density will be subjected to errors arising from the leakage signal according to equation (A1) if Γ_G is not negligible.] The output carrier spectral density is given by

$$S_o = \left| \frac{b_{10}}{2} J_0(\Delta\phi_m) + \frac{b_{e0}}{2} e^{j\theta_0'} \right|^2 \quad (A6)$$

The output sideband levels referenced to the output carrier spectral density are then given by

$$S_{\pm n} = 20 \log_{10} \frac{|J_n(\Delta\phi_m)|}{|b_{10} J_0(\Delta\phi_m) + b_{e0} e^{j\theta_0'}|} \quad (A7)$$

For example, the first sideband spectral levels of a modulator with a circulator having an isolation of 30 dB will be $-7.55 < S_{\pm 1} < -6.90$ (dB) for $\Delta\phi_m = 0.8$ radian because $b_{\epsilon_0} = 31.7 b_{1.0}$ and $J_0(\Delta\phi_m)|_{b_{\epsilon_0}} = 0.846 \pm 0.033$. The ambiguity in the spectral level depends upon the relative phase difference, θ_n , in equation (A5).

For the lossless symmetrical 3-port circulator, the isolation from port 1 to port 3 is equal to the return loss looking into port 2 when the other ports are terminated with the characteristic impedances. In an actual circulator, return loss and isolation greater than 40 dB over the output frequency band are required to avoid error in the phase modulated output.



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 was an officer in the Korean Army Signal Corps (1963-1965), a teaching

assistant at Seoul National University (1965-1967), and a research assistant at M.I.T. (1967-1969). He is a member of IEEE.

Index: unattended earth terminals, traveling wave tubes, power amplifiers

The UET multitube transmitter

A. STANDING

(Manuscript received March 25, 1975)

Abstract

The practice of using one TWT as the final amplifier in a multicarrier FM system has resulted in a steady increase in the TWT power requirement, which is currently at the 12-kW level for commercial communications satellite systems. This paper presents a different approach which uses lower power (400-W) traveling wave tubes (TWTs) multiplexed with filters to form the equivalent of a single high-power amplifier. The design of the power amplifier modules and the TWT specification are discussed with reference to a highly reliable unattended earth terminal.

Introduction

The change to an unattended station requires a new approach to transmitter design problems to provide both the necessary high reliability and a simple means of expansion to a larger station. There are two basic types of transmitters which may be used, a single, multicarrier, wideband amplifier covering the entire band of operation, or many smaller wideband tubes, each restricted to the portion of the satellite band assigned to the earth terminal, in a single-carrier mode. Each of these two basic transmitters may be realized in various ways according to the choice of subsystems.

Table 1 lists the relative advantages and disadvantages of a single tube and a modular transmitter; Figures 1a and 1b are block diagrams of the two concepts. When it is recalled that reliability must be given a significant

weight (see Table 1), a comparison of the different features clearly favors a modular design, which also permits the desired wide size variations through addition of the necessary amplifier modules.

TABLE 1. COMPARISON OF SINGLE TUBE AND MODULAR TRANSMITTERS

Type of Transmitter	Advantages	Disadvantages
Single-Tube Transmitter	<ol style="list-style-type: none"> 1. Simplicity due to single tube 2. Carrier frequency assignment flexibility 3. 1-for-1 spare 	<ol style="list-style-type: none"> 1. High power output per tube 2. Large high-voltage power supplies 3. Water cooling 4. Intermodulation and crosstalk 5. Inefficiency for small stations 6. Difficulties of remote control
Modular Type Transmitter	<ol style="list-style-type: none"> 1. Lower power tube 2. Station size flexibility 3. No intermodulation or crosstalk 4. Air cooling 5. Lower standby investment 6. High reliability due to low-power air cooled tube and low-voltage power supplies 	<ol style="list-style-type: none"> 1. Restrictions on carrier frequency assignment with wide-band transponders 2. Complex multiplexer and switching circuitry

The unattended earth terminal amplifier module described herein has been designed to meet the transmitter requirements of the UET [1].

Tube power and bandwidth

While only the band congruent with the appropriate satellite transponder bandwidth is handled by a given tube, each tube should support the full 500-MHz communications bandwidth. Hence, any tube can be used with any transponder and no frequency constraints are imposed on the system as a result of tube bandwidth.

The single-tube transmitter modules must be diplexed together and the combined output fed to the antenna. In a modular design employing many small tubes with a single carrier per tube, the tube is operated at or near saturation. The system capacity has been calculated for a single carrier per tube, a 40-MHz bandwidth, and the satellite and earth station G/Ts chosen for the UET system. These calculations, described in a previous paper [1], have resulted in a 400-W transmitter power requirement. Hence, the overall tube requirement is for a high-reliability, broadband, 500-MHz design

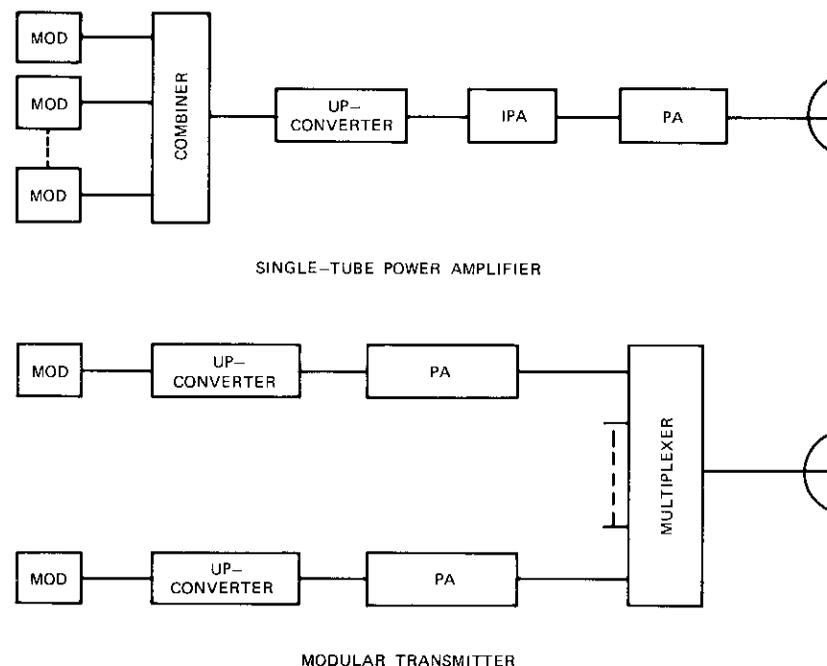


Figure 1. Comparative Transmitter Configurations

with 400-W saturated output power. This requirement can best be met by a permanent-magnet-focused, air cooled, helix-type traveling wave tube (TWT) used, for efficiency, with its collector depressed.

With a single large-capacity carrier per tube as the primary operating mode, intermodulation is no longer a problem and high-level (~10-mW) up-converters are acceptable. Studies have shown that a varactor up-converter with output levels of this magnitude can be realized using an abrupt junction varactor. Such up-converters also offer carrier-to-intermodulation ratios suitable for the transmission of television video and sound carriers through the same output tube, thus eliminating the need for a driver unit. These considerations have led to a TWT with a 55-dB large signal gain and the transmitter module shown in Figure 2.

TWT development

While the TWT in the UET is intended primarily for restricted single-carrier operation, its specification has been written to cover both the single-

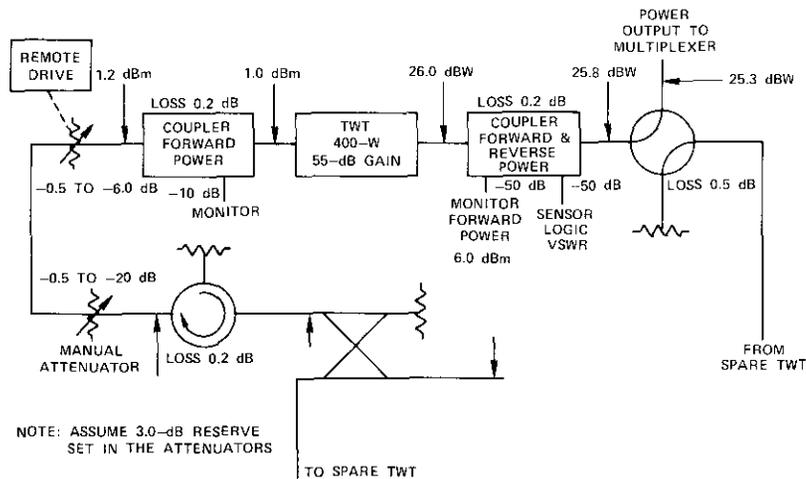


Figure 2. HPA RF Block and Level Diagram

and multiple-carrier cases. The original specifications are given in Table 2. The tube gain which has been achieved is considerably higher than the specified value, allowing a reduction in the up-converter output and hence easing the problems associated with up-converter design.

To ensure stability and allow RF switching under load without damage to the tube, the TWT has been designed to operate indefinitely without RF, with its input/output either short or open circuited. For simplicity, output switching under full RF power has been required, and the TWT can operate without damage at full (400-W) output into a short circuit. After an output short circuit, the helix current will begin to rise, tripping out the helix over current protection and removing the voltage. Consequently, RF output switching under load is possible with the TWT tripping off on the helix if the RF switch fails to operate correctly.

High-power multiplexer

The high-power multiplexer frequency stacks the outputs of a number of power tubes onto a common waveguide while defining the operating band of each tube. In essence, the multiplexer and low-power tube assembly replace the high-power tube presently in use. This replacement must be achieved with maximum efficiency and minimum spectrum loss.

The directional filter multiplexer (Figure 3) is the chosen approach to the multiplexer design. This concept has a number of advantages. All multi-

TABLE 2. TWT SPECIFICATIONS

Parameter	Specification
Output Power	400 W min.
Large Signal Gain	48 dB
Saturation Drive	+8 dBm
Small Signal Gain	55 dB
Gain Slope	0.03 dB/MHz
500-MHz Gain Slope	3.0 dB
Efficiency at Saturation	30%
Phase Sensitivity	0.6 deg/V
Amplitude Sensitivity	0.02 dB/V
Noise Figure	35 dB
2-Tone Intermodulation	
290 W	-11 dB
100 W	-23 dB
Phase Transfer Function, $\Delta\phi$	
0-dB Input	44°
-3-dB Input	39°
-6-dB Input	23°
-10-dB Input	16°
-13-dB Input	8°

plexer input ports are inherently isolated by the hybrids and are substantially independent of the filter characteristics. The division of power between two balanced filters halves the power handling requirements of each filter. The multiplexer is modular in construction, and since the modules are isolated, they can be added to or deleted from the multiplex assembly with no design changes. The major disadvantage of this design is the requirement for twice the number of filters used in the high-power tube and balanced filter pairs. The design and measured performance of the filter used in the UET has been described in a previous paper [2].

Tube protection requirements

The following major faults can cause damage in a TWT:

- oscillation due to low beam voltage applied to the TWT during either switch on or fault conditions;
- arc in the waveguide run connected to the RF output window of the TWT;
- damage, with or without oscillation, caused by excessive reflected power from a high-output VSWR;

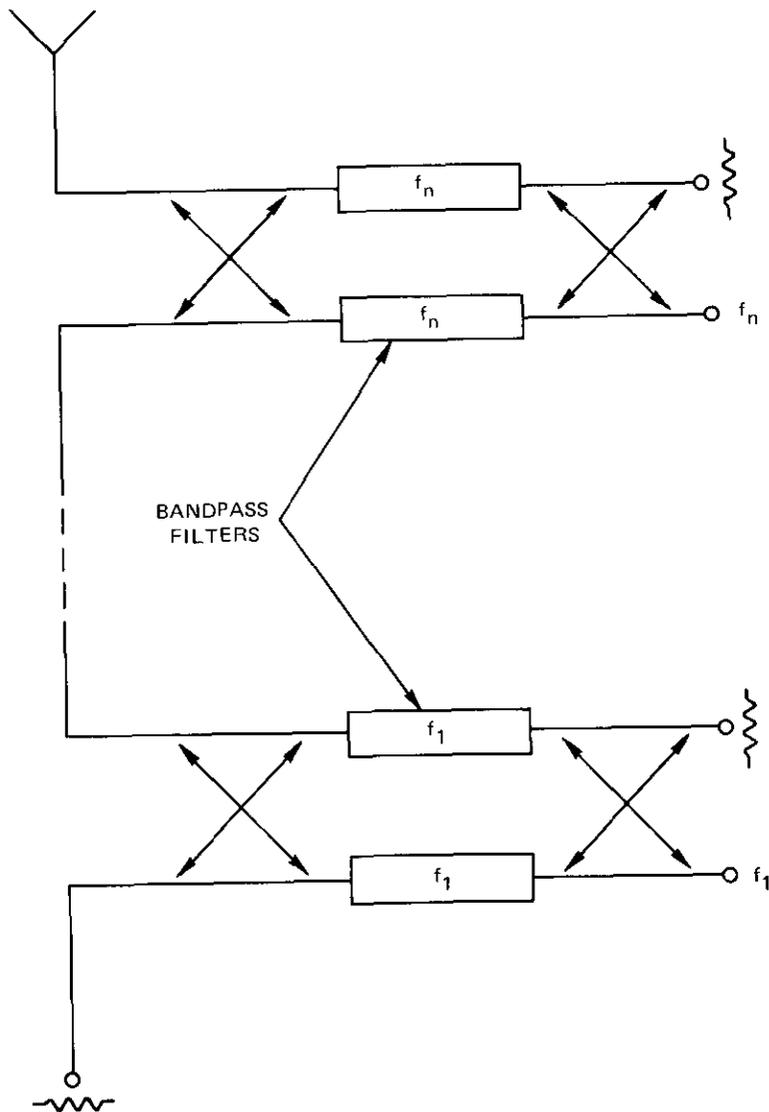


Figure 3. Directional Filter Multiplexer

d. loss of the focusing magnetic field, causing melting of the slow wave structure due to excessive current;

e. damage due to incorrect sequencing of the supply voltages during switch on, switch off, and fault conditions.

The importance of these protection requirements in terms of the operation of a 400-W tube must be assessed. It is necessary to determine the minimum protection that will ensure unimpaired tube reliability and operation.

Oscillations at low beam voltage are present in low-power helix tubes, but no protection is required provided that the dangerous voltage range is traversed in a sufficiently short time (less than 250 ms). Waveguide arcs are a problem associated only with power levels in the kilowatt range, since a parallel arc (between broad walls) cannot be supported at the 400-W level in WR-137 waveguide, irrespective of the VSWR. Thus, an arc detector in the RF output section of the HPA module is unnecessary.

The primary function of the VSWR protection in a conventional high-power amplifier is to detect any waveguide mismatch which can result in a large amount of reflected power and hence damage to the TWT window. In the 400-W TWT the RF window has been designed to handle the reflected power from a waveguide short, and the TWT internal attenuators have been designed to handle the full RF output; thus no VSWR protection is required.

A permanent magnet air cooled TWT obviously does not need protection against loss of magnetic field. However, it must be protected from loss of air cooling and incorrect power supply sequencing.

