

File Paul Lee Per
MSR microwave switch valve (MSM)
810802 CH
PNT 67m. has band to know journal entry
1974

2

COMSAT

Technical Review

Volume 1 Number 1, Fall 1971

Advisory Board Joseph V. Charyk
 William W. Hagerty
 Sidney Metzger
 Wilbur L. Pritchard

Editorial Board Pier L. Bargellini, Chairman
 Robert C. Barthle
 N. K. M. Chitre
 S. H. Durrani
 Leonard Golding
 Gary D. Gordon
 Geoffrey Hyde
 Joachim Kaiser
 Edward J. Martin
 Emeric Podraczky
 Edmund S. Rittner
 Reinhard Stamminger

Volume 1 Number 1, Fall 1971

The COMSAT TECHNICAL REVIEW is published twice a year by Communications Satellite Corporation, 950 L'Enfant Plaza, S.W., Washington, D. C. 20024. Editors: Lawrence Weekley, Jr., Gerald H. Bidlack, Leonard F. Smith.

© COMMUNICATIONS SATELLITE CORPORATION 1971

- i A STATEMENT OF PURPOSE **Joseph V. Charyk**
- 1 EXPERIMENTAL INVESTIGATIONS ON THE FUEL SLOSH OF DUAL-SPIN SPACECRAFT **Ernesto R. Martin**
- 21 NEW TYPES OF WAVEGUIDE BANDPASS FILTERS FOR SATELLITE TRANSPONDERS **A. E. Atia** and **A. E. Williams**
- 45 FREQUENCY-SHARING CONSIDERATIONS BETWEEN GEO-STATIONARY COMMUNICATION SATELLITES AND TERRESTRIAL RADIO RELAYS **J. C. Fuenzalida** and **N. K. M. Chitre**
- 79 A COMPUTER-CONTROLLED SATELLITE SIGNAL MONITORING SYSTEM **E. E. Steinbrecher** and **L. F. Gray**
- 117 A MODEL FOR NONPENETRATING PROTON DAMAGE TO SILICON SOLAR CELLS **R. A. Arndt** and **L. H. Westerlund**
- 139 ANALYSIS OF ADJACENT CHANNEL INTERFERENCE IN A MULTICARRIER FM COMMUNICATIONS SYSTEM **Marvin Wachs**
- 171 THE SPADE SYSTEM AS APPLIED TO DATA COMMUNICATIONS AND SMALL EARTH STATION OPERATION **Eugene R. Cacciamani, Jr.**
- 183 APPLICATIONS OF ERROR-CODING TECHNIQUES TO SATELLITE COMMUNICATIONS **W. W. Wu**
- 221 CTR NOTES: ATTENUATION STATISTICS AT 15.3 GHZ FOR CLARKSBURG, MARYLAND **H. D. Craft, Jr.**
- 227 TRANSLATIONS OF ABSTRACTS IN THIS ISSUE
 FRENCH 227
 SPANISH 232

A STATEMENT OF PURPOSE

After six years of successful commercial communications via satellite, it is clear that this new technology is making, and will continue to make, significant contributions to human affairs. Although satellite communications are the product of important scientific and technical advances, there is no technical periodical anywhere devoted exclusively to this subject.

COMSAT is privileged to have played a pioneering role in establishing the global commercial communications satellite system, and in the advancement of a satellite technology which is steadily growing in complexity and magnitude. The technical achievements so far have been impressive indeed. But achievement of the full potential of satellite communications depends in large measure on research and technical work yet to be done.

It is appropriate, therefore, that this Corporation, with extensive research and technical capabilities of its own, undertake publication of the first technical periodical devoted exclusively to satellite communications. This is the first issue of that journal, the COMSAT TECHNICAL REVIEW. The goals are:

- 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.

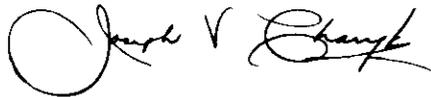
The COMSAT TECHNICAL REVIEW will cover scientific and technical aspects of satellite communications in the areas of spacecraft, earth

EXPERIMENTAL INVESTIGATIONS ON THE FUEL SLOSH OF DUAL-SPIN SPACECRAFT

ERNESTO R. MARTIN

stations and communications processing, including systems, components and techniques. The spectrum of papers will be broad, ranging from the highly specialized to the general and tutorial. Papers selected for publication by the Editorial Board will be characterized by high standards of quality, originality, timeliness and permanence of value.

It is our high hope that the COMSAT TECHNICAL REVIEW will contribute to greater technical understanding and achievement in satellite communications.



JOSEPH V. CHARYK
President

Communications Satellite Corporation
Washington, D. C.
November 19, 1971

ABSTRACT

Energy dissipated by vibrating components on the spinning rotor of a dual-spin spacecraft, such as the INTELSAT IV communications satellite, has a destabilizing effect on the dynamics of the vehicle. Experimental investigations to determine fuel slosh energy dissipation revealed unexpectedly turbulent fluid motions in the sphere-cone propellant tanks of INTELSAT IV. The liquid motion was observed inside a spinning tank subjected to angular oscillations similar to those of the spacecraft. Earlier tests in which the spinning tank was subjected to rectilinear vibration, but similar linear accelerations, had resulted in calm, rigid-body, pendulous fluid motion. It was found that the turbulent fluid motion is excited when a spinning tank is subjected to angular oscillations about an axis which is not parallel to the symmetry axis of the tank.

INTRODUCTION

A dual-spin satellite consists of a nonspinning platform coupled, through a bearing interface, to a rotor which spins about the minimum moment-of-inertia axis of the spacecraft. This unique design concept embodies several characteristics desirable for synchronous communications satellites. Among these are spin stabilization to provide gyroscopic stiffness and a nonspinning platform from which directional antennas can beam the communications power toward the earth.

In order to keep the antenna beams directed toward the earth, two conditions are necessary:

- a. the platform on which the antennas are mounted must be kept properly despun relative to the earth, and
- b. the spin axis of the rotor must remain inertially fixed.

This paper is based upon work performed under the sponsorship of the International Telecommunications Satellite Consortium (INTELSAT). Any views expressed in this paper are not necessarily those of INTELSAT.

The first condition is achieved through earth sensors and an internal control system which governs the torque of an electric motor housed in the bearing interface between the platform and the rotor. The second condition is dependent on the spacecraft dynamic stability. If the spacecraft is nutating, the coning motion of the spin axis prevents the antennas from pointing at desired targets on the earth. In the absence of external forces, the nutation angle, θ , is given by

$$\theta = \theta_0 e^{-t/\tau_s} \quad (1)$$

where t = time
 θ_0 = nutation angle at time zero
 τ_s = spacecraft time constant.

For the spacecraft to be stable, τ_s must be positive.

Using energy sink methods, other investigators [1], [2] had earlier derived the necessary conditions for τ_s to be positive. For a dual-spin vehicle with a frictionless bearing interface and axisymmetric rotor and despun platform, the necessary condition is

$$\dot{E}_r < \left[\frac{1 - (I_r/I_T)}{I_r/I_T} \right] \dot{E}_d \quad (2)$$

where \dot{E}_r = the rate of energy dissipated by vibrating components in the rotor

I_r = the moment of inertia of the rotor about the spin axis

I_T = the moment of inertia of the spacecraft about a transverse axis

\dot{E}_d = the rate of energy dissipated by vibrating components on the despun platform.

Energy dissipating passive nutation dampers mounted on the platform can provide the necessary \dot{E}_d . Stability is assured whenever equation (2) is satisfied.

Indeed, the first dual-spin spacecraft ever launched, TACSAT, exhibited an undamped nutation of about 1° [3] induced by excessive energy dissipation in its rotor. Soon afterward, the firm of Harrington, Davenport and Curtis, Inc., consultants to COMSAT, initiated a study to predict the stability of the INTELSAT IV communications satellite (Figure 1) which was being designed by the Hughes Aircraft Company. It was found that the major uncertainty was the magnitude of the energy dissipated by the hydrazine liquid propellant in the four sphere-cone

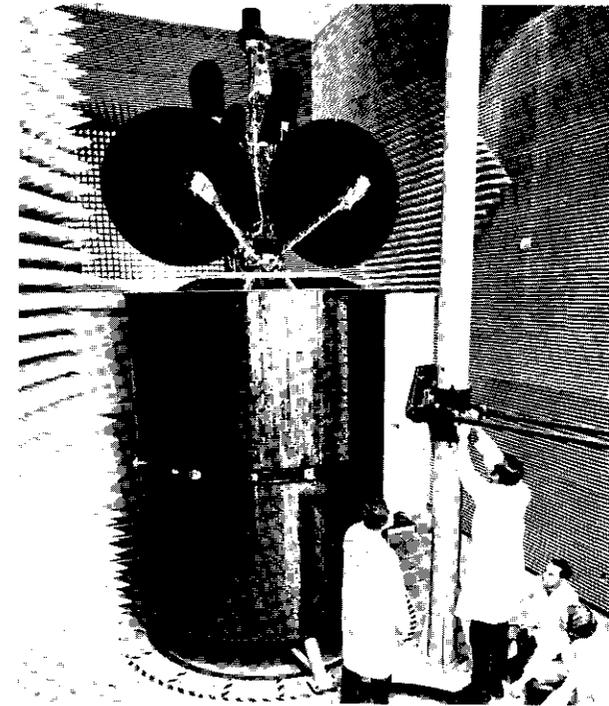


Figure 1. INTELSAT IV

tanks housed in the rotor. Therefore, it was recommended that the energy dissipation rate (\dot{E}) of the fuel slosh be investigated.

While intensive analytical studies were continued by Harrington, Davenport and Curtis, Inc., an experimental program was initiated at COMSAT Laboratories. Hughes Aircraft Company also initiated experimental investigations to measure the fuel slosh \dot{E} on INTELSAT IV. The Hughes program used an air-ball-test technique similar to the one Hughes had previously used for TACSAT and ATS-E. This paper, however, reports only the work performed at COMSAT Laboratories. COMSAT's objective was to measure the \dot{E} of the fuel in an INTELSAT IV tank subjected to motions similar to those it experiences on the spacecraft.

TEST PROGRAM: LINEAR MOTION

A test fixture was desired which would simulate the spacecraft excitation on the tank. A dynamic analysis of a general dual-spin spacecraft was performed and is presented in Appendix A. The body-fixed co-

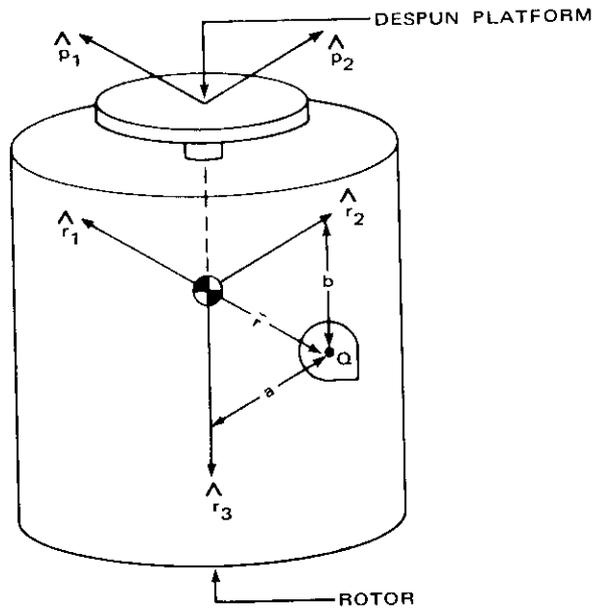


Figure 2. Dual-Spin Spacecraft Showing Coordinate System Used

ordinate system and overall configuration are shown in Figure 2. The sphere-cone tank has the outlet at the cone apex to allow full propellant depletion in both ground and orbital conditions. The center of the spherical section is defined as the center of the tank and labeled Q. The origin of the coordinate system is coincident with the center of gravity (CG) of the spacecraft. The position vector from the origin to the tank center is

$$\bar{r}_Q = a\hat{r}_2 + b\hat{r}_3 \quad (3)$$

where \hat{r}_2 and \hat{r}_3 are unit vectors in the direction shown in Figure 2. The linearized acceleration experienced by Q is equal and opposite to \bar{a}_Q , which is derived with respect to an inertial frame in Appendix A; i.e.,

$$\begin{aligned} -\bar{a}_Q = & \hat{r}_1(b\sigma^2\omega_s^2\theta \sin \lambda't) \\ & + \hat{r}_2(-b\sigma^2\omega_s^2\theta \cos \lambda't + a\omega_s^2) \\ & + \hat{r}_3[a\sigma\omega_s^2\theta(\sigma - 2) \cos \lambda't] \end{aligned} \quad (4)$$

where σ is the moment-of-inertia ratio (I_r/I_T), ω_s is the rotor spin rate, and λ' is the rotor rotation frequency given by

$$\lambda' = (-1 + \sigma)\omega_s \quad (5)$$

The first spinning test selected is shown in Figure 3. A plexiglass INTELSAT IV tank is driven with linear sinusoidal motion by the onboard motor and cam arrangement at an angle ψ from the horizontal. This angle is chosen so that the constant acceleration, which is made up of gravity and the centripetal acceleration, is perpendicular to the direction of motion. The tank is also canted by an angle ψ to preserve the same relationship between liquid free surface and tank geometry that exists in the spacecraft. The liquid is viewed by an onboard television camera with variable mounting and lenses, remotely monitored on a TV screen, and recorded on a video tape recorder. Instrumentation includes a tachometer, a pulse-per-revolution calibrator, a linear displacement transducer, and an accelerometer with its sensing axis parallel to the \hat{r}_1 axis. Slip rings allow transfer of power and signals to and from the table.

When the linear motion of the tank has an amplitude X and time function $\cos \lambda't$ at an angle ψ from the horizontal, and when the spin rate is ω_s , the linearized acceleration experienced by the center of the tank in the presence of the gravitational acceleration, g , is found to be (Appendix B)

$$\begin{aligned} g\hat{r}_3 - \bar{a}_Q = & [2X\omega_s\lambda' \cos \psi \sin \lambda't] \hat{r}_1 \\ & + [(\lambda'^2 + \omega_s^2) X \cos \psi \cos \lambda't + a\omega_s^2] \hat{r}_2 \\ & + [g - X\lambda'^2 \sin \psi \cos \lambda't] \hat{r}_3. \end{aligned} \quad (6)$$

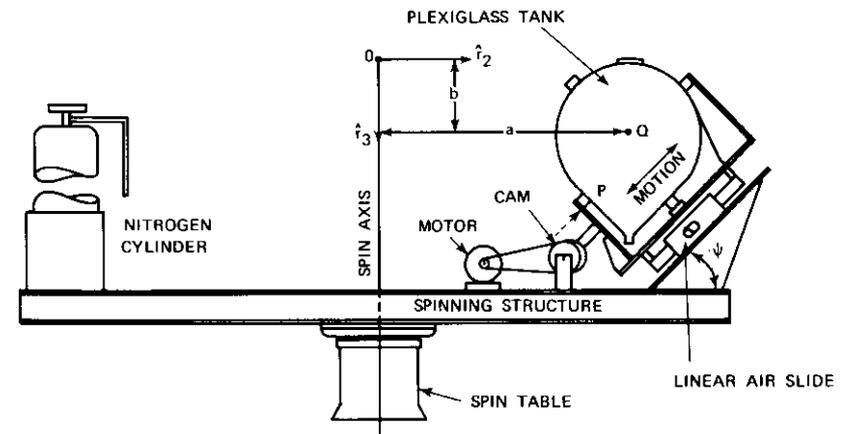


Figure 3. Linear Motion Spinning Test Fixture

Notice that, although the tank's motion is in the \hat{r}_2, \hat{r}_3 plane, there is an acceleration component in the \hat{r}_1 direction because of Coriolis effects.

Table 1 is a numerical comparison of these harmonic accelerations with those of INTELSAT IV (where $a = 23.5$ in., $b = 11$ in., and mid-life $\sigma = 0.33$) for a spin speed of 60 rpm, 1° nutation, ψ equal to the test value of 73° , and the X amplitude chosen so that the \hat{r}_3 acceleration is the same as that on the spacecraft. In the \hat{r}_1, \hat{r}_2 directions, the accelerations on the test fixture are about 10 times greater than they are on the spacecraft and in the opposite direction. Notice, however, that the amplitude and phase relationship between the \hat{r}_1 and \hat{r}_2 accelerations on the test fixture is similar to the relationship between the \hat{r}_1 and \hat{r}_2 accelerations on the spacecraft. Therefore, the test fixture, which is thought to provide an adequate simulation of the spacecraft's linear accelerations, yields conservative results because of the much greater amplitude of the \hat{r}_1 and \hat{r}_2 excitations.

TABLE 1. LINEARIZED ACCELERATIONS FOR THE
LINEAR MOTION TEST FIXTURE AND THE
SPACECRAFT

Direction	Linearized Acceleration, in./sec ²	
	Linear Motion Spinning Test	Spacecraft
\hat{r}_1	$-8.14 \sin \lambda't$	$0.83 \sin \lambda't$
\hat{r}_2	$8.81 \cos \lambda't$	$-0.83 \cos \lambda't$
\hat{r}_3	$-8.92 \cos \lambda't$	$-8.92 \cos \lambda't$

While the fixture was being designed and fabricated, one of the INTELSAT IV tanks was subjected to nonspinning unidirectional vibration of amplitudes and frequencies typical of the spacecraft. The objectives of these preliminary tests were to evaluate the effect of the sphere-cone tank shape on fuel \dot{E} and to obtain baseline \dot{E} values. The spherical section of the tank had a diameter of 17.4 inches and the conical section had an angle of 86° (axis to wall equals 43° , as shown by the tank of Figure 3).

The fluid \dot{E} measured in the vibration tests [4] was in reasonably good agreement with analytical models which assumed rigid-body fluid motion. The damping ratio of the liquid* was similar to that of a fluid in a sphere of the same diameter.

* Water was used for all the tests because its viscosity, density, and surface tension are very similar to those of hydrazine.

Results: Fluid Motion

The first series of tests was designed to study the fluid behavior and to compare it with that predicted by the rigid-body analysis. The camera was located to view the entire liquid. Free-surface motion was observed by introducing numerous floating particles, while bulk motion was studied by dropping dye on the surface and observing its migration and motion below the surface.

Throughout the range of spin speed and nutation frequency of INTELSAT IV, the liquid was very calm. Surface particles exhibited rectilinear oscillations at the driving frequency with amplitudes in the immediate neighborhood of the tank amplitude. The dye below the surface appeared almost motionless and its migration through the liquid was very slow. As would be expected, when the driving frequency approached the natural frequency of the fluid (about three times greater than the nutation frequency), its motion increased significantly.

Results: Measurement of Fluid \dot{E}

The camera is positioned to view closely point P of the tank in the direction shown by the dashed arrow of Figure 3. By measuring the oscillation amplitude of the liquid surface where it contacts the tank wall at P for different driving frequencies and tank amplitudes, one can plot the vibrational potential energy (E_p) of the liquid as a function of driving frequency and tank amplitude. Tests in which the tank is abruptly stopped and the decay history of the liquid oscillation is recorded permit calculations of its time constant, τ , and the natural frequency, ω_n . Since the fluid exhibits rigid-body pendulous motion, its energy dissipation rate is given by

$$\dot{E} = \frac{2E_p}{\tau} \left(\frac{\lambda'}{\omega_n} \right)^2 \quad (7)$$

This is the equation used to compute the fuel slosh \dot{E} from the data obtained from the described test sequences [4].

The tests performed were limited to a propellant tank which was one-third full (35 lbs.) because the \dot{E} is fairly independent of mass when the fluid exhibits pendulous motion and when the excitation frequency is smaller than the liquid's natural frequency. The results of \dot{E} vs. λ' for four discrete nutation angles are presented in Figure 4 for a spin speed of 48 rpm. Limitations in the equipment did not allow measurements at different spin speeds. The driving frequency, λ' , in terms of the inertia ratios for INTELSAT IV from start of life to end of life, is presented on the abscissa. The energy dissipation rate has been normalized by dividing it by the nutation angle squared.

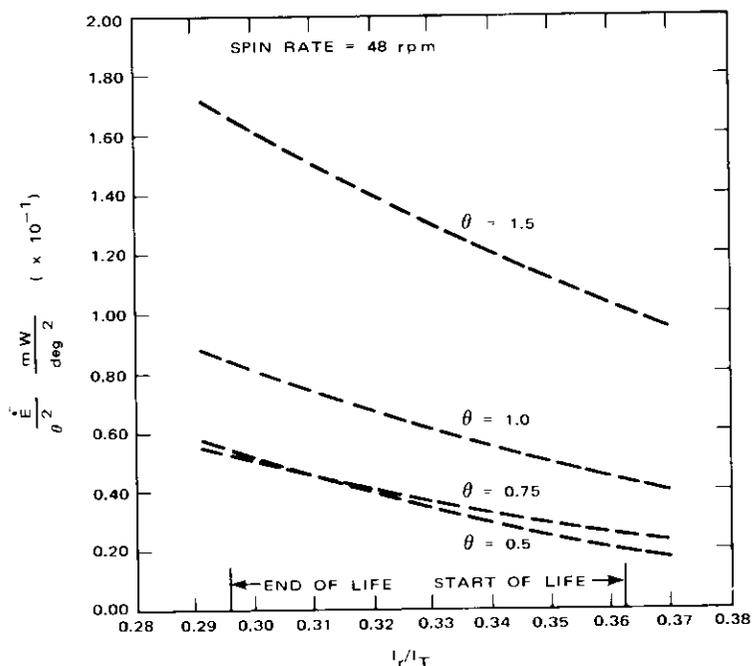


Figure 4. Results of Fuel Slosh \dot{E} on the Linear Motion Fixture

The stability margins calculated for INTELSAT IV using the preceding fuel slosh \dot{E} results were found to be adequate. There was some concern, however, that the linear motion to which the tank was subjected did not provide a sufficiently good simulation of the motion which the tank experiences on the spacecraft.

Because the spacecraft CG in inertial space is stationary, the tank should exhibit arc motion instead of the linear motion to which it was subjected in the tests. The arc motion may be separated into a translation of the center of the tank and a rotation* of the tank relative to its center. A kinematic analysis, developed in spherical polar coordinates (and briefly described in the next section), was used to gain some insight into the tank's motion on the spacecraft. It confirmed the existence of angular oscillatory components of motion. The effect which these oscillatory tank rotations would have on the liquid's behavior was discussed with fluid mechanics consultants and examined through the review of

* The term "rotation" of the tank will continue to be used in this context and should not be confused with spinning of the tank.

pertinent literature [6], [7]. The possibility was raised [8] that the presence of the tank cone and the interaction between the angular oscillation axes and the spin axis may lead to localized fluid perturbations and nonuniform flow with considerably higher \dot{E} .

A test which would subject the tank to oscillations which were more representative of the spacecraft motion was desired.

TEST PROGRAM: ANGULAR MOTION

The first task in designing a new fixture was to establish the motion of the tank on the satellite. Since the results from the kinematic analysis of the spacecraft were used, they are briefly described here.

A reference frame, T , spinning about an inertially fixed axis, \hat{i}_3 (which coincides with the vehicle's total angular momentum vector), at a prescribed constant rate (which, for small nutation angles, is nearly identical to the spacecraft's spin rate) is chosen. In a spherical coordinate system in which the origin is at the spacecraft CG (Figure 5), the motion of the tank center, Q , as seen from the T reference frame, can be expressed in terms of the angular displacements α and β . The α motion was found to be given by

$$\alpha = \alpha_0 - \theta \cos \lambda' t \quad (8)$$

where α_0 is the value of α in the absence of nutation; i.e.,

$$\alpha_0 = \cos^{-1} \frac{b}{r} = \sin^{-1} \frac{a}{r}. \quad (9)$$

No second-order terms exist in α . The equation for β contains a second-order term, but as shown in Section 5.3 of Reference [4], its amplitude is insignificant. For small nutation angles, β may be expressed as

$$\beta = \frac{\pi}{2} - \frac{b}{a} \theta \sin \lambda' t. \quad (10)$$

The dependence of the amplitude on b/a means that there is another angular oscillation about the \hat{i}_2 axis.

The spinning angular motion test fixture which was designed to subject the tank to these motions is shown in Figure 6 and was completed in October 1970. (The launch of the first INTELSAT IV spacecraft was scheduled for December 1970.) The table, spinning at a constant rate, is analogous to the T reference frame; \hat{i}_3 is coincident with the spin axis. A tubular structure supports the tank and is hinged through a pair of bearings. When the tank is in its undisplaced (null) position, the radius vector to Q is of the form $a\hat{i}_2 + b\hat{i}_3$. Thus, α_0 and the same

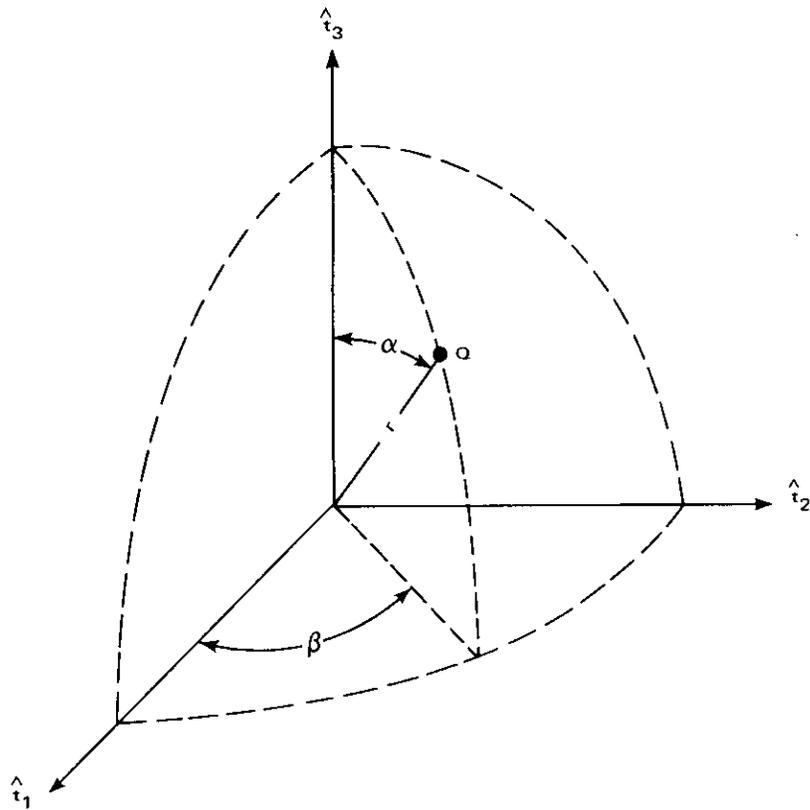


Figure 5. Spherical Coordinate System Used for Kinematic Analysis

geometry are maintained. The α motion is achieved through an onboard motor and an off-center crank arrangement which drives the tank in an arc of amplitude θ and frequency λ' .

If the spin motor is producing a constant torque, the magnitude of the spin speed will exhibit a small sinusoidal variation of frequency λ' because of the moment-of-inertia change of the fixture as the tank describes its arc motion. This small ripple, superimposed on the mean spin rate, indicates that an observer spinning at the mean spin rate (the equivalent of the T reference frame) will see the tank exhibit an angular oscillation of nutation frequency about the spin axis. This β motion is 90° out of phase with the α motion, as on the spacecraft, but its ampli-

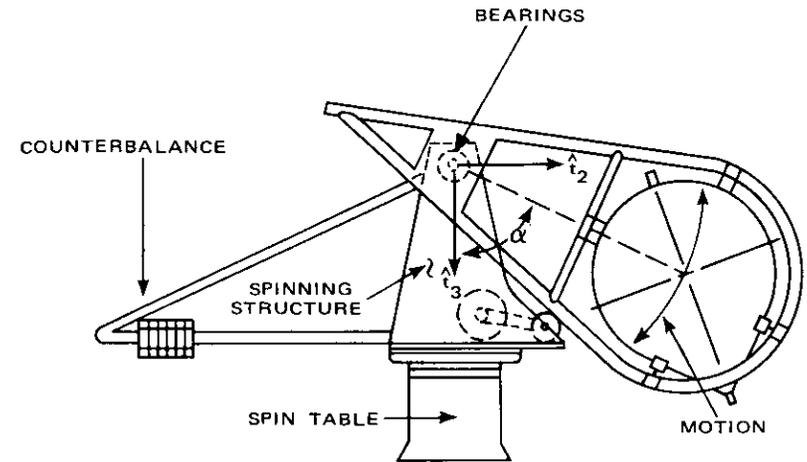


Figure 6. Angular Motion Spinning Test Fixture

the amplitude ranged between the spacecraft amplitude and twice that value. Since the initial objective of this fixture was to simulate the oscillatory motions of the tank on the spacecraft, a dynamic analysis was performed to compare the linearized accelerations to those on the spacecraft. The next section presents the results of this analysis.

Angular Motion Test Fixture Accelerations

The α arc motion to which the tank is subjected by the motor crank arrangement is of the form,

$$\alpha = \alpha_0 - \theta \cos \lambda' t \quad (11)$$

which is the same as that on the spacecraft [equation (8)]. The instantaneous position vector to Q is

$$\vec{r}_Q = (r \sin \alpha) \hat{t}_2 + (r \cos \alpha) \hat{t}_3. \quad (12)$$

The angular velocity of the fixture is

$$\vec{\omega} = (\omega_s + \omega_{sr} \cos \lambda' t) \hat{t}_3 \quad (13)$$

where ω_{sr} is the magnitude of the spin ripple.

The same method used in Appendix B is used to find the acceleration of Q in $\hat{t}_1, \hat{t}_2, \hat{t}_3$ coordinates. A transformation matrix is then used to change the acceleration to the $\hat{r}_1, \hat{r}_2, \hat{r}_3$ coordinate system fixed in the tank. The final result for the linearized accelerations felt by the center of the tank is

$$\begin{aligned}
 g\hat{t}_3 - \bar{a}_0 &= [(2b\theta\omega_s\lambda' - a\omega_{sr}\lambda') \sin \lambda't] \hat{r}_1 \\
 &+ [F_1 \cos \lambda't + a\omega_s^2] \hat{r}_2 \\
 &+ [a\theta\omega_s^2(\sigma - 2) \cos \lambda't + g] \hat{r}_3
 \end{aligned} \quad (14)$$

$$\text{where } F_1 = -\theta[(\lambda'^2 + \omega_s^2)b - g] + 2a\omega_s\omega_{sr}. \quad (15)$$

A comparison of the time-dependent accelerations of this fixture with those experienced by the tank on the spacecraft for a nutation angle of 1° , a spin rate of 60 rpm, an inertia ratio of 0.33, and a fixture spin ripple of 0.3 rpm is presented in Table 2. When the accelerations in this table are compared to the accelerations of the first test fixture (Table 1), it can be seen that the new test fixture provides somewhat better simulation of the accelerations, as well as excellent motion simulation.

TABLE 2. LINEARIZED ACCELERATIONS FOR THE ANGULAR MOTION TEST FIXTURE AND THE SPACECRAFT

Direction	Linearized Acceleration, in./sec ²	
	Angular Motion Spinning Test	Spacecraft
\hat{r}_1	$-7.05 \sin \lambda't$	$0.83 \sin \lambda't$
\hat{r}_2	$5.04 \cos \lambda't$	$-0.83 \cos \lambda't$
\hat{r}_3	$-8.92 \cos \lambda't$	$-8.92 \cos \lambda't$

Test Results

The test fixture was dynamically balanced to operate at 45, 55, and 70 rpm and at liquid weights of 20, 35, and 60 pounds with the magnitude and location of counterweights recorded for all of the nine possible conditions. As the very first test was being run, it immediately became evident that the nature of the liquid motion had drastically changed. Small but highly energetic vortices were apparent on the free surface of the fluid. The localized vortices were so strong that they deformed the otherwise smooth, free surface and created dimples which protruded into the liquid. These vortices were not stationary, but moved around the surface in circles, having more vorticity at some points than at others. Floating particles were introduced in the liquid to better observe the surface activity. The surface particles exhibited energetic oscillations which sometimes had amplitudes as large as the radius of the tank.

Even more surprising was the behavior of the bulk of the fluid. Dye-injection techniques revealed high fluid velocities below the free surface. Dye dropped on the surface was literally sucked below the surface and energetic turbulent mixing spread the dye throughout the liquid, all in an interval of between 2 and 10 seconds.

The technique used in the linear motion fixture for measuring fuel \dot{E} could not be used now because the liquid motion was no longer similar to that of a rigid body. It was decided to make a qualitative evaluation of surface particle motion and dye diffusion in order to examine the dependence of the liquid activity on spin speed, nutation frequency, nutation angle, and amount of fill. The liquid turbulence increased when any of these parameters was increased independently. No step change in liquid behavior as a function of any of these parameters was noticed; although reduced, the turbulence persisted even at very small nutation angles, such as one-tenth of a degree.

As the numerous tests were performed with different dye injection locations and the liquid's motion was recorded on videotape, a clearer understanding of the motion beneath the surface was obtained. From real-time and videotape analysis of the dye dispersion, it appeared that there was a mass flow toward the cone apex along and immediately adjacent to the cone axis. The mass of liquid then flowed away from the cone closer to the walls.

Figure 7 is a representation of a cross-section of the tank showing

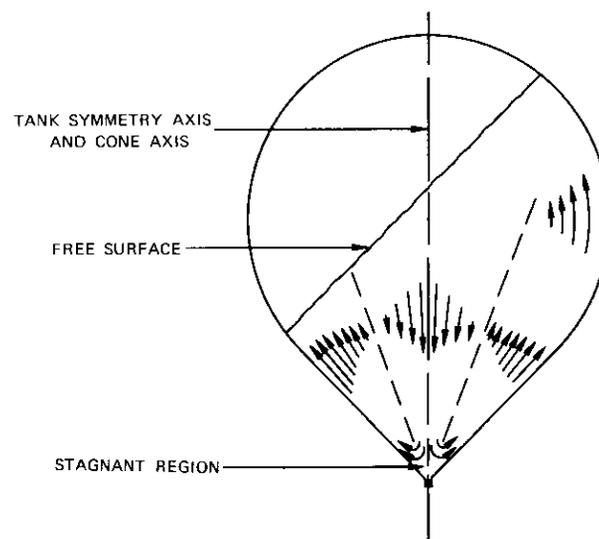


Figure 7. Turbulent Fluid Motion Within the Bulk of the Liquid

the apparent liquid motion. The arrows give an indication of relative fluid velocity. Pressurized injection of dye at the apex of the cone (which is the tank outlet) remained confined to a small volume at the apex instead of diffusing throughout the liquid; this indicated the presence of a stagnant region. When the dye was pressurized even further so that it squirted through the stagnant area into the bulk, it exhibited the motion indicated in Figure 7.

A plexiglass spherical tank of similar size was fabricated and subjected to tests. The liquid motion was calm, exhibiting rigid-body pendulous motion. The shape of the tank therefore appeared to have an important effect on the liquid motion.

The air ball tests conducted by Hughes Aircraft Company, the spacecraft contractor, successfully quantized the turbulent fluid motion energy dissipation in the INTELSAT IV tanks [5]. The measured \dot{E} values were obviously greater than had been predicted by the rigid-body analytical model. (For a spherical tank, they found that the \dot{E} was between 8 and 10 times smaller than for the sphere-cone tank.) To compensate for the larger fuel \dot{E} , the contractor incorporated design modifications on INTELSAT IV to increase the stability margins. The satellite, launched January 25, 1971, has demonstrated passive nutational stability in both transfer and synchronous orbits.

INVESTIGATIONS TO EXPLAIN TURBULENT FLUID MOTION

Following the successful launch of INTELSAT IV, an investigation was started in an attempt to explain the cause of the turbulent fluid motion. The dynamic excitation of the tank and the tank geometry appeared to be equally important in influencing the liquid motion. In order to further investigate the effect of dynamic excitation, attention was directed toward the difference between the linear motion fixture and the angular motion fixture. Although the linearized accelerations acting on the tank were about the same for both tests and the liquid properties were the same, two differences could be cited. The angular motion tests subjected the tank to arc motion and spin ripple, while the other test subjected the tank to rectilinear motion and little, if any, ripple.