In the 400-W TWT, excessive helix current must be removed in less than 100 ms to avoid tube damage. Thermal protection for the tube is required to remove the beam power in the event that the TWT exceeds a safe temperature. Hence, the necessary tube protection at the 400-W level has been reduced to the following:

- a. helix current protection within an operational interval less than 100 ms,
- b. an excessive collector temperature switch incorporated in the collector block and connected in the low-speed interlock chain,
- c. low-speed interlocking and sequencing circuits,
- d. a power supply capable of reaching its final state in less than 250 ms.

Redundancy and reliability

While the tube at the earth terminal, unlike that on the satellite, is available for replacement, unattended operation necessitates both high relia-

bility and some form of switching redundancy to meet the requirement for a 3-month availability of 0.999 [1]. Every effort should be made to increase the tube life and to reduce the random failure rate. Of the two, the latter is of greater importance for unattended operation, since unscheduled maintenance will rapidly compromise an unattended system with many stations. A 25,000-hour life, to 1-dB saturation power degradation, will provide an acceptable operational and replacement schedule. It must be remembered that, with tubes operated close to saturation, very little TWT saturated output power degradation can be absorbed by the system. However, sufficient gain margin can be provided to accommodate the tube gain loss with life.

To prevent a service interruption in case of transmitter failure and to allow off-line replacement of the faulty unit, a high-speed switching system must be used. Switches, capable of high-speed switching in the input and output lines, will provide both the necessary switching redundancy and a means of tube replacement.

The use of switched units to provide redundancy imposes a different switch requirement than that which leads to the conventional "operations to failure" approach. In the latter, each switch must remain in a preset position until a tube failure occurs; it must then switch to its alternate position until the faulty tube is replaced during maintenance. The normal operations to failure rating of a switch is derived from tests that rapidly operate the unit until switching failure occurs. The number of operations recorded to this point is the operations to failure rating. In the switched redundancy mode of operation, the number of switch operations is minimal but the switch must always operate on command. The important physical failure mechanisms are different in the two cases, indicating the need for a different switch design.

There are many different ways of connecting switches and HPA modules to achieve switched redundancy. Each connection requires a different number of tubes and switches and offers a different reliability and initial cost. The obvious connection is a one-for-one committed spare configuration. The disadvantage of this configuration is that both a tube and its spare may fail, causing a complete channel outage, while other spares are operating but unavailable.

The use of crossbar switching and uncommitted spares (Figure 4) offers increased system reliability and the ability to bypass a faulty switch. Calculations of system reliability for different configurations using 12 channels (Figure 5) clearly show that a crossbar uncommitted spare approach results in improved reliability and reduction in the number of spare modules.

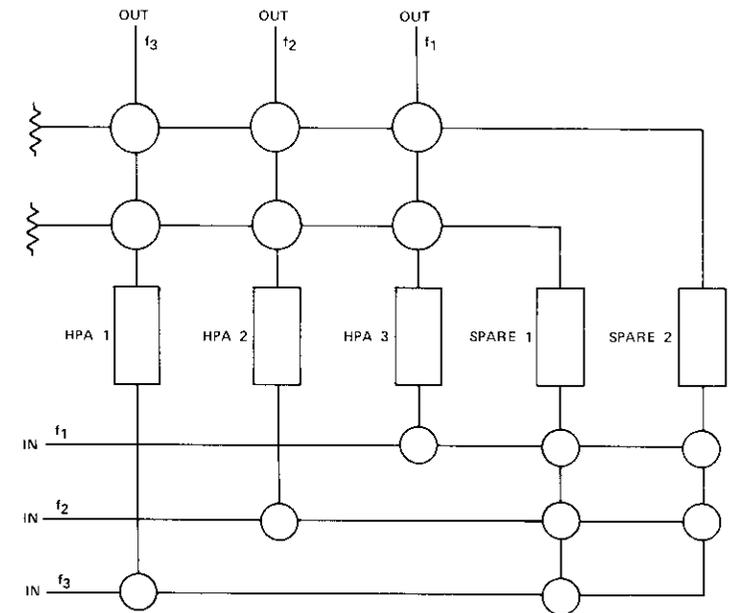


Figure 4. Crossbar Switching Redundancy

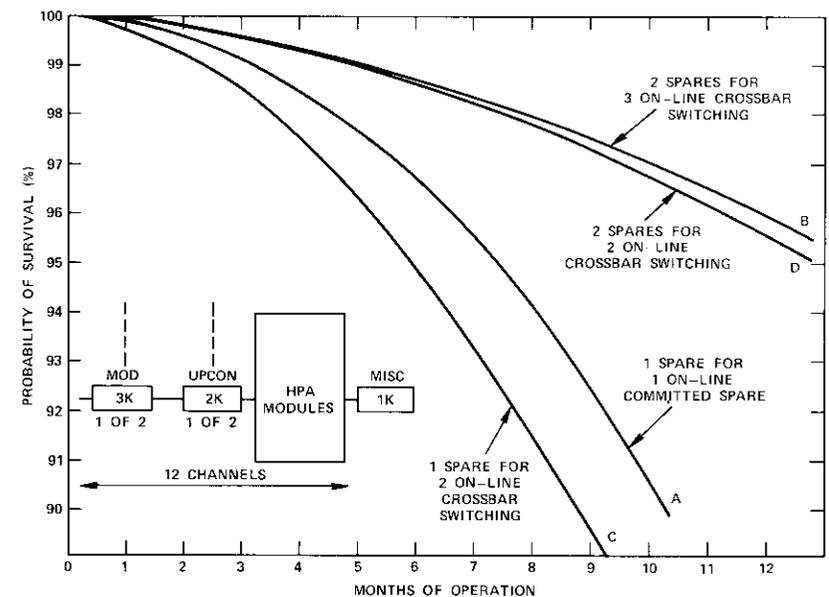


Figure 5. Reliability of Different Redundancy Configurations

It is obvious from Figure 5 that two spares for two or three on-line units will meet the 0.999 availability requirement.

System monitoring

For successful operation of an unattended multistation system, it is necessary to have available, at some central location, essential status information on each station. Such a monitoring system, to be described in a forthcoming paper, is used with the UET, and requires appropriate sensors in the individual units. The following sensors are used in the HPA:

- a. input power go/no go,
- b. output power go/no go,
- c. output power linearity,
- d. collector temperature linearity.

The implementation for the two go/no go sensors consists of an attenuator, adjustable for carrier size, feeding an RF detector and operational amplifier whose output is clamped at a convenient level. The operational amplifier feeds an analog-to-digital converter, thus providing digital data to the central monitoring station.

The linear output power sensor is similar to the go/no go units with the exception of the output clamp; thus, the operational amplifier output can represent the unit's output power. Levels are set to ensure that the RF detector operates in its square law region. The collector temperature is obtained from a thermistor-fed operational amplifier and an analog-to-digital converter.

Performance

Tests of the transmitter module (Table 3) show that all the requirements have been met or exceeded.

TABLE 3. TRANSMITTER MODULE PERFORMANCE

Item	Requirement	Test Results
Saturated Power Output (minimum) Measured at Transmitter Output Flange	400 W	>400 W
Power Output Stability (24 hr)	+0.2 dB	<±0.2 dB
Small Signal Gain (at 30-dB backoff)	55 dB minimum	62.0 dB
Saturated Signal Gain	48 dB minimum	54.6 dB

TABLE 3. TRANSMITTER MODULE PERFORMANCE

(CONTINUED)

Item	Requirement	Test Results
Frequency Band	5.925–6.425 GHz	Satisfies requirement
Gain Variation Across the Band	3 dB	<3 dB
HPA Overload	Ability to withstand a short-circuit output for 200 ms at full power	Satisfies requirement
	Ability to indefinitely withstand an output VSWR of 2:1 at full power	Satisfies requirement
	Ability to withstand an input and/or output short circuit or open circuit with no drive	Satisfies requirement
AM/PM	$\Delta\theta$	$\Delta\theta$
Phase/Input Drive		
Differential Phase		
0-dB Input Relative to Saturation	44°	36°
–3-dB Input Relative to Saturation	39°	33°
–6-dB Input Relative to Saturation	23°	24°
–10-dB Input Relative to Saturation	16°	12.5°
–13-dB Input Relative to Saturation	8°	8°
Amplitude Linearity (two unmodulated carriers spaced 10 MHz apart)		
C/I		
+54-dBm 2-Carrier Output	–11 dB	–16.7 dB
+50-dBm 2-Carrier Output	–23 dB	–25.9 dB
Noise Figure	35 dB maximum	27.5 dB

Conclusions

This paper has described the design of a 400-W TWT transmitter module capable of meeting the UET requirements previously outlined. Satisfactory operation in conjunction with a high-power multiplexer has been demonstrated.

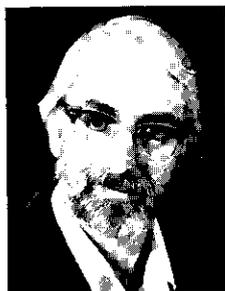
A technique for connecting a number of microwave switches and transmitted modules to yield the desired reliability through redundancy while minimizing the cost and complexity has been discussed.

Acknowledgments

The author would like to acknowledge Dr. W. K. Sones and L. Gray for numerous helpful discussions and R. Mott for much of the detailed design effort.

Reference

- [1] L. Pollack and W. Sones, "An Unattended Earth Terminal for Satellite Communications," *COMSAT Technical Review*, Vol. 4, No. 2, Fall 1974, pp. 205-230.
- [2] R. Gruner and A. Williams, "A Low-loss Multiplexer for Satellite Earth Terminals," *COMSAT Technical Review*, Vol. 5, No. 1, Spring 1975, pp. 157-178.



Arthur F. Standing received an M.S. in electrical engineering from Northeastern University in 1967. Since joining COMSAT Laboratories, he has performed extensive studies of the communications characteristics of microwave tubes and the mathematical modeling problems associated with them. He is presently a member of the Earth Terminals Department of the Microwave Laboratory and a Senior Member of IEEE.

Index: unattended earth terminals, polarization (waves), signal-to-noise ratio, spurious signals

Polarization isolation and spurious signal rejection requirements for a dual-polarized unattended earth terminal

W. SONES AND M. GROSSMAN

(Manuscript received March 14, 1975)

Abstract

In this paper the effects of cross polarization and spurious signal levels are derived in terms of signal-to-interference ratio in the top baseband channel. The problem is analyzed in terms of baseband parameters, frequency deviation, and carrier separation, since these immediately reflect the subsystem requirements and difficulties. Experimental results show satisfactory agreement with predicted values and indicate the cross-polarization objectives for a given channel noise contribution and tolerable spurious signal levels in all local oscillators.

Introduction

The unattended earth terminal discussed in a previous issue of *COMSAT Technical Review* [1] has been designed to operate in a dual-polarized frequency reuse system. Due to imperfect isolation between the orthogonal polarizations and to the absence of an effective corrective cancelling scheme, interference results ultimately in additional baseband noise which must be considered when apportioning the allowable total noise among the various contributing subsystems.

A second source of baseband noise is the in-band unmodulated spurious signals which are frequently encountered in mixers due to local oscillator leakage and impurity. This source of noise is physically unrelated to cross polarization. However since the mathematics describing the two problems are identical, it will be considered here and the allowable spurious levels derived. Both of these noise sources have been studied [2]-[5]; however, the specific adaptation of these results to a satellite system is often unclear, as in the case of computer-generated curves for line-of-sight systems. A specific analysis using satellite system parameters and values is presented here.

The paper first relates baseband power and bandwidth to modulator/demodulator sensitivity constants and carrier frequency deviation. The effects of both unmodulated and modulated interfering carriers are then derived in terms of signal-to-interference (S/I) ratios in the top baseband channel, which, for the system visualized, is the worst case. It is assumed throughout that a frequency modulated signal has a Gaussian spectral density, which is a reasonably good approximation for modulation indices above unity, but becomes progressively poorer as the index is reduced. This is illustrated with measured and calculated spectra, and the Gaussian assumption is shown to be adequate for the cases considered.

Analysis

Relationship between baseband power and modulator/demodulator constants

Assume that the baseband consists of N channels, each having a single sideband voice signal

$$s_i(t), \quad i = 1, N$$

The baseband signal is

$$x(t) = \sum_{i=1}^N s_i(t) \quad (1)$$

corresponding to an approximately white noise power of spectral density $\eta_0/2$ over the frequency band $-f_m \leq f \leq f_m$ (Figure 1). The power in the baseband signal is

$$\sigma_x^2 = N\sigma_s^2 = \eta_0 f_m \quad (2)$$

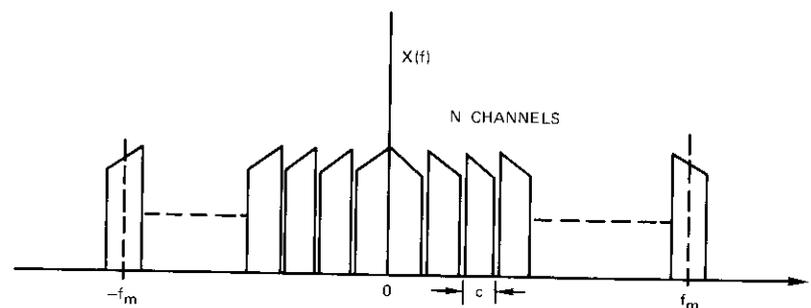


Figure 1. Two-Sided Baseband Spectrum

where σ_s^2 is the power in a signal channel. All channels are considered statistically independent.