In order to investigate the influence of the spin ripple on the fluid behavior, an open-loop control system (called the electronic ripple modulator) was designed and assembled to vary the spin ripple, ω_{sr} , on the angular motion fixture. This was accomplished by using the angular displacement transducer signal (which has a voltage proportional to the α motion) as a phase reference to modulate the voltage to the spin motor. Hence, ω_{sr} was no longer dependent on the fixture's moment-of-inertia change caused by the α motion, but could be set independently.

Tests performed with different values of spin ripple, and with no spin ripple, exhibited little difference in the fluid's motion. Although the spin ripple appeared to have little influence on fluid behavior, the linear motion test fixture was placed back in operation and subjected to tests with large values of spin ripple. As in the earlier tests with that fixture, the fluid remained calm and exhibited rigid-body, pendulous motion. Of the dynamic excitation on the tank, it appeared, therefore, that the arc motion and not the spin ripple was responsible for the fluid turbulence.

The preliminary conclusion was that turbulent fluid behavior is excited inside a spinning sphere-cone tank subjected to oscillatory arc motion, but not in a spherical tank. Further analysis of the results, however, seemed to indicate that such a general conclusion was inaccurate. The liquid had remained calm when the sphere-cone tank on the linear motion test fixture had been subjected to spin ripple, a condition which excited the tank with arc motion about the spin axis. In other words, the linear motion test fixture with spin ripple indeed subjected the tank to an arc motion, the β motion, and yet the fluid exhibited rigid-body calm motion.

Seeking an explanation of this inconsistency, we recognized that the spin axis about which these angular oscillations occurred was parallel to the tank's symmetry axis (Figure 3). This led the author to hypothesize that turbulent fluid motion is excited in a spinning tank only when the tank is subjected to angular oscillations about an axis which is not parallel to its symmetry axis. This hypothesis was not only consistent with all of the experimental results (including those from the sphere tank on the angular motion fixture), but had an even more significant implication. Although it was initially thought that the tank shape and the tank motion independently dictated fluid behavior, the hypothesis, if correct, would couple both effects into a unified and cohesive explanation of how the fluid turbulence is excited.

A test was desired which would convincingly prove or disprove the hypothesis. The angular motion test fixture could not be used because it subjected the tank to rotation about two axes and, therefore, the many variables which affect the liquid's behavior would be difficult to separate. With spin ripple, the linear motion fixture subjected the tank to rotation about the spin axis only. If one rotated the tank by 90° (counterclockwise in Figure 3), so that the cone would point radially out, away from the spin axis, and if the hypothesis were correct, turbulent fluid motion should be observed for the first time on the linear motion test fixture when it had spin ripple. Furthermore, at the limit when the spin ripple is made to disappear (by using the electronic ripple

modulator to offset changes in moment of inertia), the motion should revert to calm rigid-body-type motion.

A series of tests performed on the linear motion test fixture with the tank cone pointed radially out proved the hypothesis to be correct. When the tank exhibited its linear vibrating motion and there was no spin ripple, the fluid was very calm and behaved like a rigid body. As changes were made in the electronic ripple modulator and some ripple was added, the fluid's motion started to change and some rotation of surface particles was evident. As more spin ripple was added, the liquid motion became progressively more turbulent and resembled that observed in the angular motion tests. To obtain further confirmation of the conclusions, the tank was placed in its original position (with the cone axis straight down) and the same tests were repeated. The liquid remained calm and behaved like a rigid body, regardless of the magnitude of the spin ripple.

A possible explanation for this behavior follows. In a nonspinning environment, when one rotates a tank about its symmetry axis, the only forces which the tank exerts on the fluid are shear forces tangential to the moving tank wall. If a tank is rotated about an axis which is not a symmetry axis, the blunt body forces which the tank walls exert on the fluid are trying to change the position and geometry of the fluid. One would expect the fluid's response to these forces to be more energetic, especially in a spinning environment.

SUMMARY AND CONCLUSIONS

A sphere-cone tank and a spherical tank were subjected to tests on two spin fixtures to investigate the behavior of the liquid under simulated spacecraft motions. The linear motion test fixture subjected the spinning tanks to rectilinear vibration, while the angular motion fixture subjected the spinning tanks to angular oscillations.

Turbulent liquid motion was observed in the spinning INTELSAT IV sphere-cone tank when it was subjected to angular oscillations such as those of a nutating spacecraft. Quantization of the fuel energy dissipation rate by Hughes Aircraft Company [5] led to the incorporation of design modifications on INTELSAT IV to increase the nutational stability margins. The satellite, launched January 25, 1971, has exhibited passive nutational stability in transfer and synchronous orbits.

The principal conclusion derived from the tests is that turbulent liquid motion is excited when a spinning tank is subjected to angular oscillations about an axis which is not parallel to the symmetry axis of the tank. Thus, the liquid in a sphere will dissipate less energy than that in a

nonspherical tank; in the Hughes air-ball tests [5] the liquid in the sphere-cone dissipated between 8 and 10 times more energy than the liquid in the sphere.

REFERENCES

- [1] A. J. Iorillo, "Nutation Damping Dynamics of Axisymmetric Rotor Stabilized Satellites," unpublished presentation, *ASME Winter Annual Meeting*, Chicago, Ill., Nov. 1965.
- [2] P. W. Likins, "Attitude Stability Criteria for Dual Spin Spacecraft," *Journal of Spacecraft and Rockets*, Vol. 4, No. 12, December 1967, pp. 1638-1643.
- [3] C. R. Johnson, "TACSAT I Nutation Dynamics," *AIAA 3rd Communications Satellite Systems Conference*, Los Angeles, Calif., April 1970, AIAA Paper No. 70-455.
- [4] E. R. Martin, unpublished paper, April 30, 1971.
- [5] Private communications with J. T. Neer, Hughes Aircraft Company.
- [6] H. P. Greenspan, *The Theory of Rotating Fluids*, London: Cambridge University Press, 1968.
- [7] J. P. Vanyo, "Energy Dissipation in a Liquid Filled Precessing Spherical Cavity and Application to Satellites," Ph.D. dissertation, University of California, Los Angeles, December 1969.
- [8] Private communications with P. G. Saffman, California Institute of Technology and H. P. Greenspan, Massachusetts Institute of Technology.

APPENDIX A

SPACECRAFT DYNAMIC ANALYSIS

Consider the dual-spin vehicle shown in Figure 2. The rotor angular velocity may be written as

$$\bar{\omega} = \omega_1 \hat{r}_1 + \omega_2 \hat{r}_2 + \omega_3 \hat{r}_3 \quad (A1)$$

where \hat{r}_1 , \hat{r}_2 , and \hat{r}_3 are body-fixed unit vectors in the direction shown in Figure 2. Similarly, the platform angular velocity is given by

$$\bar{\omega}^p = \omega_1^p \hat{p}_1 + \omega_2^p \hat{p}_2 + \omega_3^p \hat{p}_3 \quad (A2)$$

The position vector from the origin (spacecraft CG) to the tank center, Q , is

$$\bar{r}_Q = a \hat{r}_2 + b \hat{r}_3 \quad (A3)$$

The linear acceleration of the tank center with respect to an inertial frame is given by

$$\bar{a}_Q = \dot{\bar{\omega}} \times \bar{r}_Q + \bar{\omega} \times (\bar{\omega} \times \bar{r}_Q) \quad (A4)$$

since the tank is in the rotor.

Performing the cross-multiplications and rearranging yields

$$\begin{aligned}\bar{a}_Q &= \hat{r}_1(b\dot{\omega}_2 - a\dot{\omega}_3 + a\omega_2\omega_3 + b\omega_1\omega_3) \\ &+ \hat{r}_2(b\omega_2\omega_3 - b\dot{\omega}_1 - a\omega_3^2 - a\omega_1^2) \\ &+ \hat{r}_3(a\dot{\omega}_1 - b\omega_1^2 - b\omega_2^2 + a\omega_2\omega_3).\end{aligned}\quad (\text{A5})$$

If, at the time $t = 0$, the vector bases, $\hat{p}_1, \hat{p}_2, \hat{p}_3$ and $\hat{r}_1, \hat{r}_2, \hat{r}_3$, are coincident, and if

$$\begin{aligned}\omega_1^P(0) &= \omega_1(0) = 0 \\ \omega_2^P(0) &= \omega_2(0) = \omega_0 \\ \omega_3^P(0) &= 0 \\ \omega_3(0) &= \omega_s\end{aligned}\quad (\text{A6})$$

$$\text{then} \quad \omega_1 = -\omega_0 \sin \lambda' t \quad (\text{A7})$$

$$\omega_2 = \omega_0 \cos \lambda' t \quad (\text{A8})$$

$$\omega_3 = \omega_s \quad (\text{A9})$$

$$\text{where} \quad \omega_0 = \sigma \omega_s \tan \theta \quad (\text{A10})$$

$$\lambda' = (-1 + \sigma)\omega_s$$

and θ is the nutation angle.

From equations (A6) through (A8), the acceleration of Q becomes

$$\begin{aligned}\bar{a}_Q &= \hat{r}_1(-b\omega_0\lambda' \sin \lambda' t - a\omega_0^2 \sin \lambda' t \cos \lambda' t - b\omega_0\omega_s \sin \lambda' t) \\ &+ \hat{r}_2(b\omega_0\omega_s \cos \lambda' t + b\omega_0\lambda' \cos \lambda' t - a\omega_3^2 - a\omega_0^2 \sin^2 \lambda' t) \\ &+ \hat{r}_3(-a\omega_0\lambda' \cos \lambda' t - b\omega_0^2 + a\omega_0\omega_s \cos \lambda' t).\end{aligned}\quad (\text{A11})$$

Linearizing in ω_0 yields

$$\begin{aligned}\bar{a}_Q &= \hat{r}_1[-b\omega_0(\omega_s + \lambda') \sin \lambda' t] \\ &+ \hat{r}_2[b\omega_0(\omega_s + \lambda') \cos \lambda' t - a\omega_s^2] \\ &+ \hat{r}_3[a\omega_0(\omega_s - \lambda') \cos \lambda' t].\end{aligned}\quad (\text{A12})$$

From equation (A10),

$$\omega_s + \lambda' = \sigma \omega_s \quad (\text{A13})$$

$$\omega_s - \lambda' = \omega_s(2 - \sigma). \quad (\text{A14})$$

By substituting these two relationships and the small angle approximation of equation (A9) in equation (A12), we obtain the linearized acceleration equation for Q relative to an inertial frame:

$$\begin{aligned}\bar{a}_Q &= \hat{r}_1(-b\sigma^2\omega_s^2\theta \sin \lambda' t) \\ &+ \hat{r}_2(b\sigma^2\omega_s^2\theta \cos \lambda' t - a\omega_s^2)\end{aligned}$$

APPENDIX B

DYNAMIC ANALYSIS OF LINEAR MOTION FIXTURE

With the tank in its null (undisplaced) position, the original of the body-fixed coordinate system (Figure 3) is arbitrarily chosen at a distance, b , above the center of the tank. The angular velocity of the spinning platform is

$$\bar{\omega} = \omega_s \hat{r}_3. \quad (\text{B1})$$

The motor-cam arrangement drives the tank with an amplitude X and a time function $\cos \lambda' t$. Relative to the body-fixed coordinate system (which is spinning), the instantaneous position vector from O to Q is, therefore,

$$\bar{r}_Q = (a - X \cos \psi \cos \lambda' t) \hat{r}_2 + (b + X \sin \psi \cos \lambda' t) \hat{r}_3. \quad (\text{B2})$$

The acceleration of the center of the tank as seen from an inertial frame in the absence of gravity is

$$\bar{a}_Q = \ddot{\bar{r}}_Q + \dot{\bar{\omega}} \times \bar{r}_Q + 2\bar{\omega} \times \dot{\bar{r}}_Q + \bar{\omega} \times (\bar{\omega} \times \bar{r}_Q). \quad (\text{B3})$$

Substitution of $\bar{r}_Q, \bar{\omega}$ and their derivatives, solution of equation (B3), and linearization of the results yields

$$\begin{aligned}\bar{a}_Q &= [-2X\omega_s \lambda' \cos \psi \sin \lambda' t] \hat{r}_1 \\ &+ [-(\lambda'^2 + \omega_s^2) X \cos \psi \cos \lambda' t - a\omega_s^2] \hat{r}_2 \\ &+ [X\lambda'^2 \sin \psi \cos \lambda' t] \hat{r}_3.\end{aligned}\quad (\text{B4})$$

An observer located at the point Q , instead of on an inertial frame, experiences an equal and opposite acceleration which, when gravity is added, is equal to

$$\begin{aligned}g\hat{r}_3 - \bar{a}_Q &= [2X\omega_s \lambda' \cos \psi \sin \lambda' t] \hat{r}_1 \\ &+ [(\lambda'^2 + \omega_s^2) X \cos \psi \cos \lambda' t + a\omega_s^2] \hat{r}_2 \\ &+ [g - X\lambda'^2 \sin \psi \cos \lambda' t] \hat{r}_3.\end{aligned}\quad (\text{B5})$$

ACKNOWLEDGMENTS

The laboratory work and the mechanical design of the angular motion test fixture were performed by Mr. Robert E. Ballard. Much of the dynamic analysis presented herein is the work of Mr. Thomas C. Patterson. Other contributors to the program were Dr. Brij Agrawal, Dr. Gary Gordon and Mr. George Huson.

I am indebted to Dr. John V. Harrington and Dr. Harold O. Curtis for their suggestions and discussions. Special thanks are due to Mr. John T. Neer of Hughes Aircraft Company for providing access to Hughes data. The cooperation of the COMSAT Project Office for INTELSAT IV is gratefully acknowledged.



Ernesto R. Martin has worked in the areas of spacecraft spin stabilization, mechanical devices, and propulsion since he joined the Spacecraft Laboratory at COMSAT Laboratories, Clarksburg, Md., in 1968. His education includes a B.S. in mechanical engineering with high honors from the University of Florida in 1967 and an M.S. in engineering from the California Institute of Technology in 1968. His present interests include advanced stabilization concepts and electric propulsion.

Index: bandpass filters, communication satellites, microwave filters, radio transponders

NEW TYPES OF WAVEGUIDE BANDPASS FILTERS FOR SATELLITE TRANSPONDERS

A. E. ATIA AND A. E. WILLIAMS

ABSTRACT

Conventional communications satellite transponder designs employ multiplexing systems based upon Chebycheff or Butterworth waveguide bandpass filters. Not only do these filters represent a significant percentage of the overall transponder weight, but the filtering functions do not possess optimum frequency selectivity, group delay, or in-band insertion loss. This paper describes how the filtering functions can be significantly improved by using orthogonal modes in either square or circular waveguide cavities. Filter weight is reduced by at least 50 percent, and optimum filter functions having flat group delay in the passband and amplitude ripple in both the passband and stopband can be realized. Experimental results obtained from the new types of filters are shown to have superior electrical performance and less weight than comparable conventional Chebycheff waveguide designs.

INTRODUCTION

High-capacity communication satellite transponders usually require many transmit channelizing filters to distribute the power over the communication band. In order to utilize the allocated frequency spectrum as efficiently as possible, guard bands should be kept very narrow and, hence, sharp cut-off filters are desirable. Further, the filters must have flat in-band group delay and small gain slope in order to minimize distortion and crosstalk [1]. Conventional direct-coupled waveguide Chebycheff filters of high order followed by group-delay equalizers have been used in the design of multiplexing systems for such transponders. For example, the INTELSAT IV multiplexer consists of twelve 36-MHz bandwidth channels, separated center to center by 40 MHz over the 3.7- to 4.2-GHz band. The frequency selectivity of each channel is achieved by a 10-cavity waveguide Chebycheff filter followed by a 5-pole equalizer [2]. These filters constitute a substantial part of the transponder's weight and volume.

This paper describes the theory and design techniques of new types of waveguide filters which possess nearly optimum electrical characteristics and, in addition, weigh significantly less than comparable conventional waveguide filters. The realization of these filters depends

on the use of degenerate modes in waveguide cavities. This idea was proposed by Ragan [3] in 1948, who described the use of dual modes in single square or circular cavities. Later, in 1951, Lin [4] showed how this could be extended to 3-, 4-, and 5-pole Chebycheff filters in a single cavity of a cylindrical cross section; and, in 1953, Currie [5] described a fifth-order Chebycheff filter in a single spherical cavity.

Although the use of more degenerate modes in a single cavity is an obviously attractive solution to reducing the size and weight of a filter, the control of more than three modes in a single cavity is difficult from a practical point of view. However, higher-order filters can be constructed by cascading either dual- or triple-mode cavities, and it is this concept that is considered in this paper. Such an idea was used by Williams [6] to construct a fourth-order elliptic function filter, in which two circular waveguide cavities, each excited by two orthogonal TE_{111} modes, were coupled together by a cross slot. A further advantage of this approach is the possibility of realizing filtering functions exhibiting ripples in both their passbands and stopbands and offering optimum selectivity [7] consistent with the lowest in-band loss.

In this paper, a synthesis procedure for a general class of transfer function which can be realized by coupling synchronously tuned cavities in an arbitrary fashion is presented. The design procedure for microwave waveguide filters and their physical realization is then described. The realizations discussed are in the form of either circular or square waveguide cavities coupled by small slots in their common walls.

Experimental results of an 8-pole, dual-mode, circular cavity filter described by Atia and Williams [8] are presented, and the design of a sixth-order elliptic filter in two triple-mode, coupled, circular cavities is described in detail.

Finally, an equalized 12-pole, square, dual-cavity design, which provides a substitute for a conventional filter-equalizer combination, is proposed. The three designs are then compared with classical waveguide filters from the point of view of electrical characteristics, weight, and size.

THEORY OF NARROW-BAND COUPLED CAVITIES

General

At microwave frequencies, it is natural to utilize the tuned cavity as one of the basic circuit elements in narrow-band applications. When this is the case, a network will be composed of an interconnection of cavities. The interconnection (or coupling) represents a means by which electromagnetic energy can be exchanged between the different

cavities. The coupling is assumed to be frequency-invariant, which is a valid assumption for narrow bandwidths. In addition to the cavity and the coupling, a third element, the resistance, will also be included.

Since the main purpose is the synthesis of filter networks, only 2-port networks consisting of lossless cavities terminated by resistances at both ports are considered. A general network of this type is shown in Figure 1, where all the cavities are tuned to the same normalized

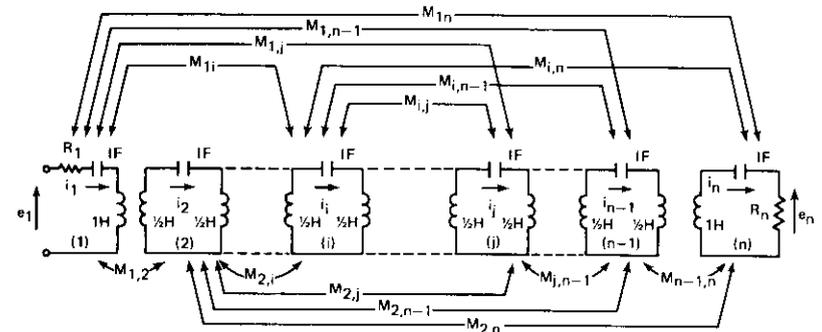


Figure 1. A General 2-Port Network Composed of Synchronously Tuned Coupled Cavities

center frequency, $\omega_0 = 1$ rad/sec. The cavities are numbered 1, 2, ..., n, with the input applied to cavity 1 and the output taken from cavity n. The coupling coefficient between the i th and j th cavities is denoted by M_{ij} . These M_{ij} 's are real (positive, negative, or zero) numbers whose values are independent of frequency.

Two questions are of primary interest. First, what is the class of transfer function that can be realized by the postulated network? Second, if a transfer function of this class is given, how can a network which possesses that transfer function be constructed?

Analysis

Let the loop current in cavity k be I_k ; then the loop equations can be written as

$$\left[R_1 \delta_{1i} + R_n \delta_{ni} + j \left(\omega - \frac{1}{\omega} \right) \right] I_i + j \sum_{\substack{k=1 \\ k \neq i}}^n M_{ik} I_k = e_1 \delta_{1i} \quad (1)$$

$$i = 1, 2, 3, \dots, n$$

where R_1, R_n are the source and load resistances, respectively; δ_{ij} is the Kronecker delta; and e_1 is the input voltage. (Note that the couplings are reciprocal; i.e., $M_{ij} = M_{ji}$.) Equation (1) can be placed in the matrix form,

$$[\lambda I - jR + M]g = -je_1(1, 0, 0, \dots, 0)^t \quad (2)$$

where

$$\lambda = \omega - \frac{1}{\omega}$$

and R is an $n \times n$ matrix whose elements are all zero, except the (1,1) and (n,n) entries which are equal to R_1 and R_n , respectively; M is a real symmetric matrix whose (i,j) entry is M_{ij} , for $i \neq j$, and $M_{ii} = 0$; I is the identity matrix; g is the current vector [$g = (I_1, I_2, \dots, I_n)^t$]; and t represents the transpose.

If the system of equations in (2) is solved for the currents, the transfer voltage ratio can be determined. It is easily seen that this ratio, $I_n R_n / e_1$, is the quotient of two generally complex polynomials in λ . The denominator of the ratio is an n -degree Hurwitz polynomial in λ which is equal to the determinant of the matrix, $[\lambda I - jR + M]$. The numerator polynomial is of order $n - 2$. This defines the class of transfer functions realizable by the general network of Figure 1.

Synthesis

Turning now to the synthesis problem, given the transfer ratio, $I_n R_n / e_1$, and the ratio of the load to the generator impedances, R_1 / R_n , one must calculate the terminal resistances, R_1 and R_n , and the coupling matrix, M , of a network which will produce the given voltage ratio. Since the denominator of the transfer ratio is of the form,

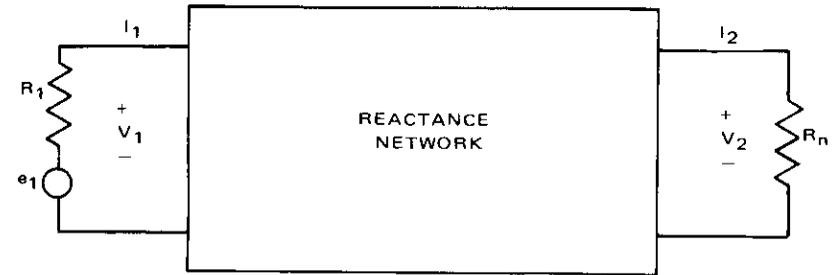
$$d(\lambda) = \sum_{k=0}^n d_k \lambda^{n-k} \quad (3)$$

then, by a well-known algorithm [9] for the computation of the inverse of the matrix $(\lambda I - A)$, one obtains

$$\begin{aligned} d_0 &= 1 \\ d_1 &= -\text{trace}(A) = -\text{trace}(-jR + M) = j(R_1 + R_n) \end{aligned} \quad (4)$$

i.e., the coefficient of λ^{n-1} of the denominator of the transfer ratio can be identified with $j(R_1 + R_n)$, provided that $d(\lambda)$ is a monic polynomial. Since R_1 / R_n is known, the terminal resistances can be calculated. The extraction of R_1 and R_n by this method automatically normalizes the impedance levels of the network.

It remains to compute a coupling matrix, M , from the short-circuit characteristics of the network. According to the procedure outlined by Darlington [7], the network short-circuit admittances, $Y_{11}(\lambda)$ and $Y_{n1}(\lambda)$, can be computed (see Figure 2). However, these admittances



$$I_1 = Y_{11}V_1 + Y_{n1}V_2$$

$$I_2 = Y_{n1}V_1 + Y_{nn}V_2$$

Figure 2. A 2-Port Network and Its Short-Circuit Admittance Relations

are equal to the short-circuit currents I_{s1} and I_{sn} , respectively, obtained by solving equation (2) with $R = 0$ and $e_1 = 1$; i.e.,

$$Y_{11}(\lambda) = -j[(\lambda I + M)^{-1}]_{11} \quad (5a)$$

$$Y_{n1}(\lambda) = -j[(\lambda I + M)^{-1}]_{n1}. \quad (5b)$$

From these equations, it is seen that the denominator of both $Y_{11}(\lambda)$ and $Y_{n1}(\lambda)$ is the characteristic polynomial of the matrix, $-M$. Since $-M$ is a real symmetric matrix, all of its eigenvalues are real, and it is orthogonally similar to a diagonal matrix of the eigenvalues. Thus, there exists an orthogonal matrix T (i.e., $TT^t = T^tT = I$) such that

$$-M = T\Lambda T^t \quad (6)$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ and λ_k ($k = 1, 2, \dots, n$) are the eigenvalues of $-M$. When equation (6) is used in (5), the partial fraction expansion of the short-circuit admittances results; i.e.,

$$Y_{11}(\lambda) = -j \sum_{k=1}^n \frac{T_{1k}^2}{\lambda - \lambda_k} \quad (7a)$$

$$Y_{n1}(\lambda) = -j \sum_{k=1}^n \frac{T_{1k} \cdot T_{nk}}{\lambda - \lambda_k} \quad (7b)$$

Since these admittances are known, it is possible to identify the residues at the poles λ_k ($k = 1, 2, \dots, n$) of $jY_{11}(\lambda)$ and $jY_{nn}(\lambda)$ by T_{ik}^2 and $T_{ik} \cdot T_{nk}$, respectively; hence, the first and last rows of the orthogonal matrix, T , can be determined.

The remaining rows of T can be constructed in a variety of ways. One possible way is to choose any $n - 2$ linearly independent row vectors, which, together with the first and last rows of the T matrix, constitute a basis for the n -dimensional real Euclidian vector space. Applying the Gram-Schmidt [10] orthonormalization process to this set of vectors with the first and last rows of T as the two starting vectors makes it possible to construct an orthonormal basis. If the positions of the first and last rows are preserved, this basis, arranged in any order, forms different possible choices of T . Finally, the M matrix can be constructed from equation (6).

Reduction of Coupling Matrix

It is clear that the M matrix is not unique, since an infinite number of T matrices having the same first and last rows can be constructed. All of these matrices, however, are similar, produce the same prescribed short-circuit admittances, and, hence, the required transfer function. However, the practical realization of a general coupling matrix is usually difficult and, in many cases, it may not be possible to provide certain couplings. Thus, in order to make this theory of practical value, a method of producing zero couplings in specified positions in the M matrix must be developed. One such method, based on Jacobi's diagonalization technique, was developed [11]. In the following, another method to accomplish the reduction of the M matrix will be described. This method depends on applying successive similarity transformations in the form of "plane rotations" to the M matrix. The angle of rotation is determined in such a way as to minimize the sum of the squares of the elements that are affected by the plane rotation and that have to be reduced to zero.

The final form of the M matrix is described by a "topology matrix," P , which is an $n \times n$ matrix of zeroes and ones, depending on whether the corresponding element of M has to be reduced to zero or not. Thus,

$$P_{ij} = 0 \text{ if } M_{ij} \text{ has to be reduced to zero} \\ = 1 \text{ otherwise.}$$

In a step of the reduction process, an orthogonal matrix, O , is formed. All of the elements of O are the same as the identity matrix, except for the four elements, O_{ij} , O_{kk} , O_{ik} , and O_{ki} , which are given by

$$O_{ii} = O_{kk} = \cos \phi; O_{ki} = -O_{ik} = \sin \phi$$

where $i, k = 2, 3, \dots, n - 1$, and ϕ is a properly chosen angle of rotation.

After rotation, the new element values of the M matrix are given by

$$\left. \begin{aligned} M'_{ik} &= M_{ik} \cos \phi + M_{ii} \sin \phi \\ M'_{ii} &= M_{ii} \cos \phi - M_{ik} \sin \phi \end{aligned} \right\} \text{ for } i \neq k, \\ i = 1, 2, \dots, n. \\ M'_{ii} &= M_{ii} \cos^2 \phi + M_{kk} \sin^2 \phi - 2M_{ik} \sin \phi \cos \phi \\ M'_{kk} &= M_{kk} \cos^2 \phi + M_{ii} \sin^2 \phi + 2M_{ik} \sin \phi \cos \phi \\ M'_{ik} &= M'_{ki} = M_{ik}(\cos^2 \phi - \sin^2 \phi) + (M_{ii} - M_{kk}) \sin \phi \cos \phi. \end{aligned} \quad (8)$$

If the function $f_{ik}(\phi)$ is defined as

$$f_{ik}(\phi) = \sum_{\{P_{ii}=0\}} M_{ii}^2 + \sum_{\{P_{kk}=0\}} M_{kk}^2 \quad (9)$$

then it is necessary to choose ϕ such that f_{ik} is minimum. It can be shown that, if $x = \cos^2 \phi$ and $F(x) = f_{ik}(\phi)$, then

$$F(x) = a_1 + a_2x + a_3x^2 \pm (b_1 + b_2x) \cdot \sqrt{x - x^2}. \quad (10)$$

The coefficients (a 's and b 's) of equation (10) are given in terms of the coupling elements in the Appendix.

Differentiating equation (10) with respect to x , and equating the derivative to zero, one obtains the quartic equation,

$$e_4x^4 + e_3x^3 + e_2x^2 + e_1x + e_0 = 0 \quad (11)$$

where the e 's are given in the Appendix. When equation (11) is solved, the root x , which satisfies $0 \leq x \leq 1$ and which gives the minimum value of $F(x)$ of equation (10), is chosen. The new M matrix is then computed and the next step of reduction starts.

The reduction procedure should continue until all elements that have to be reduced to zero become smaller (in absolute value) than a prescribed error. If the topology specified by the P matrix is realizable, then the process will usually converge in a number of iterations proportional to n^2 . This process is easily implemented on a digital computer.

DESIGN OF WAVEGUIDE BANDPASS FILTERS

The previous section described how a transfer function of the form,

$$|t(\lambda)|^2 = \frac{1}{1 + \epsilon^2 [Q^2(\lambda)/P^2(\lambda)]} \quad (12)$$

satisfying Darlington's realizability conditions and those described previously, could be synthesized in the form of terminal resistances and couplings between synchronously tuned cavities. Such a transfer function can be realized in waveguide cavities by the structure shown in Figure 3. This structure incorporates dual orthogonal TE_{101} modes. Transfer functions whose numerator and denominator polynomials, $N(\lambda)$ and $D(\lambda)$, respectively, satisfy the relation,

$$\text{order } (N) \leq \text{integer } \left\{ \frac{\text{order } (D) + 1}{2} \right\}$$

can be realized in the simpler dual-mode circular or square cavity structure shown in Figure 4.

With reference to the coupling geometries, the input and output slots represent the loaded $Q_e (= 1/R)$ of the first and last cavities; the slot dimensions are related to this parameter. Internal cavity couplings are provided by screws so as to couple energy only between the required electric cavities. For example, in the dual-mode case, the screws are placed at 45° to the electric field polarization. Intercavity coupling is provided by slots whose dimensions must be computed to give the required coupling value. The computation of the input-output slots, intercavity slots, and cavity dimensions in terms of the parameters M_{ij} and R , the filter center frequency, and the bandwidth represents the practical design of the orthogonal-mode waveguide filters.

The design of these parameters will be given for the fundamental transverse electric modes only. The designs for higher-order modes and for the transverse magnetic modes follow in a similar manner.

Internal Coupling Slots

From the general cavity structure, three types of couplings are normally present, viz., magnetic and electric field coupling at the end of the cavity and electric field coupling at the center of the cavity. With reference to Matthaei, Young, and Jones [12], it can be readily shown that coupling M_{ij} between the square or cylindrical i th and j th cavities is given by

$$M_{ij} = \frac{q\lambda_0^2}{I_c^3 a^3} \tag{13}$$

- where $q =$ magnetic polarizability (M) of the slot if M_{ij} is the magnetic coupling
- $=$ electric polarizability (P) of the slot if M_{ij} is the electric coupling
- $\lambda_0 =$ free-space wavelength

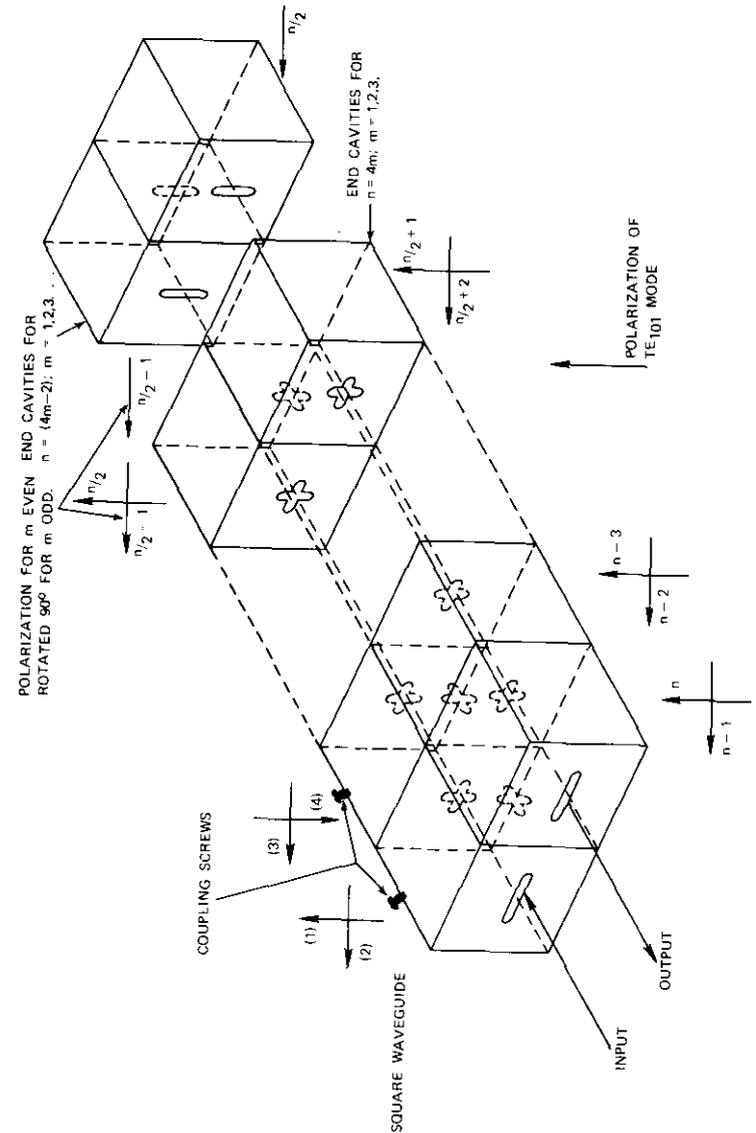


Figure 3. A General Coupled Cavity Filter

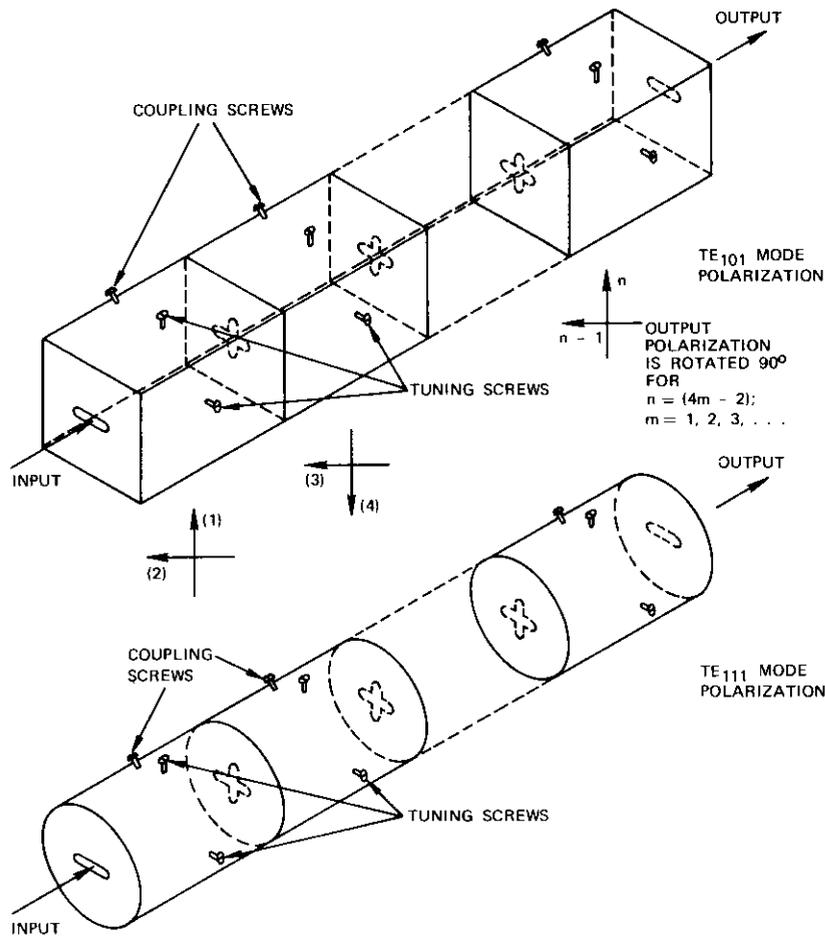


Figure 4. Longitudinal Dual-Mode Cavity Filter

l_c = cavity length
 a = the square waveguide dimension. (For the circular cavity, a^2 is replaced by $3R_0^2$, where R_0 is the cavity radius.)