In terms of the central limit theorem, a baseband signal composed of more than 10 channels has an approximately Gaussian amplitude distribution,

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{1}{2}\left(\frac{x}{\sigma_x}\right)^2\right] \quad (3)$$

The baseband signal frequency modulates a carrier with a center frequency of f_c to yield the modulated signal

$$y(t) = A \sin [2\pi f_c t + k_1 \xi(t)] \quad (4)$$

where k_1 is the modulator sensitivity constant and

$$\xi(t) = \int^t x(\tau) d\tau \quad (5)$$

Under certain conditions [5], for modulation indices greater than unity, a baseband signal having a white spectrum produces a near-Gaussian power spectral density for the modulated signal, i.e.,

$$Y(f) = \frac{A^2}{4\sqrt{2\pi}\sigma_f} \left\{ \exp\left[-\frac{(f-f_c)^2}{2\sigma_f^2}\right] + \exp\left[-\frac{(f+f_c)^2}{2\sigma_f^2}\right] \right\} \quad (6)$$

(see Figures 2 and 3), where the RF power is given by

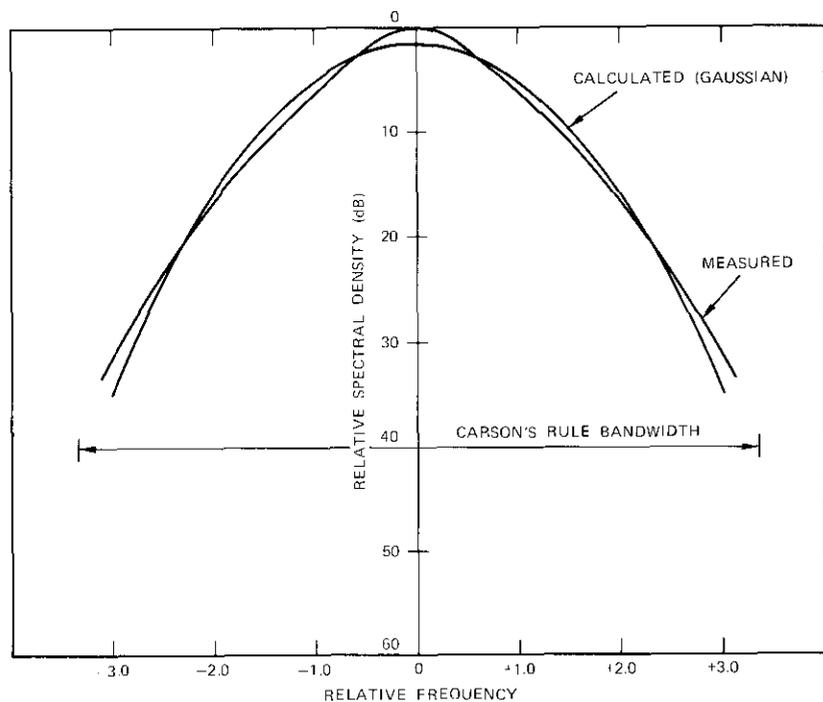


Figure 2. Spectral Density of a 612-Channel Modulated Carrier ($m = 0.75$)

$$\sigma_y^2 = \int_{-\infty}^{\infty} Y(f) df = \frac{A^2}{2} \quad (7)$$

The rms frequency deviation, σ_f , is directly related to the power in the baseband signal by

$$2\pi\sigma_f = k_1\sigma_x \quad (8)$$

With the peak frequency deviation (Δ) assumed to be 10 dB above the rms value, i.e.,

$$\Delta = \sqrt{10} \sigma_f \quad (9)$$

the probability that the baseband signal will be distorted in the frequency modulation process is

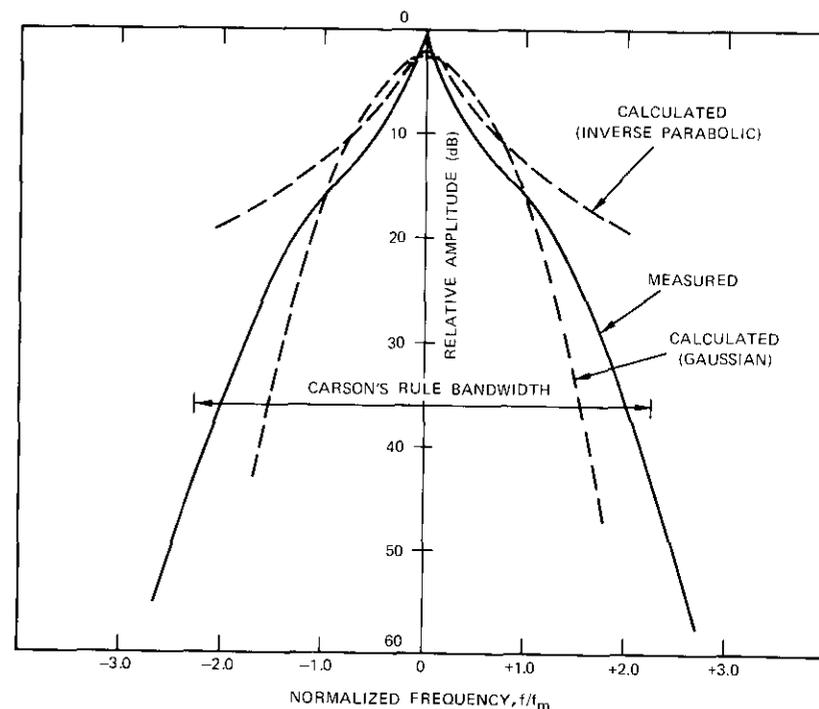


Figure 3. Spectral Density of an 1,872-Channel Modulated Carrier ($m = 0.4$)

$$P = 2 \int_{\sqrt{10}\sigma_f}^{\infty} p(x) dx = 0.0025 \quad (10)$$

Effect of an unmodulated interfering carrier

The addition of an unmodulated interfering carrier produces a signal

$$\begin{aligned} z(t) &= y(t) + B \sin 2\pi f_i t \\ &= R(t) \sin [\gamma(t) + \alpha(t)] \end{aligned} \quad (11)$$

where $R^2(t) = (A + B)^2 \cos^2 \phi(t) + (A - B)^2 \sin^2 \phi(t)$

$$\alpha(t) = \tan^{-1} [Q \tan \phi(t)]$$

$$Q = \frac{A - B}{A + B}$$

$$\begin{aligned}\gamma(t) &= \pi t(f_c + f_i) + \frac{k_1 \xi(t)}{2} \\ \phi(t) &= \pi t(f_c - f_i) + \frac{k_1 \xi(t)}{2} .\end{aligned}$$

The ideal limiter removes the amplitude component $R(t)$, while the detector yields a baseband output

$$\begin{aligned}v_u(t) &= k_2 \left\{ \frac{d}{dt} [\gamma(t) + \alpha(t)] - 2\pi f_c \right\} \\ &= k_2 \left[k_1 x(t) - \frac{B}{A} \frac{d}{dt} \sin 2\phi(t) \right] \quad (12)\end{aligned}$$

where $B \ll A$ and $k_2 = 1/k_1$ (see Appendix A). The distortion component in equation (12) is given by

$$u(t) = k_2 \frac{B}{A} \frac{d}{dt} \sin 2\phi(t) \quad (13)$$

with power spectral density

$$\begin{aligned}U(f) &= \frac{1}{4} k_2 \left(\frac{B}{A} \right)^2 \frac{4\pi^2 f^2}{\sqrt{2\pi} \sigma_f} \left[\exp - \left\{ \frac{[f - (f_c - f_i)]^2}{2\sigma_f^2} \right\} \right. \\ &\quad \left. + \exp - \left\{ \frac{[f + (f_c - f_i)]^2}{2\sigma_f^2} \right\} \right] . \quad (14)\end{aligned}$$

The maximum baseband distortion will occur in the highest frequency ($\pm f_m$) channels when the interfering frequency, f_i , has the value

$$f_i = f_c - f_m . \quad (15)$$

The power spectral density of this distortion is

$$U(f_m) = \frac{1}{4} k_2 \left(\frac{B}{A} \right)^2 \frac{4\pi^2 f_m^2}{\sqrt{2\pi} \sigma_f} \left\{ 1 + \exp \left[- \frac{1}{2} \left(\frac{2f_m}{\sigma_f} \right)^2 \right] \right\} . \quad (16)$$

Substituting for k_2 (see Appendix B) yields

$$U(f_m) = \frac{1}{4} \frac{\eta_0}{\sqrt{2\pi}} \left(\frac{B}{A} \right)^2 \left(\frac{f_m}{\sigma_f} \right)^3 \left\{ 1 + \left[\exp - \frac{1}{2} \left(\frac{2f_m}{\sigma_f} \right)^2 \right] \right\} . \quad (17)$$

The total discriminator output spectral density in the top channel is

$$V(f_m) = \frac{\eta_0}{2} + \frac{1}{4} \frac{\eta_0}{\sqrt{2\pi}} \left(\frac{B}{A} \right)^2 \left(\frac{f_m}{\sigma_f} \right)^3 \left\{ 1 + \left[\exp - \frac{1}{2} \left(\frac{2f_m}{\sigma_f} \right)^2 \right] \right\} . \quad (18)$$

In equation (18), the first term is the recovered signal, and the second term is distortion due to interference (e.g., cross polarization).

For a standard signal-to-noise (S/N) power ratio of 51 dB, the noise power ratio (NPR) is given by

$$\text{NPR} = (51 - K) \text{ dB}$$

where K is the S/N to NPR conversion factor. If the noise allowed for this interference is arbitrarily set at 10 dB below this value, contributing 800 pWp of noise to the channel, then from equation (18)

$$10 \log \left(\frac{A}{B} \right)^2 - 10 \log \left[\left(\frac{f_m}{\sigma_f} \right)^3 \frac{1}{2\sqrt{2\pi}} \right] \left[1 + \exp \left(- \frac{2f_m^2}{\sigma_f^2} \right) \right] = 61 - K .$$

That is, the minimum ratio between the required carrier and an interfering spurious carrier is

$$\begin{aligned}10 \log \left(\frac{A}{B} \right)^2 &= 61 - K + 10 \log \left[\left(\frac{f_m}{\sigma_f} \right)^3 \frac{1}{2\sqrt{2\pi}} \right] \\ &\quad \cdot \left[1 + \exp \left(- \frac{2f_m^2}{\sigma_f^2} \right) \right] . \quad (19)\end{aligned}$$

This expression has been used to calculate the required spurious interference for a range of carrier sizes. Table I compares these calculations and measurements with both conventional and PLL demodulators.

TABLE 1. DESIRED/INTERFERING IN-BAND CARRIER POWER
(WITHOUT MODULATION) FOR VARIOUS INTELSAT IV
CHANNEL CAPACITIES*

No. of Channels	Δf_{rms} (kHz)	f_{max} (kHz)	f_{ch} (kHz)	NPR (dB)	10 log (A/B) ²	
					Measured (dB)	Calculated (dB)
1,872 (Northern Electric demod)	3,181.0	8,120.0	8,002.0	42.0	48.5	47.3
612	1,996.0	2,540.0	2,438.0	42.0	36.0	38.0
(Northern Electric demod)			2,438.0	42.0	39.5	38.6
612	1,996.0	2,540.0	245.0	42.0	25.5	21.0
(GT&E PLL demod)			2,438.0	42.0	40.0	38.6
60 (spot beam, GT&E)	276.0	252.0	245.0	42.0	26.5	21.0
60 (global beam, GT&E)	546.0	252.0	245.0	45.5	39.5	38.3

* All figures are based on a desired S/N of 61 dB.

Effect of a modulated interfering carrier

Assume that the interfering signal is a carrier modulated with a baseband signal consisting of M channels,

$$y(t) = B \sin [2\pi f_i t + k_1 \zeta(t)] \quad (20)$$

where

$$\zeta(t) = \int^t x_1(\tau) d\tau$$

Adding this interfering signal to the desired signal gives

$$\begin{aligned} z(t) &= A \sin [2\pi f_c t + k_1 \xi(t)] + B \sin [2\pi f_i t + k_1 \zeta(t)] \\ &= R(t) \sin [\theta(t) + \beta(t)] \end{aligned} \quad (21)$$

where $R^2(t) = (A + B)^2 \cos^2 \psi(t) + (A - B)^2 \sin^2 \psi(t)$

$$\beta(t) = \tan^{-1} [Q \tan \psi(t)]$$

$$Q = \frac{A - B}{A + B}$$

$$\theta(t) = \pi t(f_c + f_i) + \frac{k_1}{2} [\xi(t) + \zeta(t)]$$

$$\psi(t) = \pi t(f_c - f_i) + \frac{k_1}{2} [\xi(t) - \zeta(t)]$$

Removing the amplitude component with an ideal limiter and applying the resultant signal to an ideal discriminator yields the baseband output

$$\begin{aligned} v_m(t) &= k_2 \left\{ \frac{d}{dt} [\theta(t) + \beta(t)] - 2\pi f_c \right\} \\ &= k_2 \left[k_1 x(t) - \frac{B}{A} \frac{d}{dt} \sin 2\psi(t) \right] \end{aligned} \quad (22)$$

The recovered signal is $x(t)$, while the second term represents distortion; that is,

$$v_d(t) = k_2 \frac{B}{A} \frac{d}{dt} \sin 2\psi(t) \quad (23)$$

Consider the special case of interference from an equally modulated carrier. (The cases of unequal modulation follow routinely, but have not been included here, since they would only complicate the formulation.) The distortion has a Gaussian spectrum centered on $\pm f_i$ with an rms deviation, $\sigma_d = \sqrt{\sigma_f^2 + \sigma_i^2}$. Since equal modulation has been assumed,

$$\sigma_d = \sqrt{2} \sigma_f .$$

Thus, the total spectral density is

$$U_d(f) \simeq \frac{1}{4} \left(k_2 \frac{B}{A} \right)^2 \frac{\pi^{3/2} f^2}{\sigma_f} \left[\exp \left\{ - \frac{[f - (f_c - f_i)]^2}{4\sigma_f^2} \right\} + \exp \left\{ - \frac{[f - (f_c + f_i)]^2}{4\sigma_f^2} \right\} \right] . \quad (24)$$

Substituting for k_2 , as in the unmodulated case, and considering only the highest frequency channel with a white noise signal of η_0 watts/cycle yields the distortion:

$$\left(\frac{B}{A} \right)^2 \frac{c\eta_0}{2} \frac{f_m^3}{4\sqrt{\pi} \sigma_f^3} \left[\exp \left\{ - \frac{[f_m - (f_c - f_i)]^2}{4\sigma_f^2} \right\} + \exp \left\{ - \frac{[f_m + (f_c - f_i)]^2}{4\sigma_f^2} \right\} \right] . \quad (25)$$

If the S/N ratio is again restricted to 51 dB, the S/I ratio for a maximum 800-pWp cross polarization corresponds to 61 dB. When this is related to NPR, where

$$\begin{aligned} \text{NPR} &= 61 - (\text{S/N to NPR conversion factor}) \\ &= 61 - K \end{aligned}$$

the minimum ratio between the required carrier and the interfering carrier is

$$10 \log \left(\frac{A}{B} \right)^2 = 61 - K + 10 \log \left\{ \frac{1}{4\sqrt{\pi}} \left(\frac{f_m}{\sigma_f} \right)^3 \left[\exp \left\{ - \frac{[f_m - (f_c - f_i)]^2}{4\sigma_f^2} \right\} + \exp \left\{ - \frac{[f_m + (f_c - f_i)]^2}{4\sigma_f^2} \right\} \right] \right\} . \quad (26)$$

Experimental results

Unmodulated interfering carrier tests

In one of the tests, an in-band unmodulated carrier was tested as shown in Figure 4. The test was conducted as follows:

- with no interfering carrier, the NPR of the top channel and intermediate channel was measured;
- an unmodulated carrier was then inserted at a frequency f_d so that it fell within the channel being measured;
- the level of the unmodulated carrier was increased to the point at which it degraded the NPR by 3 dB;
- the levels of the desired and interfering carriers were measured.

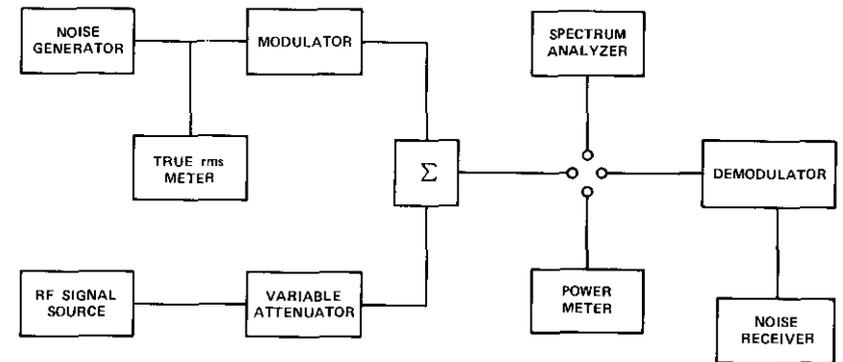


Figure 4. Unmodulated Interfering Carrier Test Setup

This test provided the S/I ratio for a given ratio between the desired and interfering carrier by allowing the distortion introduced by the interfering carrier to degrade the NPR by 3 dB, thus setting the S/I ratio equal to the S/N ratio (or the thermal noise power equal to the distortion).