For electric coupling at the center of a square cavity, the coupling is related to the electric polarizability by

$$M_{ij} = \frac{4P}{a^2} \quad (14)$$

The preceding equations are based on Bethe's theory, which holds for small apertures and zero wall thickness. For circular holes, M equals $d^3/6$ and P equals $d^3/12$, where d is the hole diameter. For slots it is appropriate to use Cohn's experimentally determined polarizabilities [13], [14]. However, when the slots become an appreciable fraction of the freespace wavelength and the aperture thickness becomes large, two corrections must be applied to the polarizabilities [15].

First, for large slot lengths, a correction of the form,

$$1 - \left(\frac{\lambda_l}{\lambda_0}\right)^2$$

where λ_l is the resonant slot length, must be applied. Secondly, if the slot thickness (t_s) is large, then a correction of $10^{-\alpha}$ must be used, where

$$\alpha = \frac{2.73 t_s A}{\lambda_l} \cdot \sqrt{1 - (\lambda_l/\lambda_0)^2} \quad (15)$$

and A is an empirically determined constant. Therefore, the new polarizabilities become

$$q_{new} = q_{old} \frac{10^{-\alpha}}{[1 - (\lambda_l/\lambda_0)^2]} \quad (16)$$

For long thin slots with width to length ratios less than or equal to 0.15, the constant A for magnetic polarizability is approximately equal to 3. For electric polarizability, A lies between 1.0 and 3.0, depending on the shape of the slot.

Input-Output Transformers

The input-output transformers of the filters are designed to realize the loaded Q_c ($=1/R$) of the first and last waveguide cavities. If a rectangular waveguide with inside dimensions a_1 and b_1 is used as the coupling transmission line, the loaded Q_c and magnetic polarizability, M , can be related by [12]

$$M = \frac{a_1 b_1^3 \lambda_0 a^2}{4\pi Q_c \lambda_0^2} \quad (17)$$

where λ_g = the rectangular guide wavelength

a = the square-cavity side dimension

and $3R_0^2$ replaces a^2 for a cylindrical cavity. Aperture size and thickness corrections given by equation (16) must be applied for large couplings and finite wall thickness.

Cavity Design

The dimensions of the cavity are determined by the center frequency of the bandpass filter. At the center frequency, the electrical length of each cavity must be equal to one-half of the guide wavelength for the particular mode under consideration. However, to obtain the physical length, this must be adjusted to allow for perturbations of the resonant frequency by the coupling slots and screws.

For the input/output slots, the normalized reactance (x/z_0) is given by

$$\frac{x}{z_0} = \frac{4\pi M}{a^2 \lambda_{gc}} \quad (18)$$

where a^2 is replaced by $3R_0^2$ for a circular cavity and λ_{gc} is the wavelength at the cavity resonant frequency. Therefore, the cavity length must be reduced by approximately

$$\left(\frac{\lambda_{gc}}{4\pi}\right) \tan^{-1} \left(\frac{2x}{z_0}\right) \quad (19)$$

In addition, a length reduction due to the internal iris and screw couplings must be made. This reduction can be calculated by equating the fractional change in resonant frequency of each cavity to the coupling M_{ij} . The final design length is then taken as the minimum physical length of the orthogonal modes occupying a single cavity. This length is reduced by approximately 2 percent so that cavity tuning screws can be provided to facilitate tuning.

ORTHOGONAL-MODE WAVEGUIDE FILTERS

A synthesis procedure has been described which realizes a class of transfer functions in the form of couplings between synchronously tuned cavities. Further, it has been shown how these parameters may be realized as slots and screws in coupled, orthogonal-mode waveguide cavities. In Reference [8], these concepts have been used in developing an eighth-order, dual-mode waveguide filter whose transfer function has ripple in both passbands and stopbands. The experimental performance of this filter is shown in Figures 5 and 6. This section shows how this theory has been applied in developing a triple-mode, sixth-order elliptic filter and a group-delay equalized filter by using six dual-mode square cavities.

Triple-Mode, Sixth-Order Elliptic Filter

The sixth-order elliptic filter will be described by first outlining the

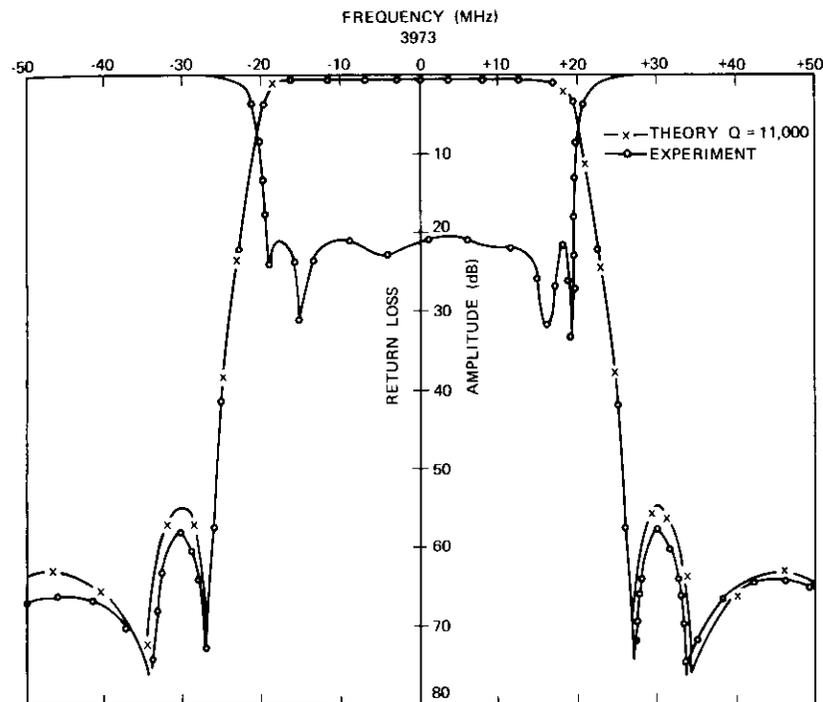


Figure 5. Transmission and Return Loss of 8-pole Filter

synthesis procedure and then the design in waveguide cavities. Let us begin with the transfer function, $t(s)$, of the form

$$|t(s)|^2 = \frac{1}{1 + \epsilon^2 \frac{s^4(s^2 + z_1^2)^2(s^2 + z_2^2)^2}{(s^2 + P_1^2)^2(s^2 + P_2^2)^2}} \quad (20)$$

where $s = j\lambda$.

The following zeros,

$$z_1 = 0.725591 \quad z_2 = 0.971439$$

and poles,

$$P_1 = 1.592692 \quad P_2 = 2.132335$$

have been chosen to produce an elliptic function response having a selectivity of 0.63. For a ripple of 0.05 dB, ϵ is equal to 21.997276.

The synthesis procedure commences by placing $|t(s)|^2$ in the form, $(\lambda/\lambda_c)^2$, and extracting the low-pass prototype transfer function:

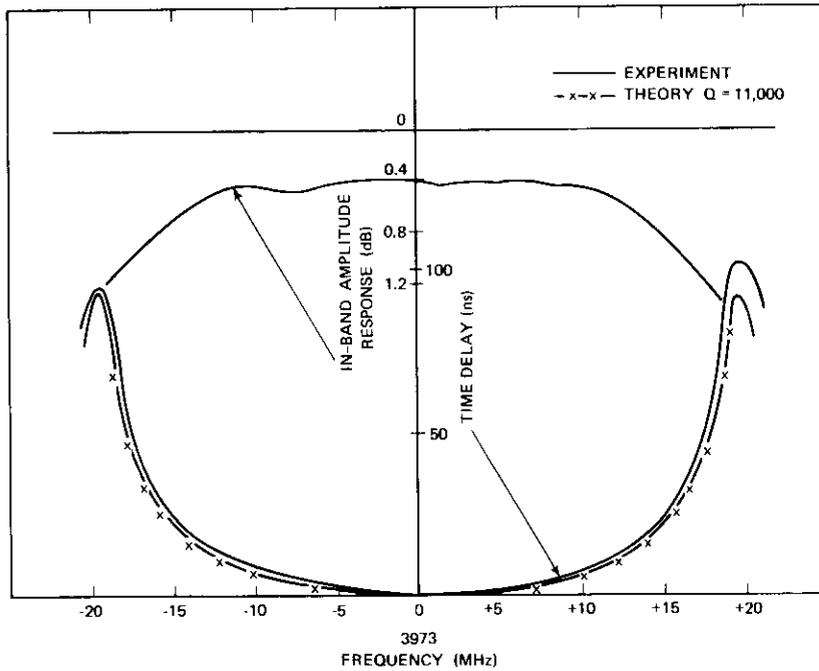


Figure 6.

Inband Transmission and Time-Delay Response of 8-Pole Filter

$$t(s) = \frac{P(s)}{\epsilon Q(s)} \quad (21)$$

$$\begin{aligned} \text{where } P(s) &= s^4 + 7.083520 s^2 + 11.533853 \\ Q(s) &= s^6 + 2.175887 s^5 + 3.837417 s^4 + 4.217645 s^3 \\ &\quad + 3.392813 s^2 + 1.794499 s + 0.524331. \end{aligned}$$

The normalized terminations,* $R_1 = R_6 = R$, are extracted as $R_1 + R_6 = 2.175887$; i.e., $R = 1.087943$.

With R known, the short-circuit input and transfer admittances of the network can be evaluated. The even-mode driving point admittance becomes

* It is convenient to compute R and the coupling matrix M by setting $\Delta\omega$ equal to 1, since the bandwidth, $\Delta\omega$, acts as a scaling factor on these network parameters.

$$Y_e(\lambda) = \frac{C_{e11}}{\lambda - \lambda_{1e}} + \frac{C_{e12}}{\lambda - \lambda_{2e}} + \frac{C_{e13}}{\lambda - \lambda_{3e}} \quad (22)$$

where

$$\begin{aligned} \lambda_{1e} &= -0.415888 & C_{e11} &= (0.68202)^2 \\ \lambda_{2e} &= 1.043595 & C_{e12} &= (0.60117)^2 \\ \lambda_{3e} &= -1.179722 & C_{e13} &= (0.41646)^2. \end{aligned}$$

Since these residues represent the squares of the first row of an orthogonal transformation, then using the Gram-Schmidt orthonormalization process with the vectors $(C_{e11}, C_{e12}, C_{e13})$, $(0, 1, 0)$, $(0, 0, 1)$ as a basis yields the general transformation matrix

$$T_e = \begin{bmatrix} 0.68202 & 0.60117 & 0.41646 \\ -0.51308 & 0.79912 & -0.31330 \\ -0.52114 & 0.0 & 0.85317 \end{bmatrix} \quad (23)$$

and the even-mode coupling matrix, M_e , becomes

$$-M_e = \begin{bmatrix} -0.02080 & 0.80081 & -0.27149 \\ 0.80081 & 0.44115 & 0.20424 \\ -0.27149 & 0.20424 & -0.97227 \end{bmatrix} \quad (24)$$

Realization of this general set of couplings in waveguide cavities would be difficult, since all possible couplings are present. However, if the coupling, M_{e13} , is reduced to zero, i.e., if $M_{11} - M_{36} = 0$, the remaining coupling matrix,

$$-M_e = \begin{bmatrix} -0.02089 & 0.84558 & 0.0 \\ 0.84558 & 0.17124 & 0.59191 \\ 0.0 & 0.59191 & -0.70236 \end{bmatrix} \quad (25)$$

can be realized in a triple-mode, 2-cylindrical-cavity structure shown in Figure 7. The input-output coupling is made from rectangular guide to the circular TM_{010} mode, and screws are used at the end of each cavity to couple M_{12} and M_{56} from this mode to the circular TE_{111} mode. Coupling between the two TE_{111} modes (M_{23} and M_{45}) is provided by screws at 45° to their electric field polarizations. Couplings M_{23} and M_{34} are provided by magnetic coupling via the central iris slot and M_{16} by electric coupling between the two TM_{010} modes via the same iris slot.

The design of a sixth-order elliptic function filter centered at 3.90 GHz for a bandwidth of 40 MHz has been made according to the procedure described previously; the calculated and experimental amplitude responses are shown in Figure 8. From the measured results, it is clear that the control of all six cavities was indeed possible, as is particularly evident from the presence of six poles in the return loss shown in Figure 8. There is, however, some discrepancy between the calculated and measured responses. The first zero of transmission is not present in the measured results, which makes the filter less sharp than it is supposed to be theoretically. This discrepancy is attributed to two factors:

a. There is uncertainty in the thickness correction factor of equation (15) for the electric field coupling (M_{10}) because the exact value of A is unknown.

b. The effect of the cylinder curvature on the input and output coupling slots (R slots) is not known. The calculations of the slots were based on Bethe's theory, assuming a flat surface.

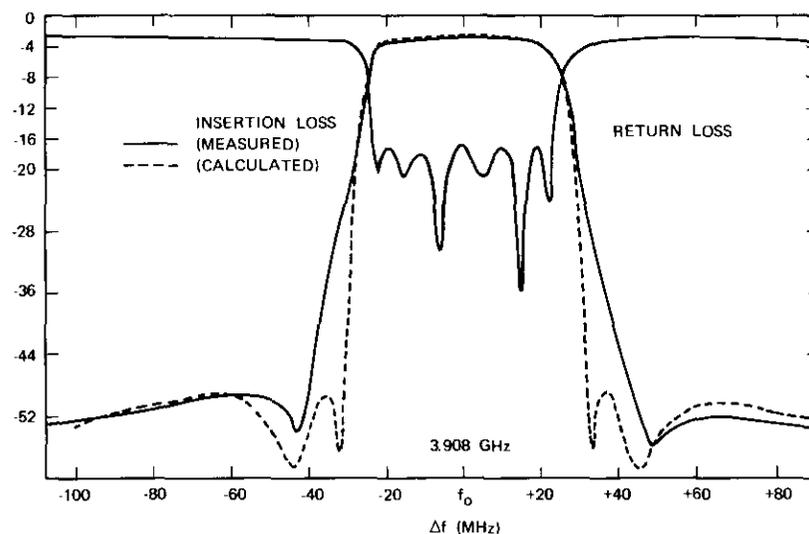
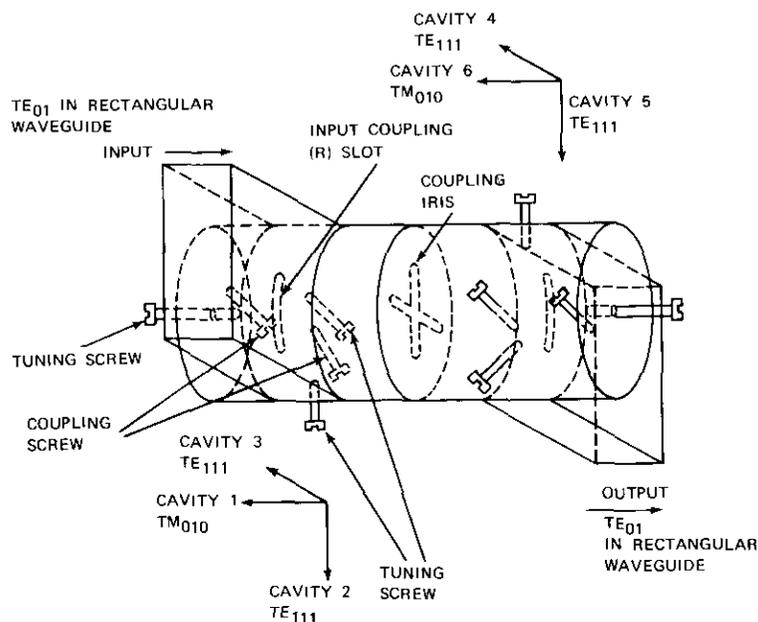


Figure 8. Amplitude and Return Loss Response of Triple-Mode Filter

Twelve-Pole Equalized Filter

Both the sixth- and eighth-order filters described previously are of the minimum phase type and, therefore, their attenuation and phase characteristics are related by the Hilbert transform. However, when filters are used for communication purposes, it is desirable that the phase be as linear as possible over the usable filter passband. This can be accomplished by either cascading the minimum phase filters by all-pass networks which provide the necessary phase (and, hence, group-delay) equalization, or by synthesizing a nonminimum phase filter function which possesses the required attenuation as well as group-delay specifications.

The latter approach has the advantages of reducing the overall number of filter sections (cavities) and, hence, the overall filter weight. Qualitatively, this approach utilizes, in a near-optimum way, all of the cavities in the filter. It is important, however, that the nonminimum phase filter function meeting the required specifications be calculated in such a way as to have the minimum number of poles. This involves the optimum solution of the approximation problem, which, in general, has not been found. There are, however, many numerical methods for the determination of these functions [16]. The details of the approximation problem will not be included here; instead, a realization of a non-



minimum phase transfer function, which is known to be near-optimum in amplitude as well as phase, will be illustrated. This transfer function has the following form [17]:

$$t(s) = \frac{N(s)}{D(s)} \quad (26)$$

where

$$D(s) = s^{12} + 2.60284 s^{11} + 6.25145 s^{10} + 9.48377 s^9 + 12.28017 s^8 \\ + 12.19458 s^7 + 10.08583 s^6 + 6.60655 s^5 + 3.46310 s^4 \\ + 1.38002 s^3 + 0.39859 s^2 + 0.07391 s + 0.00665$$

and

$$N(s) = 0.021974270 s^{10} + 0.065953794 s^8 + 0.037472439 s^6 \\ - 0.005890728 s^4 + 0.013240601 s^2 - 0.006641524.$$

The computed amplitude and group-delay characteristics of this transfer function are shown in Figure 9.

The coupled-cavity synthesis technique discussed previously can be applied directly to the transfer function. This process gives the following results for the normalized terminal resistances ($R_1 = R_n = 1.301420$) and even-mode coupling matrix of a symmetrical structure:

$$M_e = \begin{bmatrix} -0.00844 & -0.46602 & 0.00000 & -0.83480 & 0.00000 & 0.00000 \\ -0.46602 & 0.39292 & 0.17126 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 0.17126 & -0.50469 & 0.34782 & 0.00000 & 0.10450 \\ -0.83480 & 0.00000 & 0.34782 & -0.11915 & -0.50577 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & -0.50577 & 0.64748 & -0.28696 \\ 0.00000 & 0.00000 & 0.10450 & 0.00000 & -0.28696 & -0.94250 \end{bmatrix}$$

The filter can be realized in the form of six square cavities with coupling slots and screws which provide the appropriate values of the terminal resistances and coupling matrix given in the preceding equation. The configuration of this filter would be as shown in Figure 3 with $m = 3$. Dimensions of the slots and cavities are calculated according to equations (13) through (19). Realization of this filter would replace the filter-equalizer combination of INTELSAT IV and present a considerable reduction in weight and size over the conventional design, in addition to having lower loss and superior electrical characteristics.

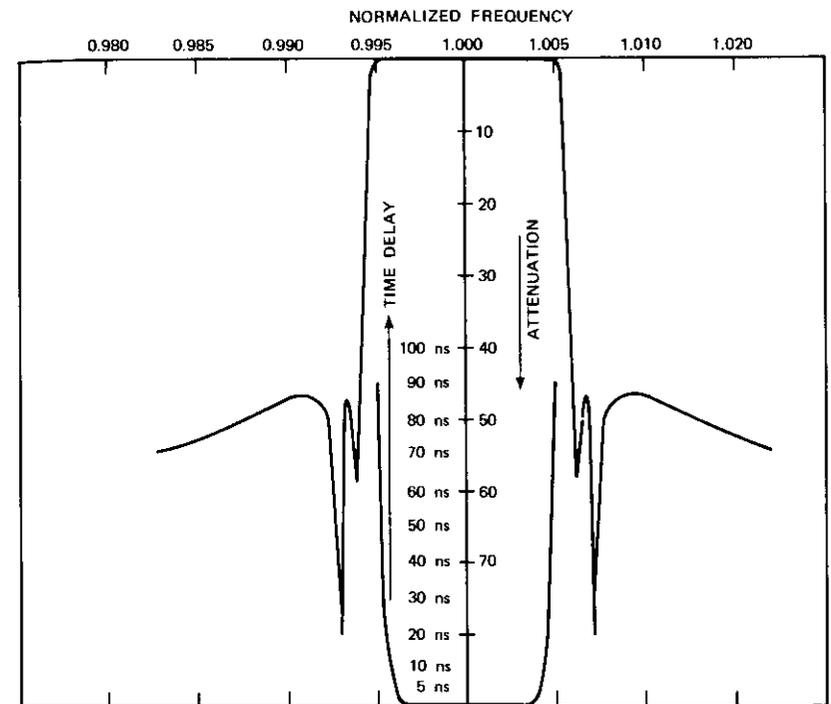


Figure 9. Theoretical Response of Equalized 12-Pole Filter

Comparison of Orthogonal-Mode Filter and Conventional Waveguide Filter

As stated earlier, in-band loss, gain slope, size, and weight for a given frequency selectivity are the important properties of filters having possible space applications. It is therefore appropriate to outline the significant parameter improvements which may be gained by employing orthogonal-mode filters rather than conventional cascaded waveguide filters.

For conventional designs, the INTELSAT IV 10-cavity Chebycheff filter followed by a 5-section time-delay equalizer is considered as the reference point. When fabricated from Invar, this filter weighs approximately 5 pounds and the equalizer 3 pounds. For the same frequency selectivity, the dual-mode 8-pole filter described in Reference [8] has been fabricated from Invar with a weight of approximately 1 pound and the triple-mode filter could be fabricated with a weight

less than 10 ounces. Not only does this represent a considerable saving in weight (and, hence, space), but the in-band filter loss (and, more importantly, the gain slope) is reduced from 1.5 dB to 0.5 dB.

The only sacrifice that has been made is a substitution of the monotonic increase of attenuation in the stopband of the Chebycheff filter by a minimum attenuation of approximately 50 dB for the 8-pole filter and approximately 40 dB for the 6-pole filter. Nevertheless, both filters would have to be equalized using conventional techniques, i.e., by adding to the filter a group-delay equalizer. This would reduce the percentage weight gained for a given transponder application.

On the other hand, if the filter-equalizer design employs a technique whereby an equalized filter transfer function is directly synthesized in the form of coupled cavities, then the full potential weight and space savings can be achieved. For example, a filter transfer function that has the same selectivity and group-delay characteristics as the INTELSAT IV filter-equalizer combination can be realized in six square-coupled cavities. Such an integrated filter-equalizer could be fabricated from Invar with a weight of less than two pounds. Further, for a Q of 10,000, the loss would be about 0.9 dB compared to 2.5 dB for the INTELSAT IV filter-equalizer combination. A significant improvement in in-band gain slope would therefore be realized. The weight reductions and improvement in electrical performance are shown in Table 1.

CONCLUSIONS

New types of bandpass filters for communication satellite transponders have been shown to possess characteristics superior to those of conventional designs. The use of orthogonal modes in waveguide cavi-

TABLE 1. WEIGHT COMPARISON BETWEEN CONVENTIONAL AND ORTHOGONAL-MODE FILTERS

Filter	Weight	In-Band Loss
10-Pole Conventional INTELSAT IV Chebycheff	5 lb.	1.5 dB
8-Pole Dual Orthogonal Mode	1 lb. 2 oz.	0.4 dB
6-Pole Triple Orthogonal Mode	10 oz.	0.35 dB
10-Pole Conventional INTELSAT IV Chebycheff Followed by 5-Pole Equalizer	8 lb.	2.5 dB
12-Pole Dual Orthogonal Mode	1 lb. 12 oz.	0.9 dB

ties permits the new filter designs to have at least 70 percent less weight than conventional ones. The in-band loss of the new filters is less than half of the loss of the corresponding conventional design of the same center frequency, bandwidth, and selectivity.

The theory and method of design of these new types of filters have been presented. Good agreement between the theory and experimental results demonstrates the practicality of using these new filters in future satellite transponders.

REFERENCES

- [1] A. L. Berman and C. E. Mahle, "Nonlinear Phase Shift in Traveling-wave Tubes as Applied to Multiple Access Communications Satellites," *IEEE Transactions on Communication Technology*, COM-18, No. 1, Feb. 1970, pp. 37-48.
- [2] T. A. Abele and H. Wang, "An Adjustable Narrow Band Microwave Delay Equalizer," *IEEE Transactions on Microwave Theory and Techniques*, MTT-15, No. 10, Oct. 1967, pp. 566-574.
- [3] G. L. Ragan, *Microwave Transmission Circuits* (MIT Radiation Laboratory Series, Vol. 9), New York: McGraw-Hill, 1948, pp. 673-677.
- [4] W. Lin, "Microwave Filters Employing a Single Cavity Excited in More Than One Mode," *Journal of Applied Physics*, Vol. 22, No. 8, August 1951, pp. 989-1001.
- [5] M. R. Currie, "The Utilization of Degenerate Modes in a Spherical Cavity," *Journal of Applied Physics*, Vol. 24, No. 8, Aug. 1953, pp. 998-1003.
- [6] A. E. Williams, "A Four-cavity Elliptic Waveguide Filter," *IEEE Transactions on Microwave Theory and Techniques*, MTT-18, No. 12, Dec. 1970, pp. 1109-1114.
- [7] S. Darlington, "Synthesis of Reactance 4-poles Which Produce Prescribed Insertion Loss Characteristics," *Journal of Mathematics and Physics*, Vol. 18, No. 4, Sept. 1939, pp. 257-353.
- [8] A. E. Atia and A. E. Williams, "Narrow Bandpass Waveguide Filters," accepted for publication, *IEEE Transactions on Microwave Theory and Techniques*.
- [9] J. A. Frame, "Matrix Functions and Applications, Part IV: Matrix Functions and Constituent Matrices," *IEEE Spectrum*, Vol. 1, No. 6, June 1964, pp. 123-131.
- [10] H. L. Hamburger and M. E. Grimshaw, *Linear Transformations in n-dimensional Space*, London: Cambridge University Press, 1951.
- [11] A. E. Atia and A. E. Williams, unpublished paper, Sept. 22, 1970.
- [12] G. L. Matthaei, L. Young and E. M. T. Jones, *Microwave Filters, Impedance-matching Networks, and Coupling Structures*, New York: McGraw-Hill, 1964.
- [13] S. R. Cohn, "Determination of Aperture Parameters by Electrolytic-

tank Measurements," *Proceedings of the IRE* (U.S.A.), Vol. 39, No. 11, Nov. 1951, pp. 1416-1421.

- [14] S. B. Cohn, "The Electric Polarizability of Apertures of Arbitrary Shapes," *Proceedings of the IRE* (U.S.A.), Vol. 40, No. 9, Sept. 1952, pp. 1069-1071.
- [15] S. B. Cohn, "Microwave Coupling by Large Apertures," *Proceedings of the IRE* (U.S.A.), Vol. 40, No. 6, June 1952, pp. 696-699.
- [16] J. G. Linvill, "The Approximation with Rational Functions of Prescribed Magnitude and Phase Characteristics," *Proceedings of the IRE* (U.S.A.), Vol. 40, No. 6, June 1952, pp. 711-721.
- [17] K. Wittmann, G. Pfitzenmaier and F. Künemund, "Dimensionierung reflexionsfaktor—und laufzeitgebener versteilter Filter mit Überbrückungen," *Frequenz*, Band 24, Nr. 10, Oktober 1970, pp. 307-312.

APPENDIX

The coefficients a and b of equation (10) are given by

$$\begin{aligned} a_1 &= c_2 + d_2 & b_1 &= c_3 + d_5 \\ a_2 &= c_1 - c_2 - 2d_2 + d_3 & b_2 &= d_4 - d_5 \\ a_3 &= d_1 + d_2 - d_3 \end{aligned}$$

where

$$\begin{aligned} c_1 &= 2 \sum_{\{l|P_{lk}=0, l \neq i, k\}} M_{lk}^2 + 2 \sum_{\{l|P_{li}=0, l \neq i, k\}} M_{li}^2 \\ c_2 &= 2 \sum_{\{l|P_{lk}=0, l \neq i, k\}} M_{li}^2 + 2 \sum_{\{l|P_{li}=0, l \neq i, k\}} M_{lk}^2 \\ c_3 &= 4 \sum_{\{l|P_{lk}=0, l \neq i, k\}} M_{lk}M_{li} - 4 \sum_{\{l|P_{li}=0, l \neq i, k\}} M_{lk}M_{li} \\ d_1 &= M_{ii}^2 \delta_{o, P_{ii}} + M_{kk}^2 \delta_{o, P_{kk}} + 2M_{ik} \delta_{o, P_{ik}} \\ d_2 &= M_{ii} \delta_{o, P_{ii}} + M_{kk} \delta_{o, P_{kk}} + 2M_{ik} \delta_{o, P_{ik}} \\ d_3 &= (2M_{ii}M_{kk} + 4M_{ik}^2) \delta_{o, P_{ii}} + (2M_{ii}M_{kk} + 4M_{ik}^2) \delta_{o, P_{kk}} \\ &\quad + 2[(M_{ii} - M_{kk})^2 - 2M_{ik}^2] \delta_{o, P_{ik}} \\ d_4 &= -4M_{ii}M_{ik} \delta_{o, P_{ii}} + 4M_{kk}M_{ik} \delta_{o, P_{kk}} + 4M_{ik}(M_{ii} - M_{kk}) \delta_{o, P_{ik}} \\ d_5 &= -4M_{kk}M_{ik} \delta_{o, P_{ii}} + 4M_{ii}M_{ik} \delta_{o, P_{kk}} - 4M_{ik}(M_{ii} - M_{kk}) \delta_{o, P_{ik}} \end{aligned}$$

The coefficients e of equation (11) are given by

$$\begin{aligned} e_0 &= b_1^2 \\ e_1 &= 6b_1b_2 - 4b_1^3 - 4a_2^2 \\ e_2 &= 4a_2^2 + ab_2^2 - 20b_1b_2 - 16a_2a_3 + 4b_1^2 \\ e_3 &= -8(2a_3^2 + 3b_2^2 - 2a_2a_3 - 2b_1b_2) \\ e_4 &= 16(a_3^2 + b_2^2). \end{aligned}$$

ACKNOWLEDGMENT

This paper is based upon work performed under the joint sponsorship of the International Telecommunications Satellite Consortium (INTELSAT) and the Communications Satellite Corporation (COMSAT). Any views expressed in this paper are not necessarily those of INTELSAT.

Ali E. Atia was born in Cairo. He received the B.S. degree with honors from Ain Shams University, Cairo, and the M.S. and Ph.D. degrees from the University of California, Berkeley, all in electrical engineering. From 1962 to 1964 he was a lecturer in the Department of Electrical Engineering, Ain Shams University. From 1965 to 1968 he was a research assistant at the University of California, where he was engaged in research on frequency-independent antennas. From 1968 to 1969 he was a Teaching Fellow and Assistant Professor in the Department of Electrical Engineering and Computer Sciences, University of California. At present he is a member of the technical staff at COMSAT Laboratories, Clarksburg, Maryland.



Albert E. Williams received the B.E. degree in electrical engineering from the University of Western Australia, Nedlands, in 1962 and the Ph.D. from University College, London, in 1966. In 1966 he was a joint recipient of the IEE (London) Sylvanus P. Thompson Premium award. From 1966 to 1968 he was a lecturer in the Department of Electrical Engineering, University of Western Australia. He is currently a member of the technical staff, COMSAT Laboratories, Clarksburg, Maryland, where he is actively engaged in the development of mixers and lightweight filters for satellite transponders.



to assume that the radio-relay traffic consists mostly of FDM telephone channels and that the modulation is low-index FM. Furthermore, to retain flexibility in the use of the bands, independent frequency plans are considered for both services. Thus the requirement is to protect the radio-relay channel at all baseband frequencies against the worst locations for any modulation of the satellite's communication carrier. Conversely, the satellite communications system needs to be protected against the large residual carriers resulting from the low-index frequency modulation of the radio-relay system. These considerations have led to the specification of a limit on the power spectral density emanated from the satellite system [2] and to a direct power limitation on the terrestrial radio-relay transmitter [3].

In the case of frequency sharing between a satellite system and a terrestrial system, there are four undesired signal paths, as shown in Figure 1; the two paths involving the satellite will be considered here. In principle, a determination of sharing feasibility consists of calculating the interference resulting from different typical system configurations and propagation conditions, and establishing the resulting system constraints for tolerable interference.

The purpose of this paper is to elucidate the technical reasons behind the limitations placed on the emissions from both the radio-relay and satellite transmitters in the presently shared 4- and 6-GHz frequency bands. Such limitations are also investigated for other frequency bands that may be shared in the future.

Geostationary satellite interference into terrestrial radio-relay receivers is first discussed. The concept of a power flux-density at the surface of the earth is introduced and a rationale for its specification in a narrow band of frequencies is presented. The "interference transfer factor," or the system's sensitivity to interference, is defined and is shown to be dependent on both the frequency of operation and the elevation angle of the interfering satellite. A suggested permissible dependence of power flux-density on elevation angle and frequency is derived.

Next, the interference caused by terrestrial radio-relay transmitters into a satellite receiver is investigated. It is shown that both the terrestrial transmitter power and e.i.r.p. (effective isotropic radiated power) need to be limited. The frequency dependence of these limits is presented. Finally, a summary and conclusions are given.

SATELLITE INTERFERENCE INTO TERRESTRIAL RADIO SYSTEMS IN SHARED FREQUENCY BANDS

The CCIR has recommended [4] that, in any telephone channel of a 2500-km hypothetical reference circuit (hrc) [5] for radio-delay

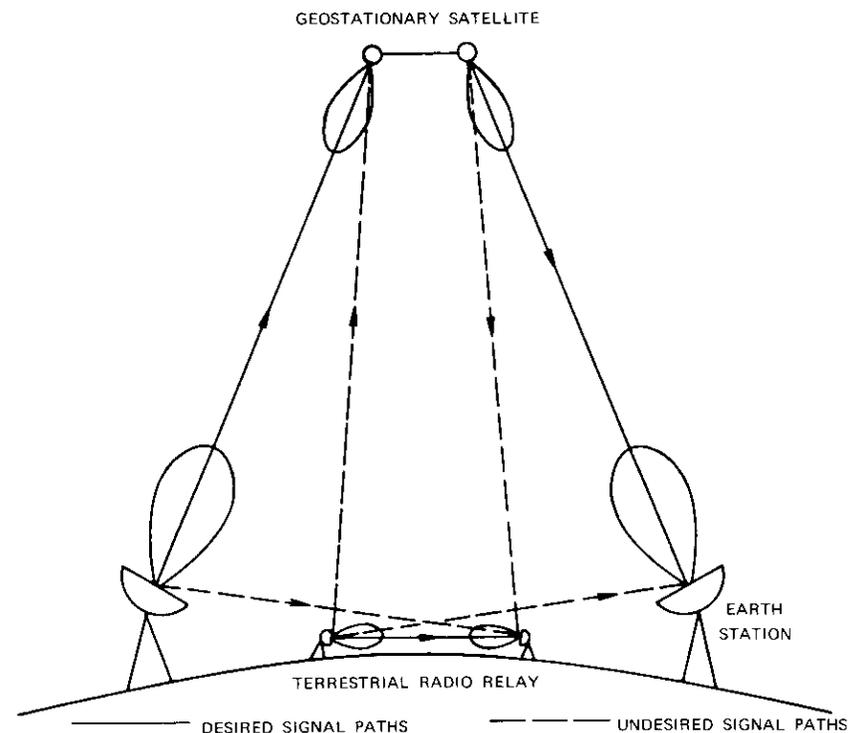


Figure 1. Interference Geometry Between the Satellite System and the Terrestrial System

systems, the interference noise power caused by the aggregate of emissions from communications satellite systems should not exceed

- 1000-pW0p mean value in any hour,
- 1000-pW0p 1-minute mean power for more than 20 percent of any month, and
- 50,000-pW0p 1-minute mean power for more than 0.01 percent of any month.