Modulated interfering carrier tests

The effect of introducing modulation into the interfering carrier was tested as shown in Figure 5. The difference between the wanted and interfering carrier frequencies was varied and for each variation the ratio of desired to unwanted carriers was measured at the point at which the distortion degraded the desired NPR by 3 dB. The results are shown in Figures 6 and 7 for 612 and 1,872 channels. In both cases there was satisfactory agreement between calculated and measured values.

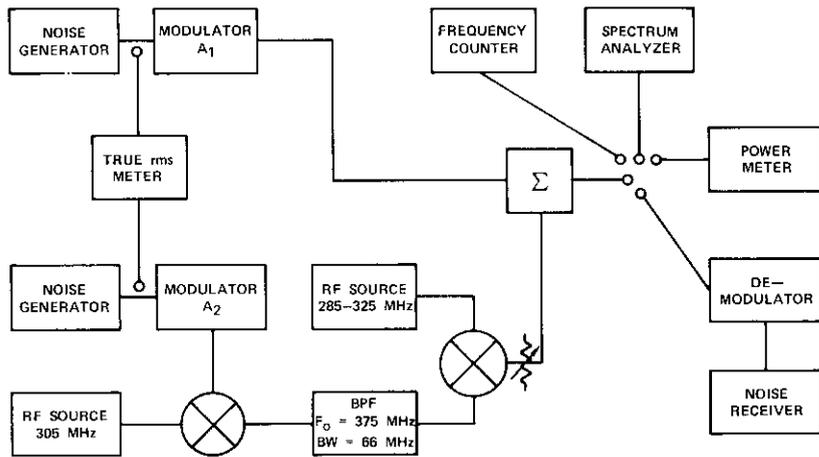
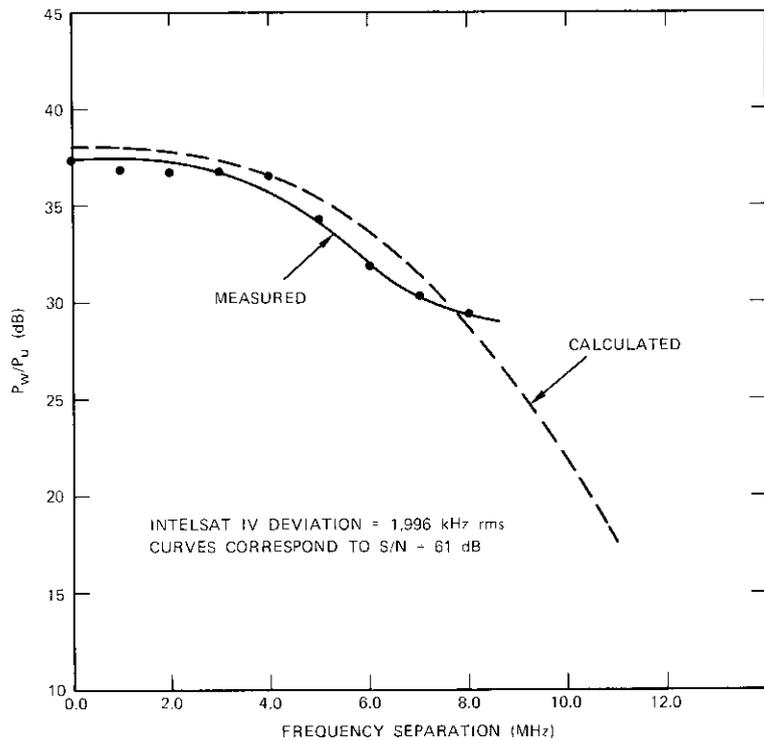
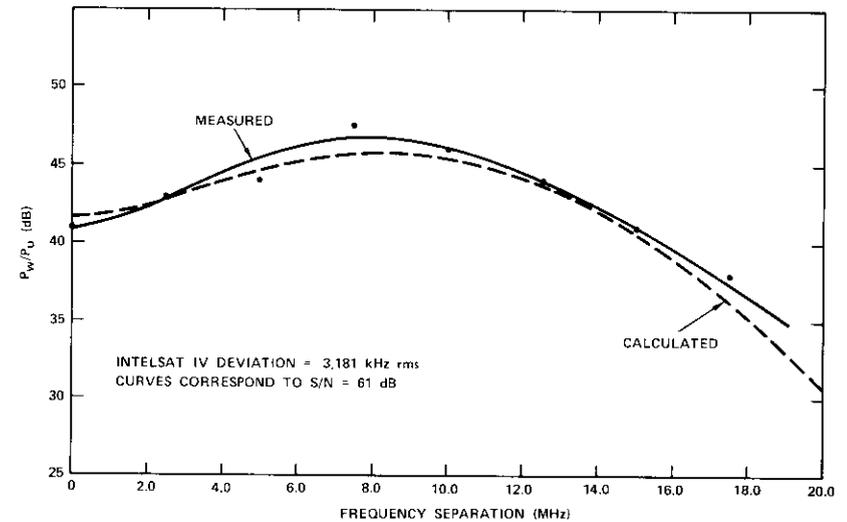


Figure 5. Modulated Interfering Carrier Test Setup

Figure 6. P_w/P_u vs Frequency Separation (612 channels)Figure 7. P_w/P_u vs Frequency Separation (1,872 channels)

For the case in which both wanted and interfering carriers were modulated with 1,872 channels and the frequency spacing was 20 MHz, the results show that the interfering carrier would have to be suppressed by at least 33 dB to prevent the top channel from being degraded with more than 800 pWp of noise. This would be improved by approximately 4 dB if pre-emphasis were included, although 3 dB of the improvement would be lost in the presence of a second interfering carrier in the adjacent channel on the other side of the wanted carrier.

From these results it must be concluded that, for the unattended earth terminal to be operated with a frequency reuse plan similar to that suggested for the INTELSAT IV system, the total system isolation between orthogonal transmissions should not be less than 32 dB (33 dB minus 4-dB pre-emphasis plus 3 dB for the second interfering carrier). If it is further assumed that the coupling in cross-polarized signals is equally divided between the irreducible system orientation errors, including Faraday rotation, and the achievable isolation in the antenna and feed, the unattended earth terminal antenna subassembly objective should be 35 dB.

Curves in Figures 8 and 9 indicate the expected interference in the top channel caused by carriers of similar size separated by half the RF occupied band (half the Carson's Rule bandwidth) and zero spacing.

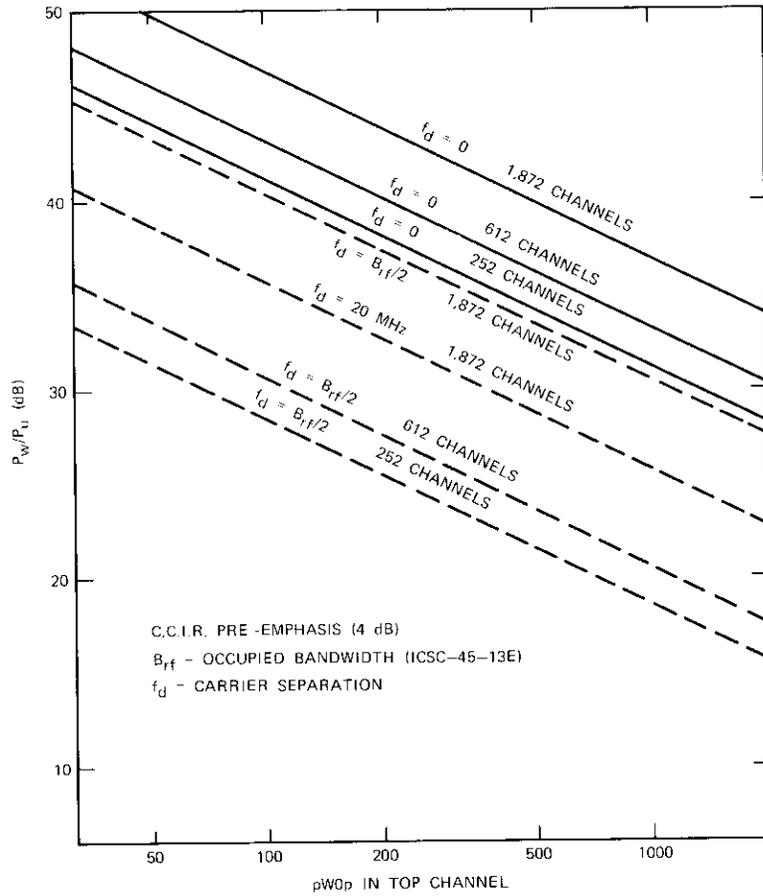


Figure 8. P_w/P_u vs $pW0p$ in Top Channel (INTELSAT spot beam)

Acknowledgments

The authors would like to acknowledge Mr. Marvin Wachs, Dr. Pier L. Bargellini, and Dr. Henry Meyerhoff for numerous enlightening discussions. They would also like to thank Mr. Wachs for permission to use his spectral measurements of an FM carrier and Mr. Alfred Barnes who performed all the experiments.

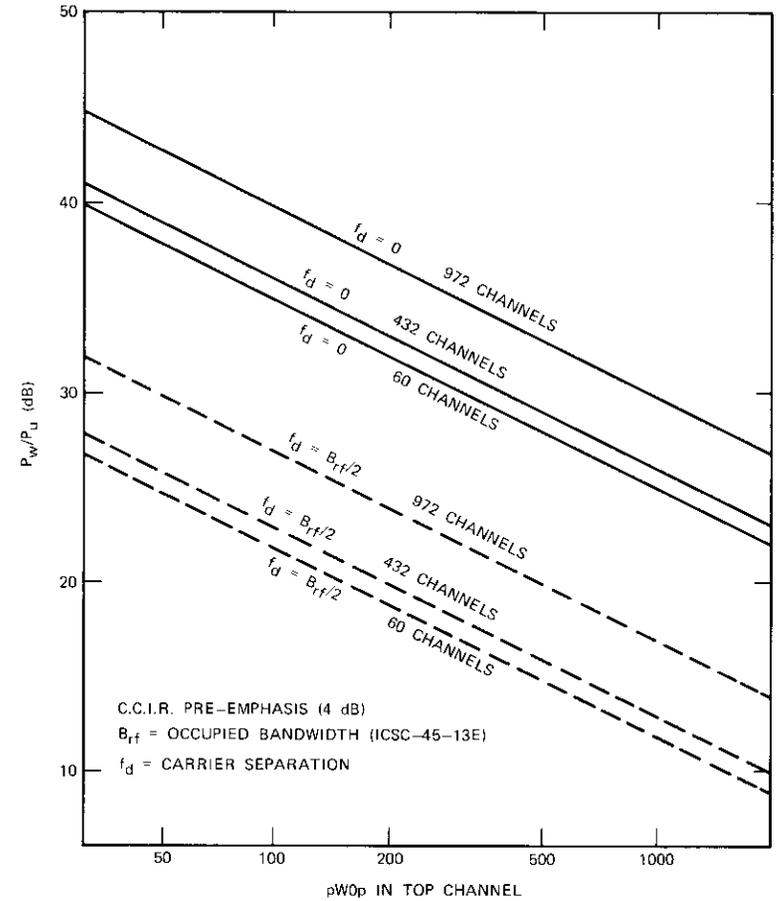


Figure 9. P_w/P_u vs $pW0p$ in Top Channel (INTELSAT IV global beam)

List of symbols

- A Amplitude of carrier
- B Amplitude of interfering carrier
- c Bandwidth of single voice channel plus guardband
- f_c Carrier frequency
- f_i Interfering carrier frequency
- f_m Highest frequency channel in baseband signal

k_1	Modulator sensitivity constant
k_2	Discriminator constant
N	Number of channels in baseband signal
$p(x)$	Amplitude distribution of baseband signal
$s_i(t)$	Single-sideband voice channel
$U(f)$	Power spectral density of $u(t)$
$u(t)$	Distortion component
$V(f)$	Power spectral density of $v(t)$
$v(t)$	Ideal discriminator output
W	Total baseband width
$X(f)$	Baseband spectrum
$x(t)$	Baseband signal
$Y(f)$	Power spectral density of $y(t)$
$y(t)$	Modulated carrier
$z(t)$	Combined modulated carrier and interfering carrier
Δ	Peak frequency deviation of modulated carrier
$\zeta(t)$	Integral of interfering modulating signal
$\eta_0/2$	Equivalent white baseband noise spectral density
$\xi(t)$	Integral of modulating signal
σ_f	rms frequency deviation of modulated carrier
σ_i	rms frequency deviation of interfering modulated carrier
σ_s^2	Power in $s_i(t)$
σ_x^2	Power in baseband signal
σ_y^2	Power of modulating carrier

References

- [1] L. Pollack and W. Sones, "An Unattended Earth Terminal for Satellite Communications," *COMSAT Technical Review*, Vol. 4, No. 2, Fall 1974, pp. 205-231.
- [2] R. G. Medhurst et al., "Distortion in Frequency Division Multiplex F. M. Systems Due to an Interfering Carrier," *Proc. IEE*, Vol. 105, May 1958.
- [3] R. Hamer, "Radio Frequency Interference in Multi-Channel Telephony F. M. Radio Systems," *Proc. IEE*, Vol. 108, Pt. B, January 1961, pp. 75-89.
- [4] R. G. Medhurst, "R.F. Spectra and Interfering Carrier Distortion in F.M. Trunk Radio Systems with Low Modulation Ratios," *IRE Transactions on Communication Systems*, CS-9, June 1961, pp. 107-115.
- [5] J. A. Mullen and D. Middleton, "Limiting Forms of F.M. Noise Spectra," *Proc. IRE*, Vol. 45, No. 6, June 1957, pp. 874-877.

Appendix A. Derivation of equation (12)

In equation (12) it has been shown that the baseband output after ideal limiting and detection is

$$v(t) = k_2 \left\{ \frac{d}{dt} [\gamma(t) + \alpha(t)] - 2\pi f_c \right\} \quad (A1)$$

Substituting for $\gamma(t)$ and $\alpha(t)$ and carrying out the indicated differentiation yields

$$v(t) = k_2 \pi (f_c + f_i) + k_2 \frac{d}{dt} \tan^{-1} [Q \tan \phi(t)] + k_2 2\pi f_c + \frac{k_1 k_2 x(t)}{2} \quad (A2)$$

It is convenient and does not restrict the generality of the analysis to assume that $k_1 k_2 = 1$ and that the modulator and discriminator are exactly centered on f_c .

The second term of equation (A2) may be written as

$$k_2 \frac{d}{dt} \tan^{-1} [Q \tan \phi(t)] = k_2 \frac{Q \sec^2 \phi(t)}{1 + Q^2 \tan^2 \phi(t)} \frac{d}{dt} \phi(t) = \frac{k_2 Q [2\pi f_c - 2\pi f_i + k_1 x(t)]}{\cos^2 \phi(t) + Q^2 \sin^2 \phi(t)} \quad (A3)$$

If it is assumed that the interfering signal is small compared to the wanted signal, namely $A \gg B$, equation (A3) becomes

$$k_2 \frac{d}{dt} \tan^{-1} [Q \tan \phi(t)] \simeq \frac{k_2}{2} [2\pi f_c - 2\pi f_i + k_1 x(t)] \cdot \left[1 - \frac{2B}{A} \cos 2\phi(t) \right]$$

Thus, discriminator output $v(t)$ is obtained from equation (A2), which, after some algebraic manipulation, becomes

$$v(t) = x(t) - k_2 \frac{B}{A} \frac{d}{dt} \sin 2\phi(t) \quad (A4)$$

or equation (12).

Appendix B. Derivation of equation (16)

To determine the distortion power appearing in the top channel, it is necessary to eliminate the discriminator sensitivity constant k_2 as follows. Since σ_s^2 is the single-sideband baseband power per channel, then

$$N\sigma_s^2 = \sigma_x^2 = \eta_0 W .$$

Now the rms deviation

$$2\pi\sigma_f = k_1\sigma_x = k_1\sqrt{\eta_0 W} .$$

With $k_1 k_2 = 1$ and $W \sim f_m$, the top baseband frequency,

$$\frac{1}{k_1} = k_2 = \frac{\sqrt{\eta_0 f_m}}{2\pi\sigma_f} .$$

The distortion term [equation (16)] then becomes

$$U(f_m) = \frac{1}{4} \left(\frac{B}{A} \right)^2 \frac{\eta_0 f_m^3}{\sqrt{2\pi} \sigma_f^3} \left\{ 1 + \exp \left[-\frac{1}{2} \left(\frac{2f_m}{\sigma_f} \right)^2 \right] \right\} . \quad (B1)$$

William K. Sones received a B.Sc. with honors from London University, and an M.S. and a Ph.D. with distinction from Johns Hopkins University. His professional career includes seven years at Westinghouse, Baltimore, where he participated in the design of HF, VHF, and microwave communications systems. At Johns Hopkins University five years were spent jointly as supervisor in charge of the communications and data processing group at their radiation laboratory and as professor in the school of graduate engineering, where his main interests were optimum detection and adaptive maximum likelihood techniques.



In the six years Dr. Sones has been at the COMSAT Laboratories he has worked principally on earth terminal design for satellite systems. He is presently a Senior Staff Scientist.



Melvyn Grossman received a B.E.E. from the City College of New York in 1961 and an M.S. from George Washington University in 1974. Prior to joining COMSAT, he was a design engineer with Space Communications Laboratory, ITT Defense Communications Division, Nutley, New Jersey, where he engaged in PLL and FMFB receiver design. He joined the RF Transmission Laboratory of COMSAT Laboratories in 1969 and is presently engaged in the design of low-noise amplifiers, FDM/FM modulators, and threshold extension receivers for earth terminal use. He is a member of the IEEE.

CTR Notes

Design of a lossy waveguide filter

M. H. CHEN AND C. MAHLE

(Manuscript received April 18, 1975)

Introduction

In a satellite transponder, system constraints require many channelization filters characterized by a flat passband response, a high rolloff rate, and high out-of-band rejection. The flat passband response is necessary to maintain a low level of intelligible crosstalk and the high out-of-band signal rejection is needed to control the so-called multipath [1] effect. At 4 GHz, channelization filters with 1-percent fractional bandwidth (40 MHz) and unloaded Q 's of the order of 10,000 for the fundamental modes of waveguide cavities are typical.

Future applications at 12 GHz may require filters with smaller fractional bandwidth, while the realizable unloaded Q 's of waveguide cavities in the fundamental mode may decrease to about 3,000 to 5,000. This will lead to excess gain slope in the filter's passband. One solution to this dilemma which is applicable to input multiplexer filters* is a tradeoff in

*An input multiplexer [2] is generally used to channelize a band into channels followed by individual power amplifiers (TWTAs). Additional insertion loss can be tolerated at this point in the transponder since the signal level is usually low.

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

Dr. Ming H. Chen is a member of the technical staff in the Transponders Department of the Microwave Lab, COMSAT Laboratories.

Dr. Christoph E. Mahle is Manager, Transponders Department, Microwave Lab, COMSAT Laboratories.

which mid-band insertion loss is traded for flat in-band gain slopes to achieve reasonable values of usable bandwidth with low Q 's. Of course, the out-of-band selectivity must be preserved during this tradeoff.

This paper presents a design procedure which uses an equal predistortion technique [3]-[5] for the design of a bandpass filter with flat in-band amplitude response in a low- Q situation. The predistortion technique was introduced by Darlington [3] and has been discussed in several textbooks [4], [5]. However, this technique has not been fully utilized in lossy filter design. In a recent publication [6] Craig developed design tables for lossy filters that require the solution of n variables from n simultaneous equations for an $(n + 1)$ -pole Chebychev filter. Because a great deal of computation is required if n is large, its application is limited to filters of low order. The currently available design data are applicable only to Chebychev filters with no more than five poles [6]. Furthermore, to the authors' knowledge, no microwave filters have been developed using this technique.

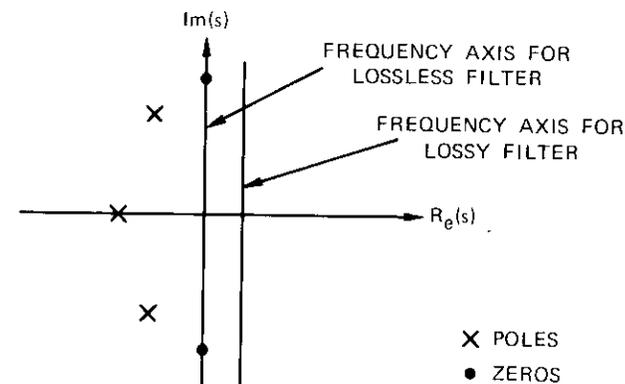
This paper presents a systematic design method for lossy filters utilizing the predistortion technique. The design method, which is based upon the realization of a lossless bandpass filter with a prescribed passband reflection, differs from the commonly available realization method that assumes no reflections at mid-band. However, a synthesis procedure for a general coupled resonator transmission network developed by Saito [7] has been found to be applicable to this design.

Theory and design procedure

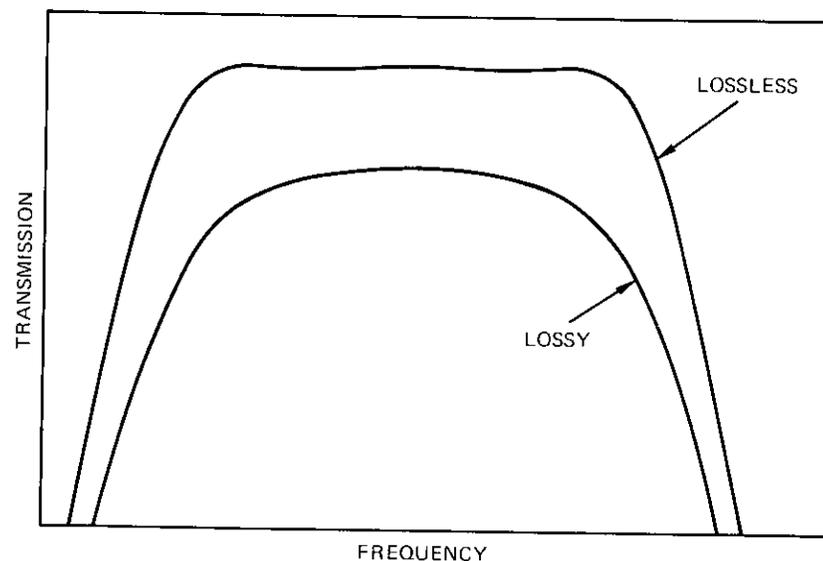
Conventional bandpass filter design is based upon a lossless synthesis which provides a flat in-band response for an infinite Q realization. The introduction of losses causes the band edges to round off (see Figure 1) because the frequency axis is shifted from the imaginary axis to the right half of the complex S plane, thus destroying the correct pole and zero relationship. However, a flat transmission response under lossy conditions may be obtained by restoring a correct pole-zero relationship through the use of predistortion techniques (see Figure 2).

The predistortion design procedure consists of the following steps:

- a. the poles and zeros of the filter's transfer function are found according to the lossless synthesis method;
- b. all of the poles are moved toward the positive direction along the real axis by $1/(Q_u \times BW)$, where Q_u is the estimated unloaded cavity Q and BW is the filter's percentage bandwidth;

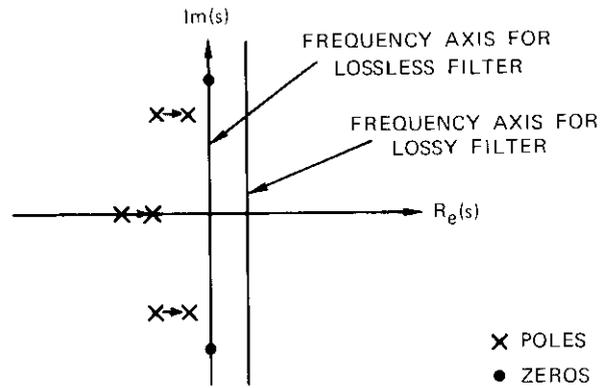


(a) POLE - ZERO DIAGRAM

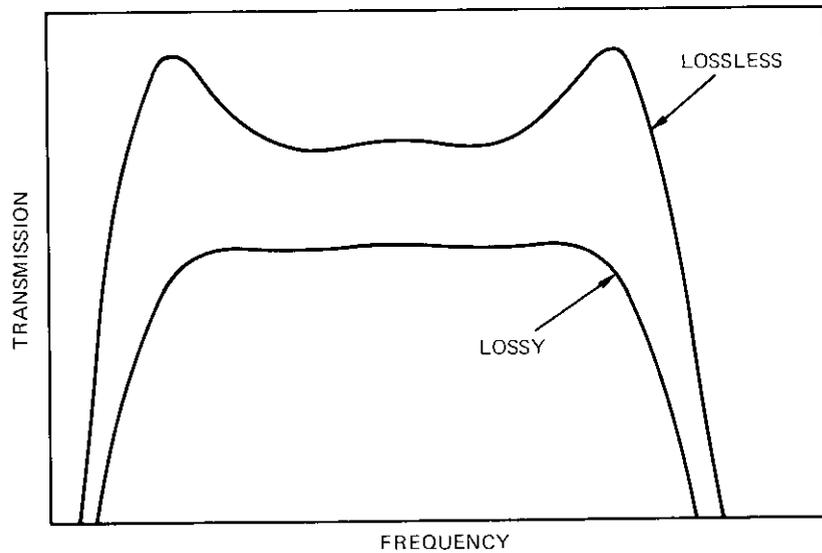


(b) TYPICAL TRANSMISSION RESPONSE

Figure 1. Typical Transmission Responses and the Pole-Zero Diagram for Lossless and Lossy Filters of Lossless Design



(a) POLE - ZERO DIAGRAM



(b) TYPICAL TRANSMISSION RESPONSE

Figure 2. Typical Transmission Responses and the Pole-Zero Diagram for Lossless and Lossy Filters of Predistorted Design

c. a lossless network is designed with respect to these new zeros and poles.

To satisfy the realizability conditions for a filter, no poles or zeros may be moved into the right half of the complex S plane. This requirement constrains the maximum allowable movement of poles in a predistorted design. Consequently, the minimum required unloaded Q may be determined. Figure 3 shows the relationship between the minimum required unloaded Q and BW for Chebychev filters with different numbers of poles.

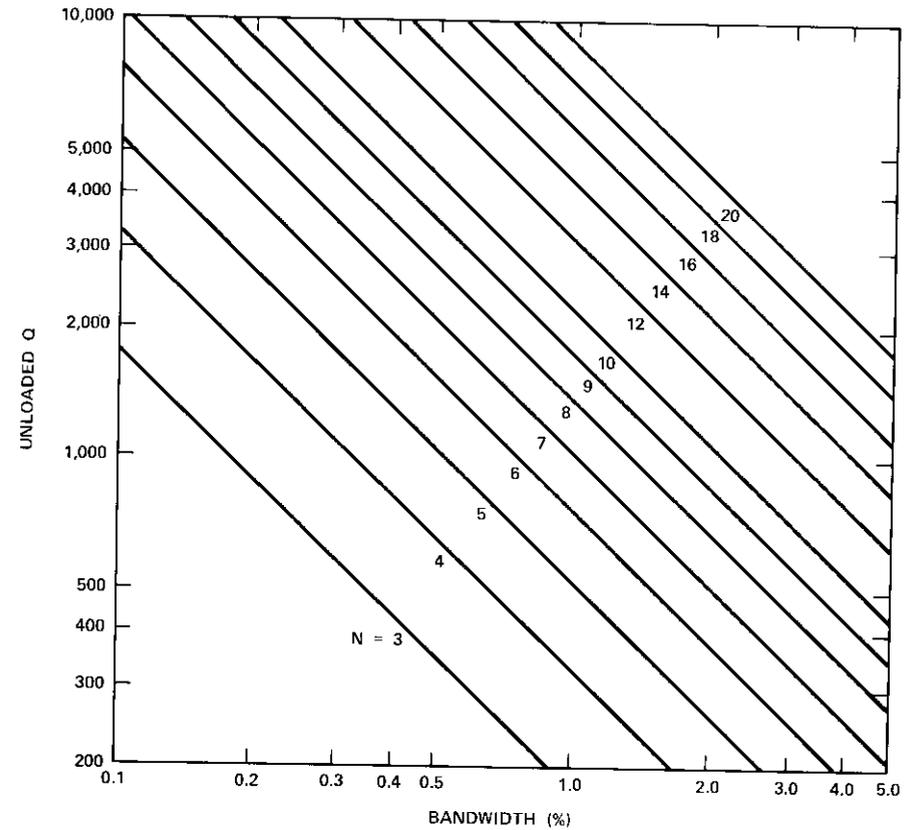


Figure 3. The Minimum Required Unloaded Q for Applying the Full Predistortion to Chebychev Filters with Different Numbers of Poles and Various Bandwidths

Any filter which has an unloaded Q higher than the value shown in Figure 3 may be predistorted to have a flat in-band response over the design bandwidth provided that additional mid-band insertion loss can be tolerated.

In a filter in which Q is lower than the minimum required value, only a partial predistortion may be applied. Although a partial predistortion cannot flatten the transmission response completely, the in-band gain slope may still be improved. Furthermore, partial predistortion yields a mid-band insertion loss which is smaller than that obtainable with full predistortion. It may then be advantageous to use partial predistortion for cases in which the mid-band insertion loss and the in-band gain slope are of concern. A tradeoff between insertion loss and in-band gain slope may be necessary.

A lossless filter which exhibits the horseshoe-shaped transmission response shown in Figure 2 does not have full transmission at center frequency. Consequently, it must have some reflection at the band center. The center frequency response of a bandpass filter is related to zero frequency in a low-pass filter inserted between the generator and a load. The required mismatch must be contributed by unequal impedances of load and generator; therefore, an asymmetrical structure will result. The transfer function for the lossless network may be derived from the new pole and zero locations resulting from the predistorted design. Thus the realization technique developed by Saito [7] may be applied to obtain an asymmetrical lossless network.

Insertion loss and selectivity

The predistorted design provides a flat in-band response in a lossy structure by restoring the correct pole and zero relationship. Therefore, out-of-band selectivity is not affected by the predistorted design. For a Chebyshev filter, there is an exact mathematical transformation between the lossless design for a lossless filter and the predistorted design for a lossy filter. This transformation enables the predistorted design to restore flat in-band response as well as good selectivity to a lossy filter.