Any limitation of satellite emissions proposed to protect the radio-relay systems should be consistent with this recommendation.

Interference Due to Geostationary Satellites

At terrestrial radio-relay systems, geostationary satellites, unlike moving satellites, produce an interference which is invariant, except when changes due to the uncorrelated fading of the radio-relay and satellite

signals occur. The limitations on satellite emissions should therefore be based on the lower, long-term, permissible interference power level, i.e., 1000 pW0p.

The hrc for a radio-relay system comprises 50 stations; hence, the interference introduced by all geostationary satellites into any radio-relay station should be, at worst, 1000 pW0p, and the average interference should be less than 20 pW0p.

Traditionally, the limitations of satellite emissions have referred not to the e.i.r.p. of the satellites, but to the power flux-density in a unit of bandwidth produced at the surface of the earth. This approach eliminates consideration of the modulation techniques used by the satellite and allows the satellite system designer more flexibility in frequency planning, antenna coverage, etc.

The exact baseband interference into an angle-modulated carrier is determined by the convolution of the two RF spectra. Since terrestrial radio relays using FDM/FM typically use a low modulation index, the resultant interference at baseband is primarily determined by the power spectrum of the interference. Any interfering carrier spike will therefore cause interference in just one channel.

A worst-case interference calculation consists of equating the interference-to-thermal noise ratio at baseband to the ratio of RF interference power in the worst 4-kHz slot-to-thermal noise power. This is equivalent to representing the interfering signal by a band of white noise with a power density equal to the highest power density in any 4-kHz slot of the actual interfering signal. For terrestrial radio systems employing digital modulation techniques, such an interference calculation technique is overly conservative, since, for these systems, it is the total power in the receiver bandwidth, and only to a lesser extent the distribution in the band, that determines the performance degradation.

Interference into a Radio-Relay Station from One Geostationary Satellite

The baseband interference noise power caused by an interfering satellite transmission into a particular radio-relay receiving station can be expressed as

$$N_i = N_{th} \cdot \frac{F(\theta) G_r(\alpha)}{S_i} \quad (1)$$

where

$$S_i = \frac{4\pi k T b}{\lambda^2} \quad (2)$$

The isotropic interference transfer factor, S_i , is particular to the radio-relay system and is usually frequency dependent. It is evident that the

actual degradation produced at that receiver will depend critically on α , the angle between the main beam of the radio-relay antenna and the location of the satellite as seen from the station. Another factor, S_m , the on-beam transfer factor, is defined by

$$S_m = \frac{S_i}{G_r(0)} \quad (3)$$

Terrestrial radio-relay systems must point their antennas close to the horizon and are therefore quite insensitive to interfering signals arriving at high elevation angles. The worst interference condition exists when a radio-relay antenna points toward the geostationary orbit and a geostationary satellite is located just above the horizon at that azimuth (critical azimuth). The resulting interference is then directly proportional to the power flux-density at the particular (low) elevation angle caused by the interfering satellite. This particular interference situation can often be avoided by designing the route of the terrestrial radio system in such a way as to avoid pointing near the geostationary satellite orbit.

The geostationary satellite orbit presents some unique geometrical properties. It is visible to all points on the surface of the earth except the polar caps. For a given location on the earth, the geostationary satellite orbit can be thought of as a curved line in the sky touching the local horizon at two different azimuths, as indicated in Figure 2. The maximum elevation of the geostationary satellite orbit as well as the two critical azimuths are a function of latitude only.

Figures 3 and 4 show the geometrical relation between a geostationary satellite and a terrestrial radio-relay receiving station. The azimuth, γ_s , and the elevation angle, θ , of the direction to a geostationary satellite are expressed by

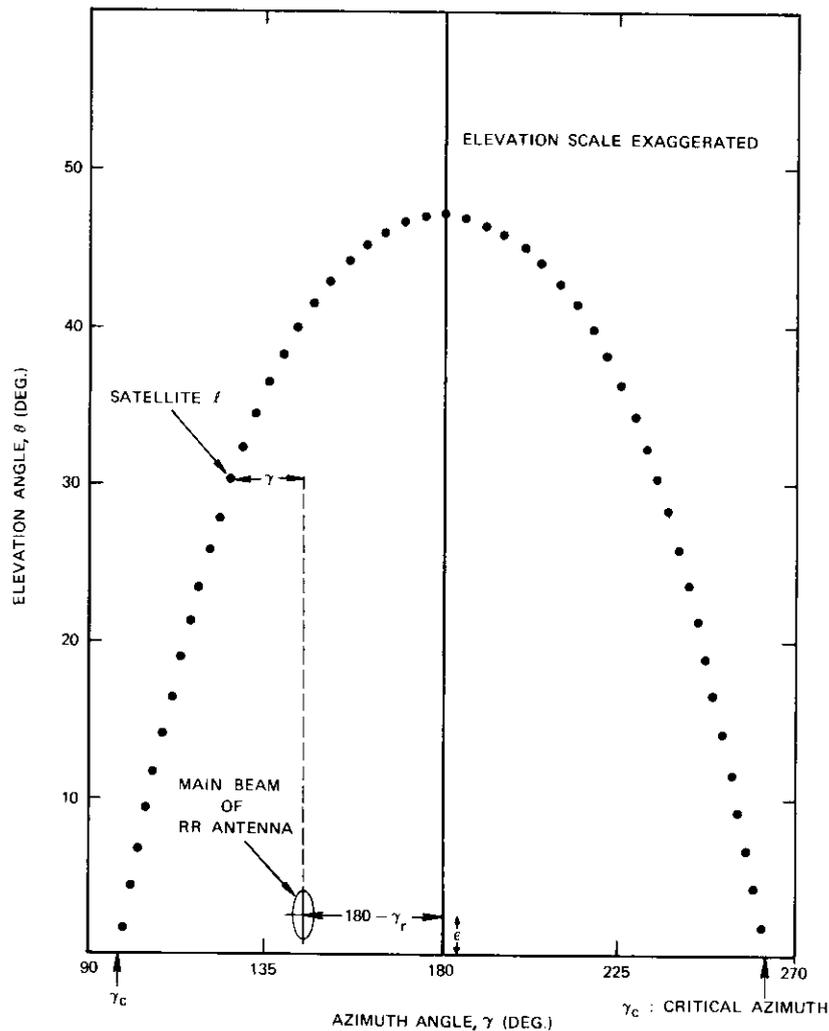
$$\gamma_s = \arctan \left\{ \pm \frac{\tan \delta}{\sin \phi} \right\} \quad (4)$$

$$\theta = \arcsin \left(\frac{\cos \phi \cos \delta - r}{\sqrt{r^2 + 1 - 2r \cos \phi \cos \delta}} \right) \quad (5)$$

The longitude difference for which a geostationary satellite appears on the horizon is obtained by letting θ equal zero in equation (5); i.e.,

$$\delta_c = \arccos \left(\frac{r}{\cos \phi} \right) \quad (6)$$

Introducing δ_c , obtained from equation (6), into equation (4) results in the "critical" azimuths, γ_c , as a function of the latitude ϕ of the



NOTE: DOTS REPRESENT GEOSTATIONARY SATELLITES SPACED 3° APART

Figure 2. The Geostationary Satellite Orbit as Seen from the Surface of the Earth (latitude $\approx 37^\circ$)

receiving stations, as shown in Figure 5. Also shown in Figure 5 is the band of azimuths corresponding to the directions in which the angle between the main beam and the geostationary satellite orbit is less than 2° .

The discrimination angle, α , between the main beam direction of the radio-relay antenna and the direction of a geostationary satellite is a function of the satellite elevation angle, the orientation of the antenna with respect to the subsatellite point, and the elevation angle, ϵ , of the radio-relay antenna. If γ is the difference between the azimuth toward the satellite and the azimuth of the main beam of the radio-relay antenna, the discrimination angle, α , can be shown to be

$$\alpha = \arccos [\cos \theta \cos \gamma \cos \epsilon + \sin \theta \sin \epsilon]. \quad (7)$$

The preceding relationships permit a calculation of the interference noise once the radio-relay antenna pattern and the flux-density variation of satellite emissions as a function of elevation angle are known. The following assumptions were made to derive the interference curve plotted in Figure 6. The radio-relay beam was assumed to be

$$\begin{aligned} 10 \log G_r(\alpha) &= 40 \text{ dB} & 0^\circ < \alpha < 0.575^\circ \\ &= 34 - 25 \log \alpha \text{ dB} & 0.575^\circ < \alpha < 57.5^\circ \\ &= -10 \text{ dB} & 57.5^\circ < \alpha < 180^\circ \end{aligned} \quad (8)$$

The flux-density was assumed to be

$$F(\theta) = -152 + \frac{\theta}{15} \text{ dB } (W/m^2 \cdot 4 \text{ kHz}). \quad (9)$$

Finally, the radio-relay interference transfer factors were taken as

$$\begin{aligned} 10 \log S_i &= -124.3 \text{ dB} \\ 10 \log S_m &= -164.3 \text{ dB} \end{aligned} \quad (10)$$

As expected, the interference produced by a geostationary satellite is quite negligible if the radio-relay antenna happens to be pointed a few degrees away from the critical azimuth; it is still within the permissible limit, even if the antenna points at the satellite located at the critical azimuth.

Interference Produced by a String of Geostationary Satellites

To extend the computation to include a large number of geostationary satellites, it is necessary to sum all the contributions; i.e.,

$$(N_i)_n = \frac{N_{th}}{S_i} \sum_{i=1}^n F(\theta_i) G_r(\alpha_i) \quad (11)$$

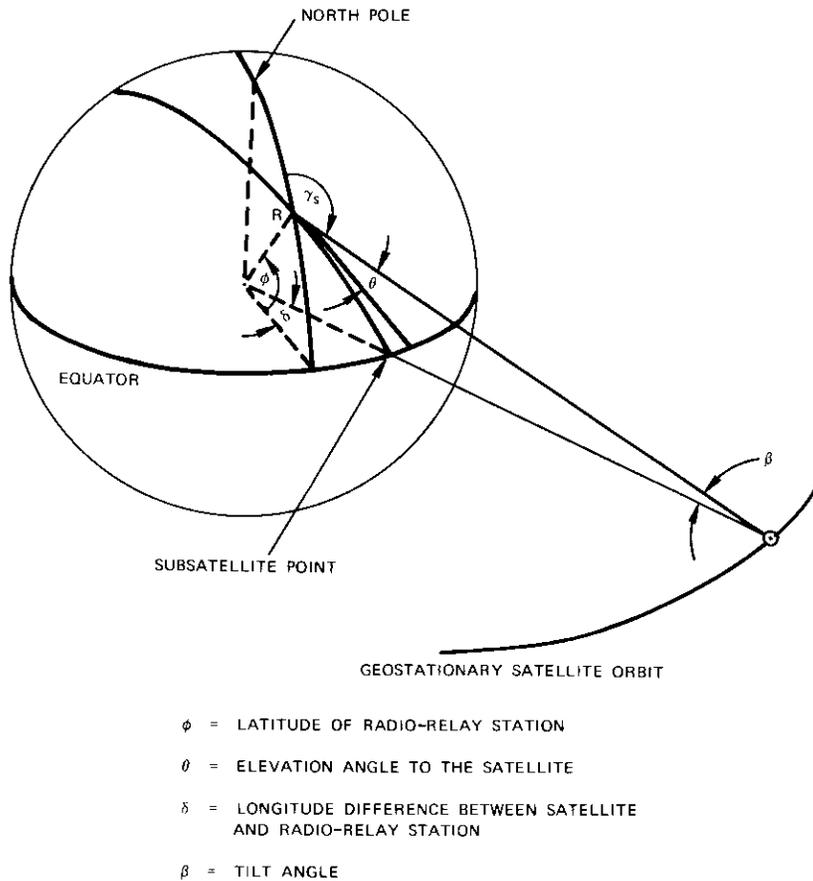


Figure 3. Geometry of a Geostationary Satellite and a Radio-Relay Receiving Station

where the subscript l refers to the satellite number. The results of a computation assuming satellites spaced at 3° of orbit arc are also shown in Figure 6. Since only one satellite can appear at the critical azimuth and since the on-beam interference is by far the largest of all the interference contributions, the degradation in a radio-relay beam pointing at the critical azimuth is not substantially changed from that caused by a single satellite. However, for a radio-relay beam pointed well away from the critical azimuth, the increase in interference noise is considerable. This again is to be expected, since, in this case, all satellites produce approximately the same amount of interference.

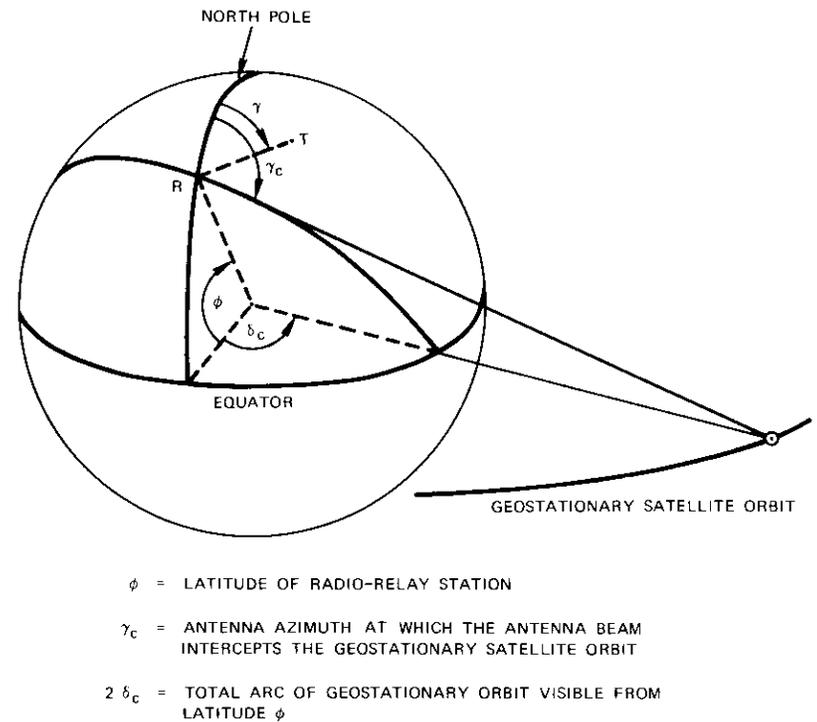


Figure 4. Radio Relay and Critical Azimuth

It can now be argued that, if the flux-density limit is raised for high elevation angles, but kept constant for low elevation angles, the interference for the radio relay pointing at the critical azimuth will remain substantially unchanged. However, this will increase the interference into radio-relay stations pointing far away from the critical azimuth. As an example, a computation was made for a flux-density limit given by

$$\begin{aligned}
 F(\theta) &= -152 \text{ dB } (W/m^2 \cdot 4 \text{ kHz}) & 0^\circ < \theta < 2^\circ \\
 &= -154 + \theta & 2^\circ < \theta < 20^\circ \\
 &= -134 \text{ dB} & 20^\circ < \theta < 90^\circ.
 \end{aligned} \quad (12)$$

The interference noise produced by a string of satellites is plotted as a dashed line in Figure 6.

Interference into an hrc

The total amount of interference noise produced in a radio-relay hrc can be expressed as a sum of the interference noise at each station: i.e.,

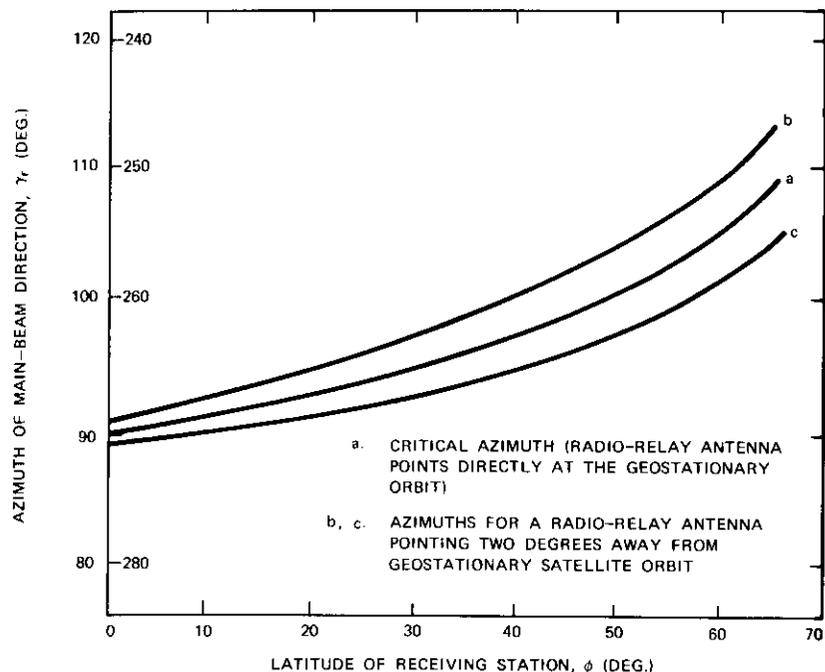


Figure 5. Critical Azimuth vs. Receiving Station Latitude

$$(N_i)_{nm} = \frac{N_{th}}{S_i} \sum_{k=1}^m \sum_{l=1}^n F(\theta_{lk}) G_r(\alpha_{lk}) \quad (13)$$

where the subscript k refers to the radio-relay station number.