To demonstrate the usefulness of the predistorted design two typical 10-pole Chebyshev filters have been compared, a conventional design with high Q , and a predistorted design with low Q . Both designs have a 38.4-MHz design bandwidth and a center frequency of 4 GHz. Figure 4 shows the calculated amplitude responses for both cases. It can be seen that both filters have almost identical performance with the exception of

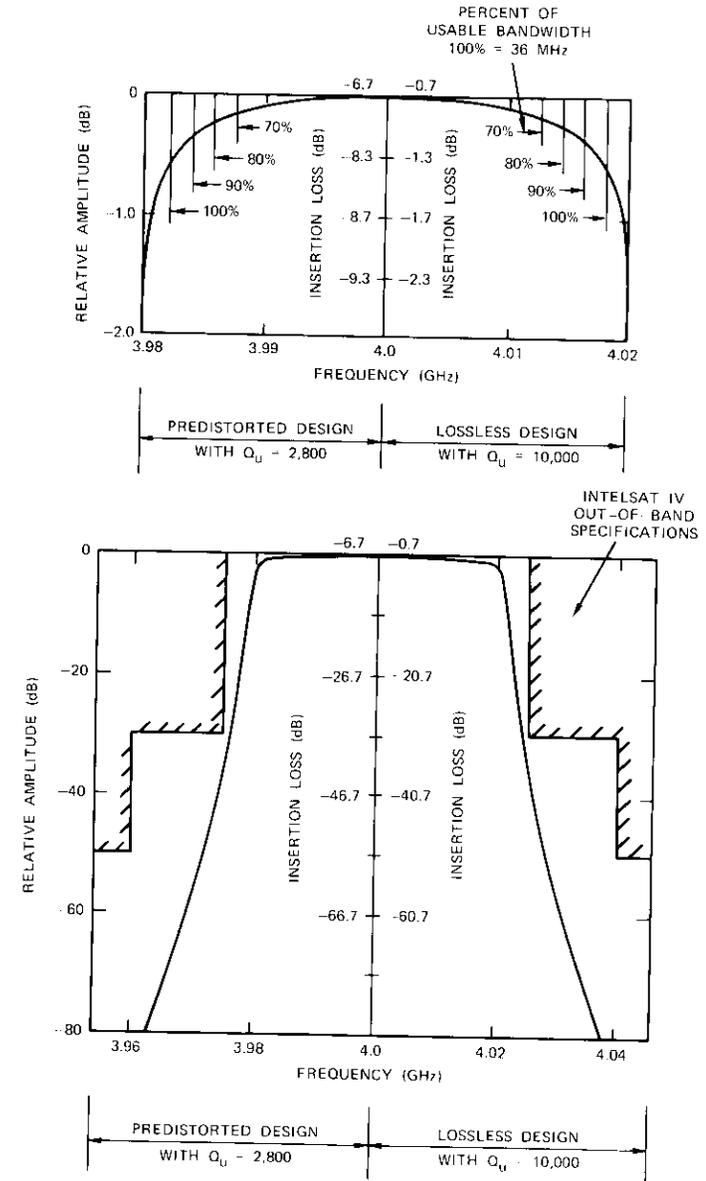


Figure 4. Comparison of Transmission Response for a High- Q Filter of Conventional Design and a Low- Q Filter of Predistorted Design

insertion loss. The predistorted filter can be realized with a cavity Q of 2,800 instead of 10,000. Figure 5 shows the relationship between ohmic losses, reflection losses, and Q for a 10-pole predistorted filter. The reflection loss is the additional mid-band loss incurred in the predistorted design due to the tradeoff between loss and flat passband.

Experimental model

To verify this design technique, an experimental model of a predistorted 3-pole Chebychev filter was designed. This filter consisted of three cavities which were nickel plated* to provide artificially high loss ($Q_u = 3,420$). The design parameters were selected as follows:

- center frequency = 4.0 GHz
- bandwidth = 5.4 MHz
- in-band ripple = 0.05 dB
- predistortion = 0.2

A 3-pole Chebychev filter with 0.05 dB in-band ripple should have three poles at $-0.569 \pm j1.311$ and -1.137 on the complex S plane according to the conventional lossless design. When the equal predistortion technique is applied, these poles will be moved to $-0.369 \pm j1.311$, and -0.937 according to the design procedure described in the second section of this paper. Therefore, the transfer function for the predistorted 3-pole filter is derived from these new poles:

$$T(s) = \frac{1}{0.668s^3 + 1.125s^2 + 1.704s + 1.158} \quad (1)$$

where
$$s = j \frac{[(f/f_0) - (f_0/f)]}{BW}$$

A coupling equation is derived from equation (1) by applying the realization procedures developed by Saito [7]:

$$\begin{bmatrix} E \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} s + 1.17 & j0.912 & 0 \\ j0.912 & s & -j1.06 \\ 0 & -j1.06 & s + 0.511 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} \quad (2)$$

*This nickel plating was used to obtain higher loss in the experimental model and hence simulate the low Q of a practical filter at higher frequencies.

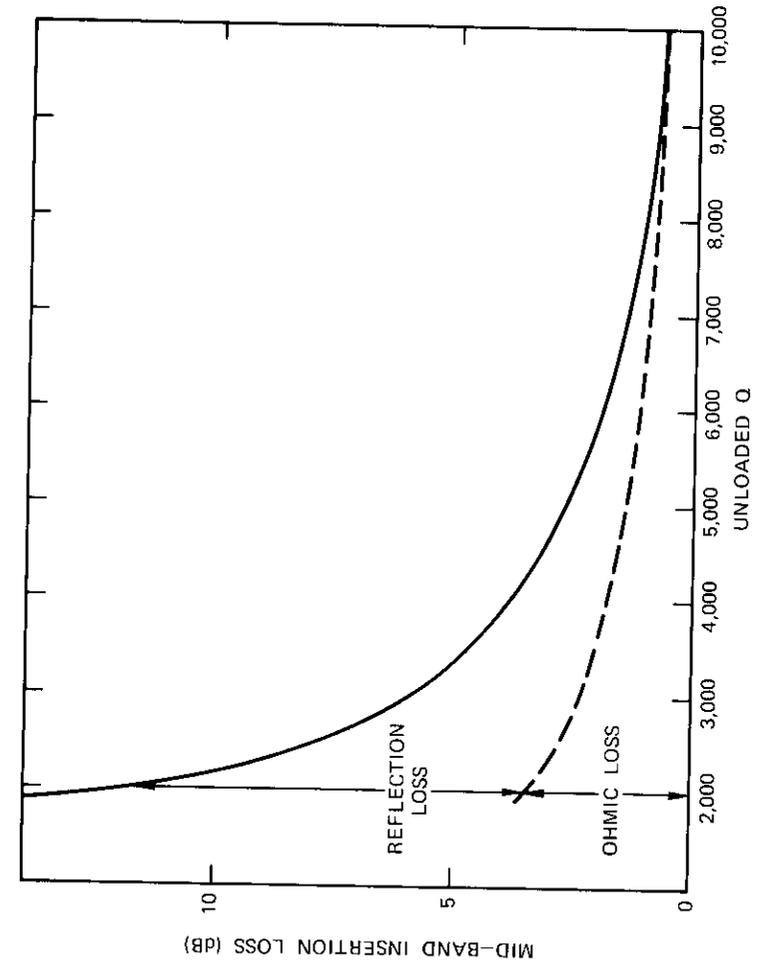


Figure 5. Required Mid-band Insertion Loss vs Cavity Q for a 10-Pole Chebychev Filter to Meet the In-band Gain Slope Specification for INTELSAT IV

Since the input impedance, the output impedance, and the coupling coefficients have been provided by equation (2), a 3-pole filter can be designed without difficulty. The configuration and the dimensions are given in Figure 6.

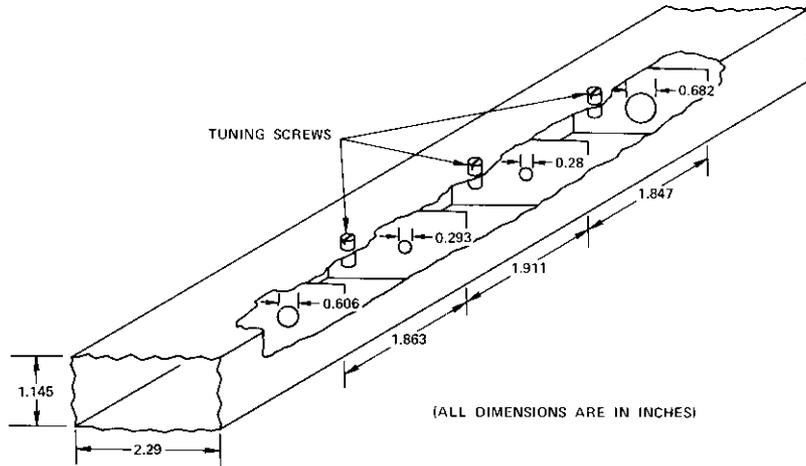


Figure 6. Dimensions of the Experimental Filter

This experimental model has achieved the calculated performance. A flat transmission response with a mid-band insertion loss of 4 dB has been obtained. The return loss, selectivity, and group delay also show good agreement with the computed results (see Figure 7). For comparison, the transmission responses of this experimental filter and those of a similar filter of conventional design are plotted in Figure 8. For example, if the useful bandwidth of the filter is defined to have a gain slope less than 0.25 dB/MHz, the predistorted design provides 100-percent usable bandwidth, while the lossless design results in only 60-percent usable bandwidth. Therefore, a significant savings in usable bandwidth can be realized with the technique described herein.

Conclusion

It has been demonstrated that the predistortion technique is applicable to lossy waveguide filter design and useful for input multiplexer applications. The predistorted design may be applied to Q -limited microwave filters to meet in-band gain slope specifications which would ordinarily require a high- Q filter. It may also be applied to a high- Q filter design to further enhance the useful bandwidth.

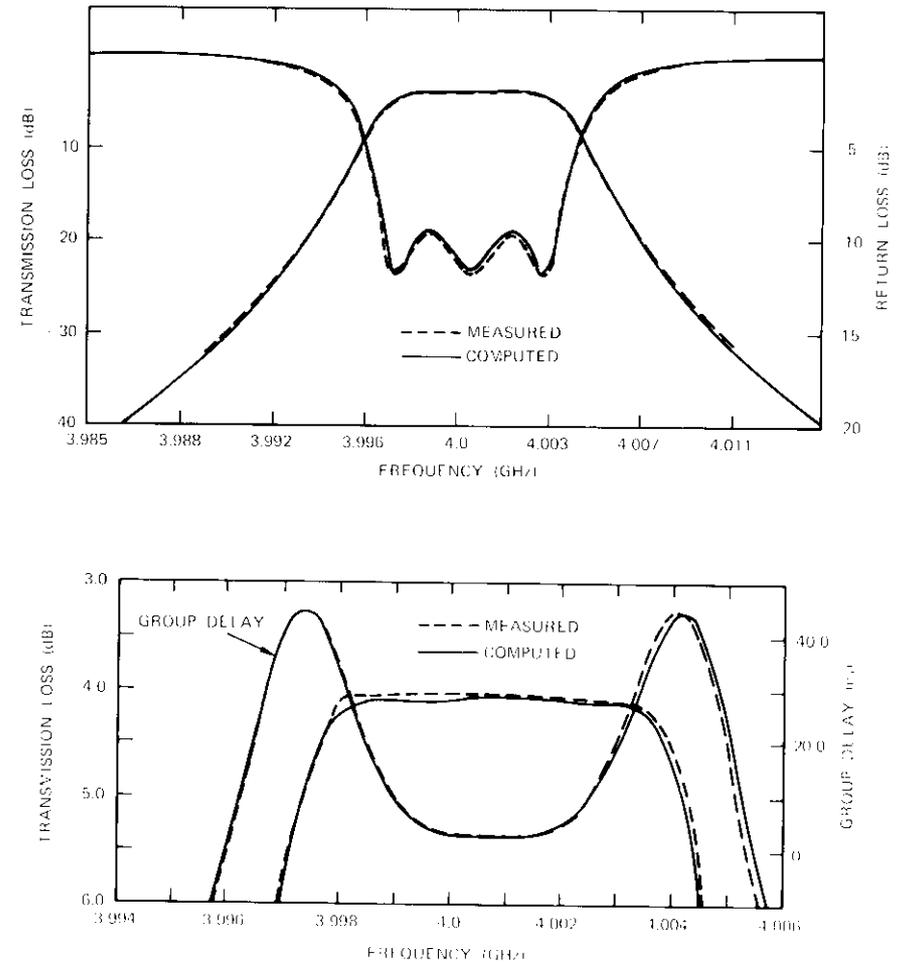


Figure 7. Computed and Measured Performance of a 3-Pole Experimental Filter with Unloaded Q of 3,420

Acknowledgment

The authors wish to thank Dr. A. E. Williams for suggesting the subject, F. Assal for helpful discussions, R. Kessler for his assistance in the construction of the experimental model, and H. Chang for his effort in cavity plating.

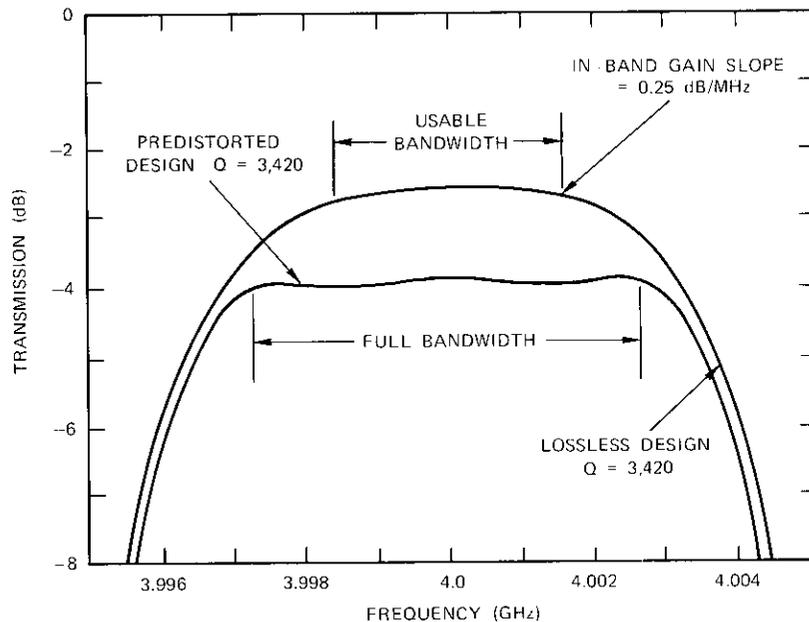


Figure 8. Transmission Responses for 3-Pole Chebychev Filters of Lossless and Predistorted Design

References

- [1] "The INTELSAT IV Communications System," *COMSAT Technical Review*, P. L. Bargellini, editor, Vol. 2, No. 2, Fall 1972, pp. 437-537.
- [2] E. T. Jilg, "The INTELSAT IV Spacecraft," *COMSAT Technical Review*, Vol. 2, No. 2, Fall 1972, pp. 271-389.
- [3] S. Darlington, "Synthesis of Reactance-four-poles which Produce Prescribed Insertion Loss Characteristics," *Journal of Mathematics and Physics*, Vol. 18, September 1939, pp. 257-353.
- [4] L. Weinberg, *Network Analysis and Synthesis*, New York: McGraw-Hill, pp. 92-95.
- [5] E. Guillemin, *Synthesis of Passive Networks*, New York: John Wiley & Sons, Inc., 1957, pp. 537-547.
- [6] J. Craig, *Design of Lossy Filters*, Cambridge, Mass.: M.I.T. Press, 1970.
- [7] M. Saito, "Synthesis Procedure for General Coupled-resonator Transmission Networks," *Electronics Communication in Japan*, Vol. 53-A, No. 5, 1970, pp. 25-33.

11.7- and 20-GHz radiometric measurements at a Brazilian equatorial site

H. J. MEYERHOFF

(Manuscript received April 22, 1975)

Introduction

To obtain statistical propagation attenuation data at 11.7 and 20 GHz in an equatorial region, an INTELSAT 3-meter receiving terminal was operated in the radiometric mode. Manaus, Brazil, was selected as a suitable site for the experiment because of the frequency of heavy rainfall during the available 6-month period from November through April.

The antenna was pointed approximately due east at the geostationary orbit at an elevation angle of around 55°. The radiometers measured increases in sky noise temperature that resulted from the passage of clouds and rain across the antenna beam. These measurements, together with rain gauge markings, time markings, and equipment calibrations, were recorded on a stripchart. The temperature readings at both 11.7 and 20 GHz were converted to equivalent fade attenuations in dB, and the recordings were eventually analyzed by hand to derive the statistical data.

Radiometric measurements of sky noise temperature at 15.3 GHz were described in a previous issue of the *COMSAT Technical Review* [1].

Satellite communications at frequencies greater than 10 GHz suffer attenuation fades due to precipitation along the satellite-earth-station paths.

The system designer must know the percentage of time during which certain attenuation levels are exceeded on the earth-station-satellite path

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

Dr. Meyerhoff is Senior Staff Scientist, System Studies Division, at COMSAT.

so that he can allow appropriate system margins to reduce the cumulative time of fade outages as well as the duration of the fades and their distribution.

It was assumed that the statistical attenuation data derived at Manaus might be applied qualitatively to other equatorial regions with similar rainfall rates. Hence, the experiment included a rain gauge, with backup rainfall measurements provided by the nearby Amazon Research Institute Meteorological Station.*

Increases in sky noise temperature were measured continuously at 11.7 and 20 GHz and recorded simultaneously on a recorder running at 5 millimeters per minute. The tipping-bucket rain gauge printed a tick on the recorder for every 0.01 inch of rain. Equipment was calibrated daily and also before and after major storms.

Analysis of propagation data

The increases in noise temperature that were recorded on the strip-charts at 11.7 and 20 GHz were converted to attenuation fades using calibrated measurements closest to the storms. The chart was visually scanned to yield statistical propagation attenuation data. Figure 1 gives the attenuation statistics for the 5-month period from November 30th through April 27th, and the worst month statistics at both 11.7 and 20 GHz. During this period there was an average of 10 major storms† per month (10 in December, 9 in January, 11 in February, 7 in March, and 12 in April). An approximate relationship between duration and number of storms was found to be

$$T \approx 1,000 \times 10^{-n/a} \quad (1)$$

where T = storm duration in minutes
 n = number of storms exceeding that duration
 a = factor which varies for each of the given months with a mean of 6.3 and a standard deviation of 1.5.

Recorder traces of representative storms are shown in Figure 2.

*However, the usefulness of the rain data was limited because the rain gauge was unreliable in heavy storms.

†A major storm is defined here as one which has a fade greater than 8 dB for the 20-GHz measurement.

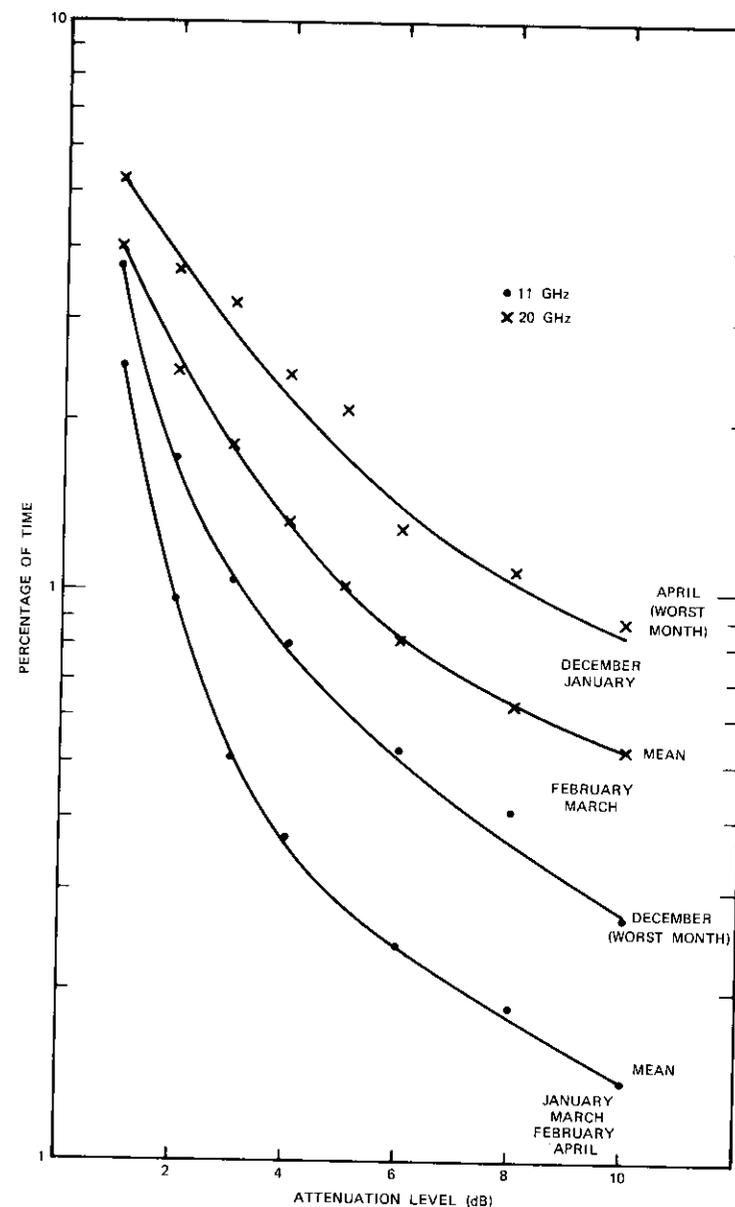


Figure 1. Attenuation Statistics for November 30, 1973, through April 27, 1974

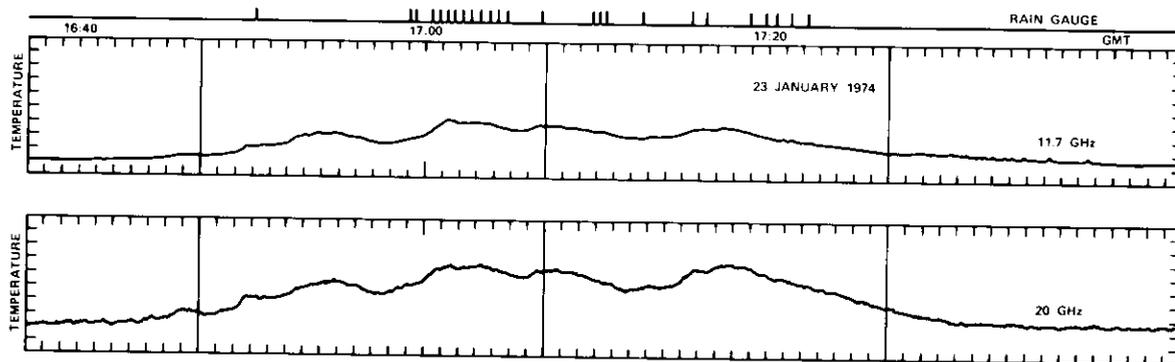
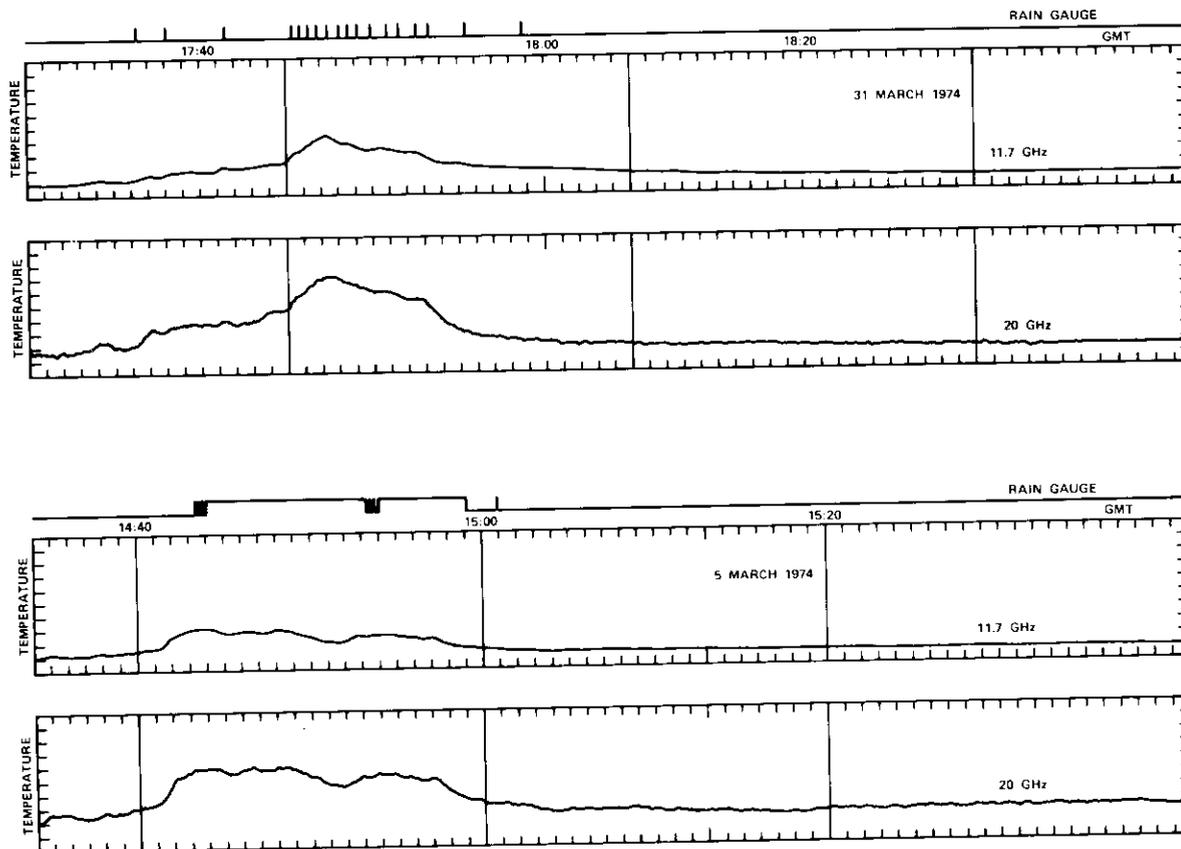


Figure 2. Stripchart Recordings for Typical Storms

Major storm histograms correlate with the attenuation statistics shown in Figure 1. Statistical attenuation data are also presented in the average monthly fade duration histograms for the five months. The approximate relationship between fade duration, T , in minutes, the number of fades, n , exceeding T minutes in duration, and the fade attenuation, x , in dB is

$$T \approx 70 \times 10^{-nz/b} \quad (2)$$

where b varies for each month with a mean of 81 and a standard deviation of 33.

The relationship between the number of major storms, n_s , and the number of fades of x -dB intensity, n_f , is derived as

$$n_f x \approx 13n_s - 94 \quad (3)$$

The daily histograms of major storms taken over the 5-month period are presented in Figure 3.

Interpretation of statistical data

CORRELATION OF ATTENUATION FADES WITH HEAVY RAINFALL RATES AT THE EARTH STATION

An examination of the major storms indicates that in many cases severe attenuation has been recorded prior to the registration of heavy rainfall. This can be explained readily by the fact that radiometric measurements of increases in sky noise temperature are related to rainfall rates, not only at the earth station site where the rain gauge is located, but also along the satellite-earth-station path. However, in most cases, the inception of the increase in sky noise temperature coincides with the observer's recording of the commencement of rainfall, usually at a rate which is too slow to activate the rain gauge.

CORRELATION OF ATTENUATION FADES WITH RAINFALL RATES AT A NEARBY METEOROLOGICAL STATION

The designer of a communications satellite system that is sensitive to rain attenuation might use as a guide the rainfall statistics for a site near a proposed earth station location. In fact, this type of data was used to select Manaus as a representative equatorial site. The total monthly rainfall at the nearby meteorological station is indicated in Table 1.

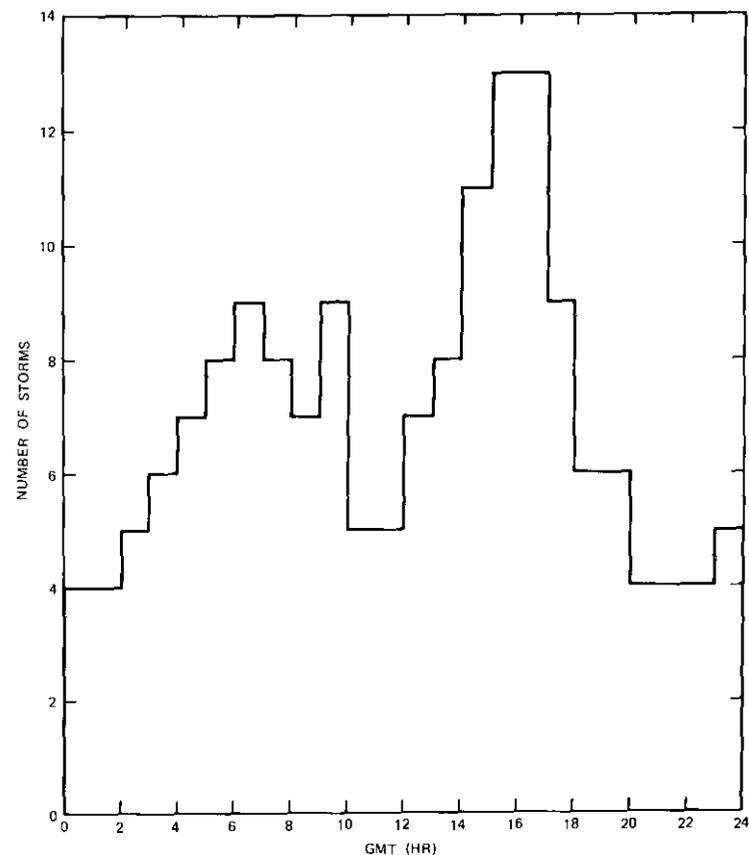


Figure 3. Distribution of Major Storms over a 24-Hour Period (December 1973-April 1974)

TABLE 1. MONTHLY RAINFALL AT A METEOROLOGICAL STATION NEAR MANAUS

Month	Rainfall (mm)
December 1973	322
January 1974	142
February 1974	249
March 1974	265
April 1974	335

The meteorological station also recorded the time intervals during which rain fell so that a statistical profile of estimated rain rate could be derived. The approximate relationship between time, T , in minutes and rainfall rate, R , in mm/hr was

$$T \approx c \times 10^{-R/7} \quad (4)$$

where c varied for each month with a mean of 3.1 and a standard deviation of 2.5.

CORRELATION OF ATTENUATION FADES AT 11.7 AND 20 GHz

Examination of stripchart recordings for all major storms indicates correlation between increases in sky noise temperature at 11.7 and 20 GHz. Measurements at 11.7 GHz are less subject to saturation, since the approximate 10-dB dynamic range of the radiometers is more often exceeded at 20 GHz. Whereas the correlation between attenuation fades at the two frequencies is good, the quantitative relationship between 11.7- and 20-GHz attenuation fades varies considerably. That is, a fade of 2 dB at 11.7 GHz may correspond instantaneously to a fade of between 4 and 10 dB at 20 GHz for different storms. It is suspected that this discrepancy is due mainly to the inaccuracies in the radiometric measurements of sky noise temperature and the estimation of attenuation from these measurements. This discrepancy is also evident in the attenuation statistics of Figure 1, where April corresponds to the worst month at 20 GHz and the best at 11.7 GHz.

The mean curves in Figure 1 were used to correlate attenuation at these two frequencies; the result is shown in Figure 4. These data concur somewhat with the relationship between rain attenuation at different frequencies as indicated in C.C.I.R. Document 5/154-E (Figure 5). A comparison of C.C.I.R. data with those measured at Manaus is presented in Table 2.

TABLE 2. COMPARISON OF C.C.I.R. AND MANAUS RAIN ATTENUATION DATA

Rain Rate (mm/hr)	Attenuation (dB)		Ratio	
	11.7 GHz	20 GHz	Manaus	C.C.I.R.
5.6	1	2	2	4.5
7.2	1.5	3.5	2.33	4.2
8.5	2	5.2	2.6	3.7
9.6	2.5	7.6	3	3.4
10.1	2.8	9.5	3.4	3.3

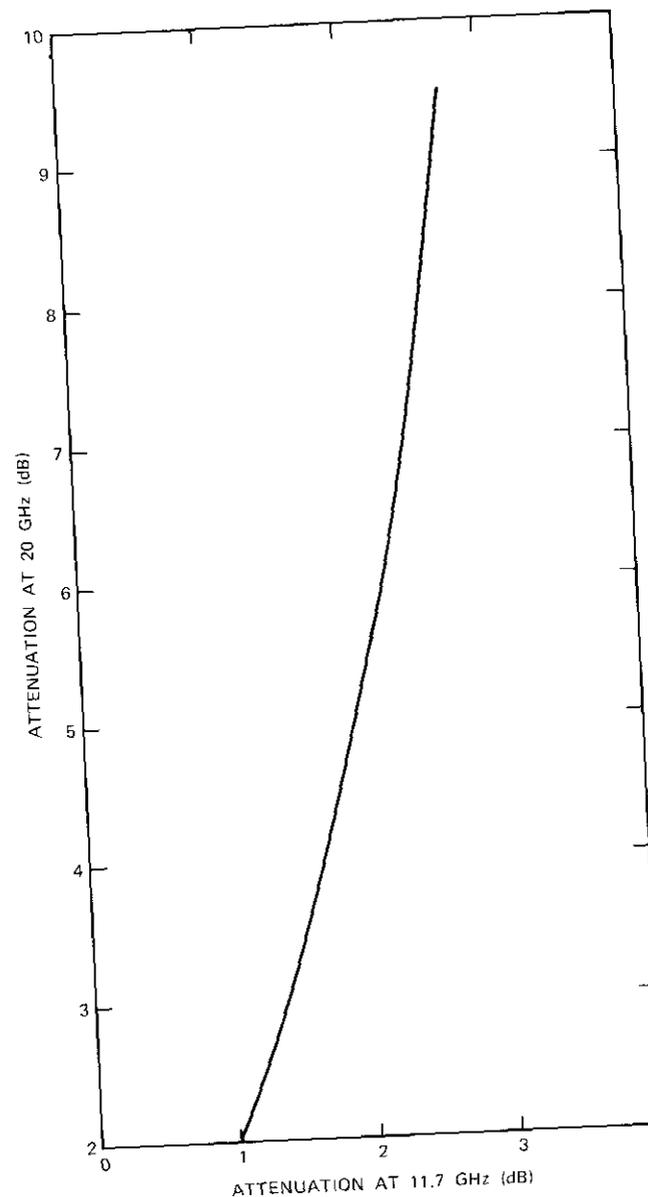


Figure 4. Correlation of 11.7- and 20-GHz Attenuation over Five Months

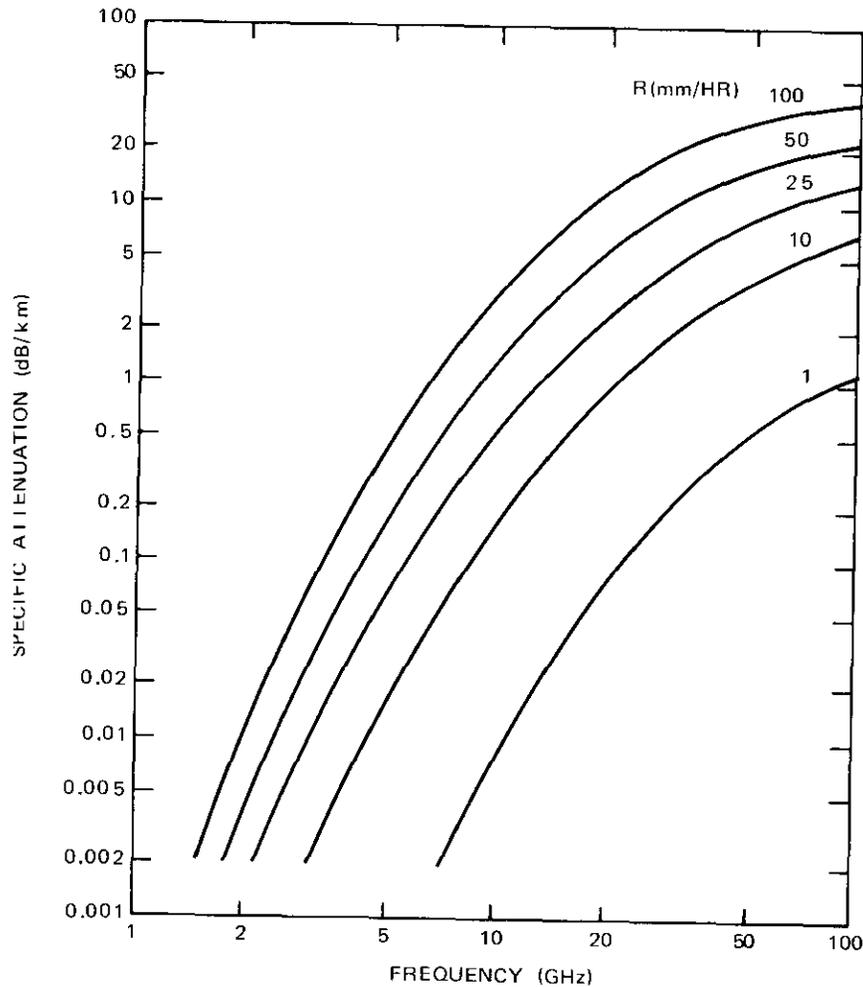


Figure 5. Variation of Specific Attenuation with Frequency

Evaluation of experimental deficiencies

The radiometric method of obtaining precipitation attenuation data is severely limited in terms of dynamic range and accuracy. However, in the absence of a satellite source, this method is perhaps the best available

for obtaining statistical information for satellite systems calculations. Possible improvements in the measuring and recording techniques are suggested in the following subsections.

Attenuation fades less than 1 or 2 dB occur over 95 percent of the time, and need not be recorded. The remaining data are significant and should be recorded at a speed slower than 5 millimeters per minute.

The exponential relationship between sky noise temperature and attenuation fade leads to cramping of the scale at higher attenuation levels. The accuracy of the recordings can be improved by compensating for this exponential characteristic (see Figure 6). The scale of the recordings should also be twice that used in the experiment described herein so that the statistical data can be more easily read.

The experiment has indicated the low reliability of the tipping-bucket rain gauge in very heavy rain and the difficulty of reading stripchart recordings of heavy rain rates when the apparatus was functioning. A slower chart speed with the threshold device recommended previously will greatly improve the method of rain rate recording, but the measurement instrument must be made more reliable.

Comparison with a similar experiment at Utibe, Panama

Qualitatively the heavier rainfall in the equatorial region (Manaus) over the 5-month period yielded attenuation statistics at 11.7 GHz which were similar to those obtained over a 9-month period at 15.3 GHz in a tropical region (Utibe). For various levels of availability, the corresponding margin at these two sites is given in Table 3.

TABLE 3. MARGIN VS AVAILABILITY AT MANAUS AND UTIBE

Availability (%)	Margin (dB)	
	Manaus, 11.7 GHz	Utibe, 15.3 GHz
99	2	1.5
99.3	2.3	2.2
99.5	3	3.2
99.6	3.7	4
99.7	4.7	5.2
99.8	7.4	7.3
99.85	9.5	8.7

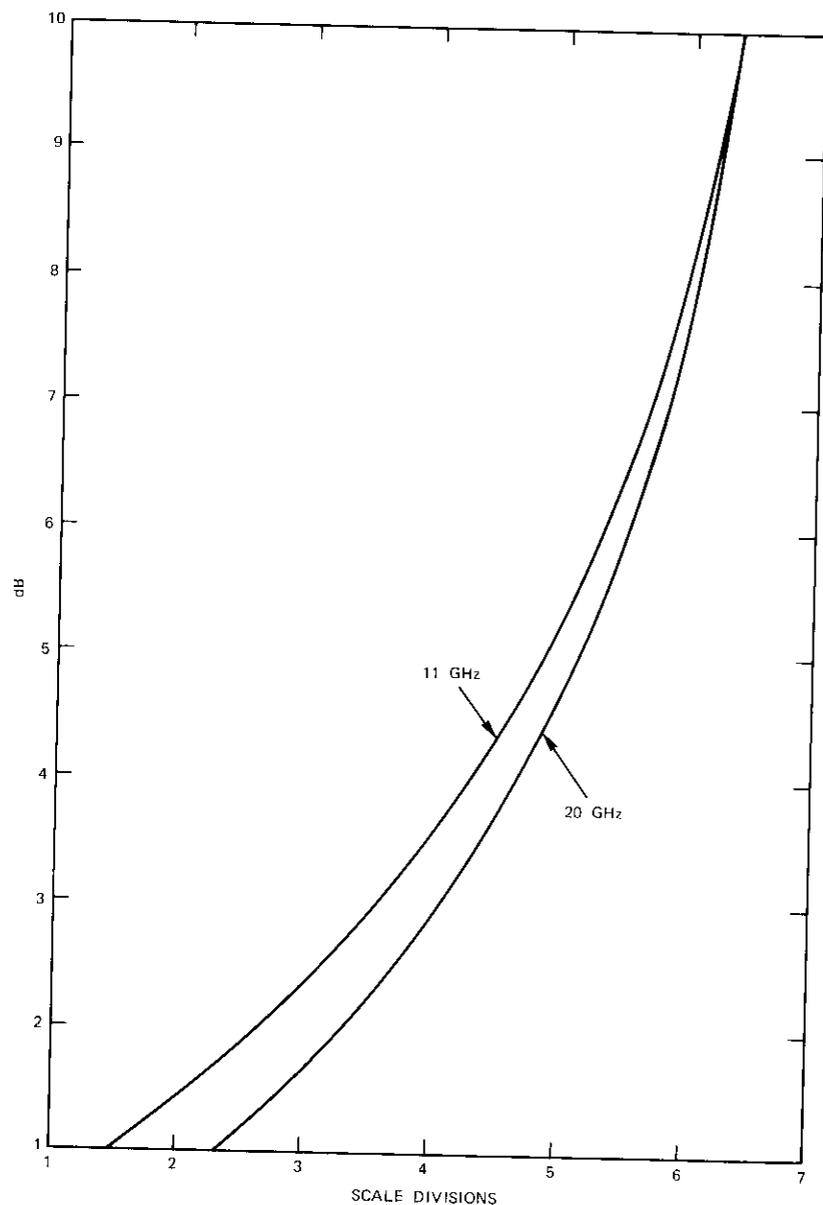


Figure 6. Typical Sky Noise Temperature Attenuation on a Decibel Conversion Scale

Conclusion

Radiometric measurements of sky noise temperature at 11.7 and 20 GHz were conducted at an equatorial site during five rainy months using a 3-meter receiving terminal. Various simple analytical expressions derived from histograms of major storms and related attenuation fades showed correlation between attenuation at the two frequencies.

Qualitatively the heavy rainfall in the equatorial region of Manaus yielded many deep attenuation fades and attenuation statistics that were considerably more detrimental than those obtained from a previous experiment carried out at a tropical site near Utibe, Panama.

Acknowledgment

Transportation, installation, and operation of the equipment were carried out successfully by A. H. da Costa, J. Ornelas, and the EMBRATEL staff at Manaus. Manual data reduction was executed by K. Mekhayarajjanon with assistance by K. Cain, while the overall equipment was the responsibility of W. Mercer.

Reference

- [1] H. J. Meyerhoff, A. J. Buige, and E. A. Robertson, "15.3-GHz Precipitation Attenuation Measurements Using a Transportable Earth Station at Utibe, Panama," *COMSAT Technical Review*, Vol. 4, No. 1, Spring 1974, pp. 169-186.

Translations of Abstracts in this issue

Comparaison entre les cellules solaires Violettes et les cellules solaires Non Réfléchissantes de COMSAT

J. F. ALLISON, R. A. ARNDT, ET A. MEULENBERG

Sommaire

Cet article compare les propriétés de la cellule solaire Violette et celles de la cellule améliorée de deuxième génération appelée cellule COMSAT Non Réfléchissante. Ces propriétés comprennent les caractéristiques courant-tension, les coefficients de température, la réflectivité par rapport à la longueur d'onde, la réponse spectrale, l'énergie produite par rapport à un flux intégré d'électrons de 1-MeV, l'énergie produite en fonction de l'angle d'incidence de la lumière, le facteur d'absorption d'énergie solaire et le pouvoir émissif énergétique, ainsi l'adhésion au substrat. L'énergie produite par la cellule violette est supérieure d'au moins 30 pour cent à celle des cellules de satellite classiques, tandis que la cellule non réfléchissante donne une amélioration de plus de 50 pour cent. Dans l'un et l'autre cas, cette amélioration relative se maintient lorsque la cellule est exposée aux rayons ultraviolets et aux rayonnements pénétrants spatiaux.

Analyse de l'étalement des spectres de puissance des Signaux MDP par un modèle de chaîne de Markoff

C. DEVIEUX

Sommaire

On emploie un modèle de chaîne de Markoff pour calculer les spectres de signaux MDP à deux états de phase lorsque les voies sont linéaires et