An exact computation requires knowledge of the location and antenna pointing direction for each radio-relay station, as well as the location of each geostationary satellite. A solution which gives a good approximation, with reasonable assumptions, is required. This solution should be sufficiently flexible to include different flux-density limits.

The worst-case solution to this problem can be derived by assuming that all 50 antennas of the hrc point at the critical azimuth and that there is always a satellite located within the main beam of the antenna. However, the probability of such an hrc being installed is exceedingly small. An examination [6] of a total of almost 6000 antenna beam directions has shown that 127 are oriented so that their beam axes are within 2° of the geostationary satellite orbit. A similar number is obtained if the antenna pointing directions are considered to be randomly and uniformly distributed. It is therefore reasonable to assume that

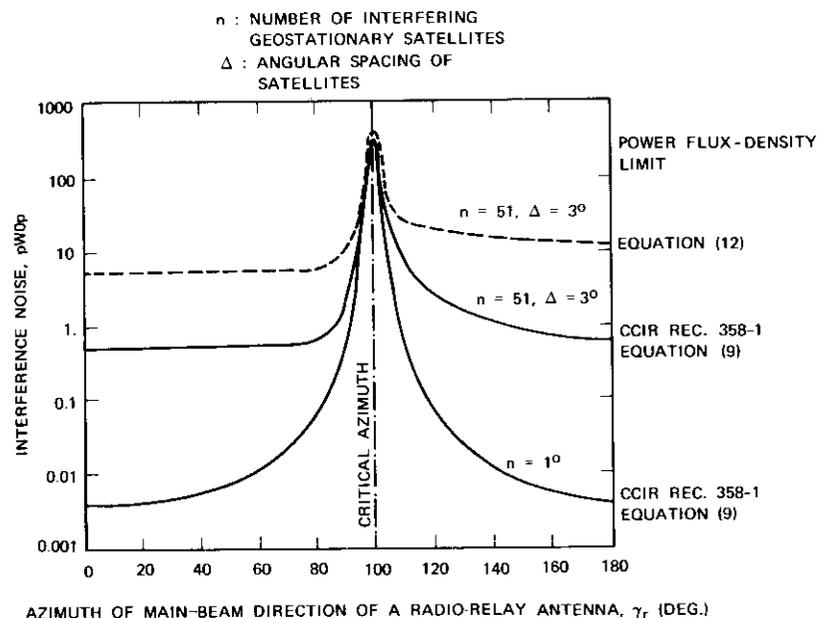


Figure 6. Interference to a Station at 50° Latitude

there is a certain amount of randomness in the antenna pointing directions of a radio-relay hrc.

The locus of the geostationary satellite orbit in coordinates centered on the antenna pointing direction, i.e., in the θ vs. γ plane, can now be plotted for each station, as shown in Figure 7. The loci appear as arcs; their shift with respect to the origin indicates the random pointing of the radio-relay antenna beam and the variation in their maximum elevation angle indicates the random choice of latitudes of the radio-relay station. Quite possibly, one of the orbits will pass through the origin. If 50 radio-relay stations located at about 50° latitude and a geostationary satellite population with an orbital spacing of 6° are assumed, the satellites can be represented as more than 1250 points which will be, to a large extent, randomly distributed in this presentation.

The summation over k and l can be replaced by an integration if a probability density function for the distribution of points in the scatter diagram of Figure 7 is assumed. The expected total interference noise, i.e., the average over all possible radio-relay hrc's, is then

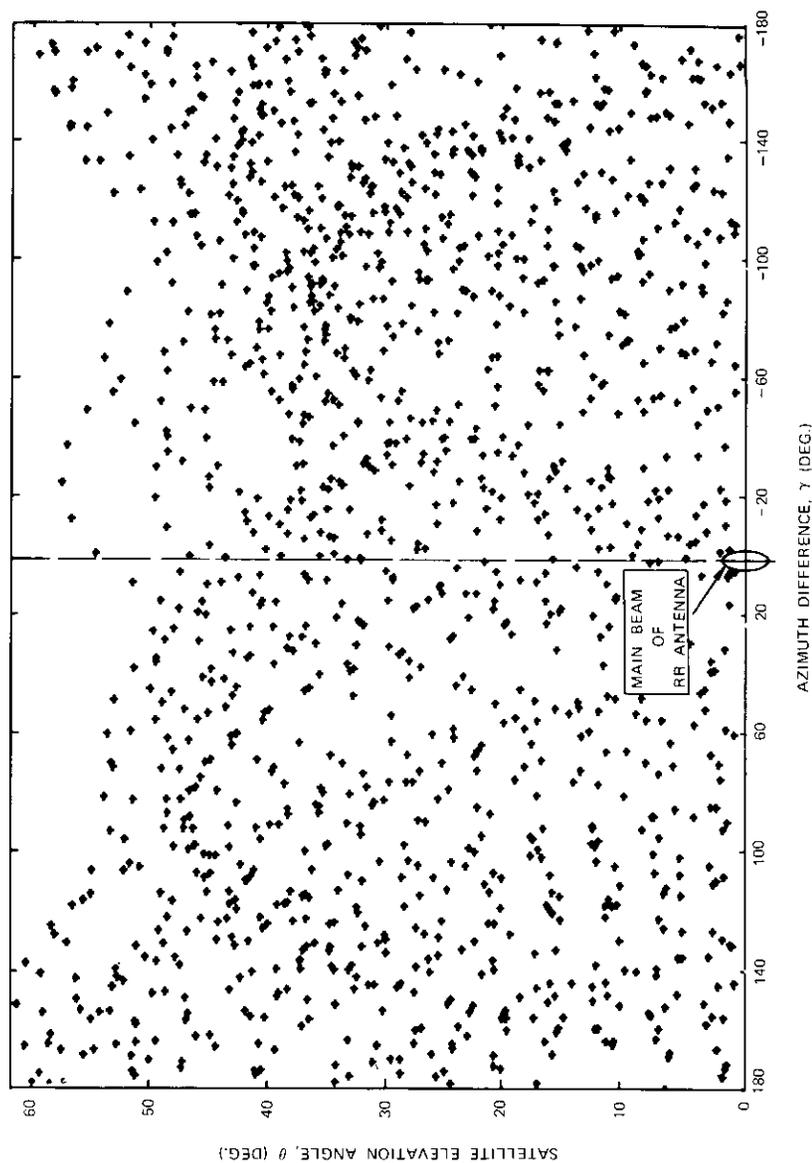


Figure 7. Scatter Diagram of Geostationary Satellites Plotted in Coordinates Centered on the Radio-Relay Antenna Beams

$$\begin{aligned} \bar{N}_i &= E[(N_i)_{nm}] \\ &= \frac{N_{th}}{S_i} \cdot nm \int_{\theta=0}^{\pi/2} \int_{\gamma=-\pi}^{+\pi} F(\theta) G_r(\theta, \gamma) P(\theta, \gamma) d\theta d\gamma. \end{aligned} \quad (14)$$

Furthermore, if it can be assumed that the distributions in the two axes are independent, then

$$P(\theta, \gamma) = P(\theta) P(\gamma). \quad (15)$$

Equation (14) reduces to

$$\bar{N}_i = \frac{N_{th}}{S_i} \cdot nm \cdot \int_{\theta=0}^{\pi/2} F(\theta) \bar{G}_r(\theta) P(\theta) d\theta \quad (16)$$

where

$$\begin{aligned} \bar{G}_r(\theta) &= E[G_r(\theta)] \\ &= \int_{\gamma=-\pi}^{\pi} G_r(\theta, \gamma) P(\gamma) d\gamma. \end{aligned} \quad (17)$$

This procedure reduces the antenna gain function, which is initially a function of γ and θ to a new function, $\bar{G}_r(\theta)$, which may be interpreted as an "average" elevation-dependent gain pattern.

The effect of the elevation angle, ϵ_i , of the radio-relay antenna beams can be incorporated by retaining equation (16) for the average interference and redefining $\bar{G}_r(\theta)$ as

$$\bar{G}_r(\theta) = \int_{\gamma=-\pi}^{\pi} \int_{\epsilon=\epsilon_{min}}^{\epsilon_{max}} G_r(\theta, \gamma, \epsilon) P(\gamma) P(\epsilon) d\gamma d\epsilon. \quad (18)$$

In a strict sense, \bar{N}_i is the average of the interference noise in all possible radio-relay hrc's.

In order to estimate the gain of an "average" radio-relay antenna as a function of the elevation angle of the satellite, some assumptions about the distribution of γ and ϵ are necessary. The antennas in the different hops of a particular radio-relay route may tend to point in the general direction of the route. The angle γ is a difference between the azimuth of the radio-relay antenna pointing direction and the azimuth of the direction to the sub-satellite point. The probability densities of both angles are difficult to estimate. A simple assumption is that γ is uniformly distributed between $-\pi$ and $+\pi$.

The sidelobe level patterns of the different antenna types and sizes used by terrestrial radio systems vary a great deal. Moreover, the pattern of each antenna exhibits fine structure. In the present case, the interference noise is formed by the superposition of many "entries." Every geostationary satellite contributes with a fraction of the total

acceptable noise at every receiving station of the terrestrial radio-relay link. The "smoothing" of the antenna pattern that is obtained by this superposition justifies the use of average patterns to neglect the side-lobe peaks and to fill in the nulls by following the general trend of the pattern.

Let the antenna pattern of a single terrestrial station be represented by equation (8) and let the radio-relay elevation angles be randomly distributed between $+3^\circ$ and -3° . The "average" radio-relay antenna pattern versus elevation angle is shown in Figure 8 for three different conditions: no pointing within 2° of the geostationary satellite orbit, polarization discrimination for satellite emissions entering the main beam of the radio-relay antenna, and unrestricted pointing and no polarization discrimination. It may be observed that such an average gain pattern is isotropic at an elevation angle slightly less than 5° and that it asymptotically decreases to -10 dB with higher elevation angles.

Since the average gain decreases considerably at high elevation angles, it can be seen from equation (16) that the average contribution of interference noise from satellites located at high elevation angles

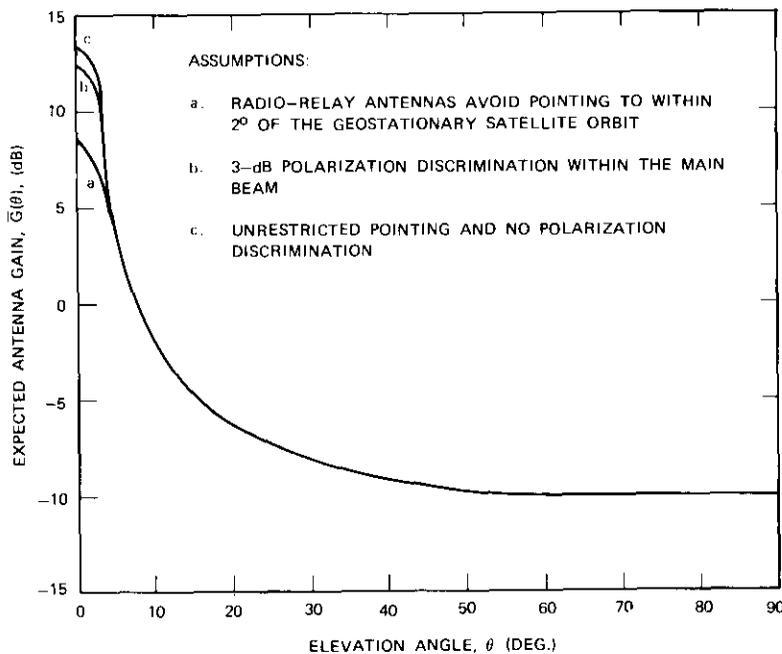


Figure 8. Expected Gain vs. Elevation Angle

will be quite small for a flux-density function, such as that given by CCIR Recommendation 358-1 [2], which increases slowly in comparison. One way to equalize the contributions from all elevation angles is to make $F(\theta) \cdot \bar{G}_r(\theta)$ a constant, i.e., to increase the flux-density limit as rapidly as the average gain function decreases.

In Figure 9, power flux-density curves corresponding to the three curves presented in Figure 8 are shown. With power flux-density limits such as those shown in Figure 9, the interference noise in an average hrc would be 1000 pW0p if

$$m = 50 \text{ hops}$$

$$n = 23 \text{ interfering satellites}$$

$$10 \log S_i = -124.7 \text{ dB}$$

$$N_{ih} = 25 \text{ pW0p per hop.}$$

Under these conditions, a significant number of hrc's would receive more than the 1000-pW0p interference noise received by an average hrc. The interference noise would also increase if there were a larger number of satellites considered. There are two available options to take these factors into account:

a. The power flux-density limit can be reduced by a certain factor at all elevation angles, while an elevation angle dependence resulting in constant $F(\theta) \cdot \bar{G}_r(\theta)$ is maintained.

b. A power flux-density limit can be synthesized such that $F(\theta) \cdot \bar{G}_r(\theta)$ is a given function of the elevation angle, θ .

A limit for which $F(\theta) \cdot \bar{G}_r(\theta)$ depends on the elevation angle appears in the SJM report [7], Geneva, 1971. Figure 10 shows the corresponding $F(\theta) \cdot \bar{G}_r(\theta)$ versus elevation angle. It can be observed that, at low elevation angles, a higher average interference results. Consequently, the interference noise into an average hrc is reduced without imposing a more severe power flux-density limit at the horizon. The limit at low elevation angles plays an important role, since it essentially constitutes the controlling limitation for all global-coverage antennas.

In another statistical approach taken by May and Pagonis [8], the location of the radio-relay stations was chosen along a random trendline and the antenna pointing directions were assumed to be uniformly distributed within $\pm 25^\circ$ of the trendline direction. A deterministic solution to equation (13) could therefore be found for any given flux-density limit for an hrc following the trendline. Generating a large

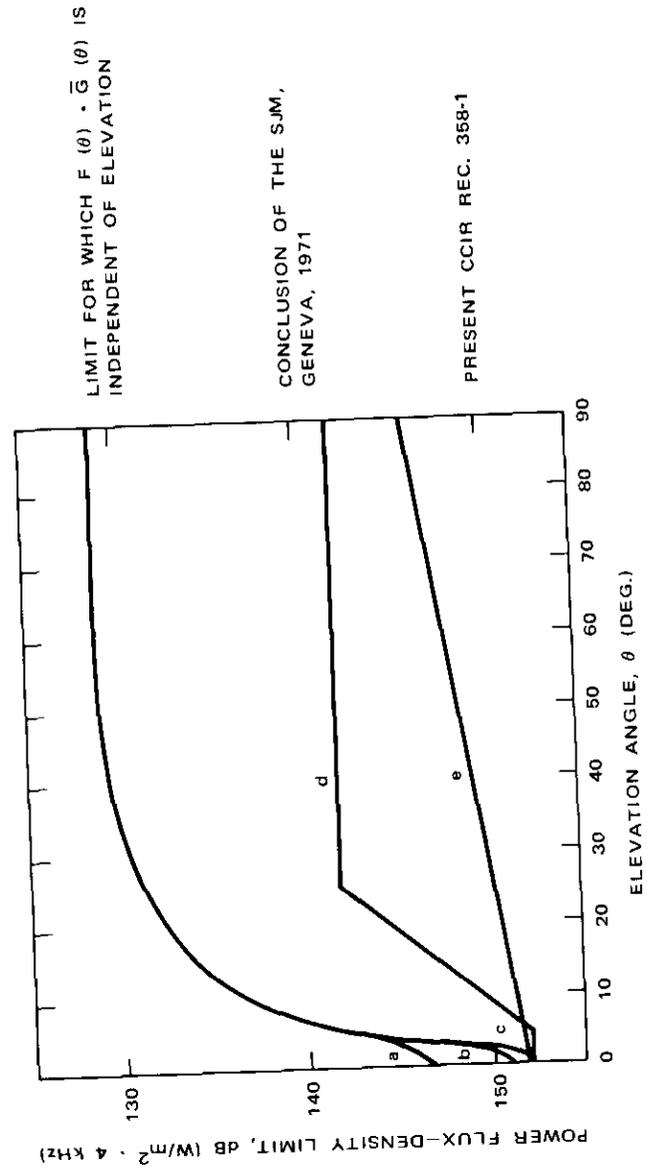


Figure 9. Power Flux-Density Limits vs. Elevation Angle

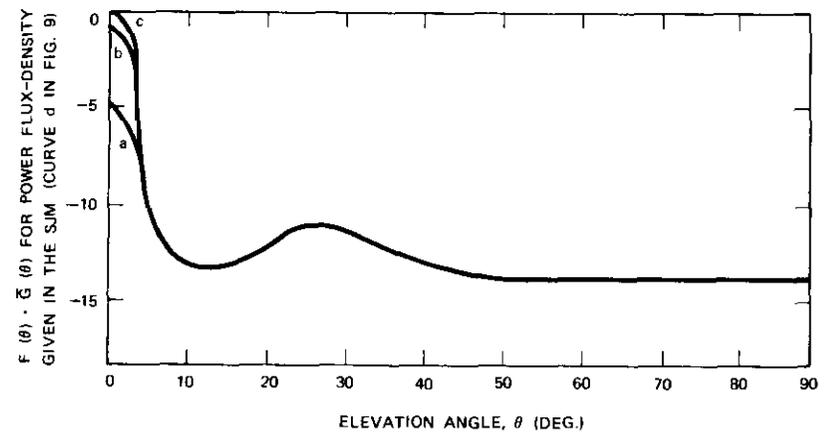


Figure 10. Relative $F(\theta) \cdot \bar{G}(\theta)$ Using the Power Flux-Density Limit of the SJM, Geneva, 1971

number of random trendlines made it possible to obtain not only the average interference into an hrc, but also the cumulative distribution function for the interference. Their results with an escalated flux-density limit substantiate the contention that the permissible power flux-density from geostationary satellites can be relaxed with increasing elevation angles at a faster rate than that indicated in the present CCIR Recommendation 358-1.

Reported calculations of interference [9], [10] into existing systems for the present and escalated flux-density limits have resulted in similar conclusions. The CCIR has recognized this phenomenon and it appears that [11] future recommendations will consider an increase in the permissible power flux-density limit of at least 10 dB between 5° and 25° of elevation.

Use of the Power Flux-Density

It was implicitly assumed during the previous discussions that all of the satellites radiate the full permissible flux-density toward all areas of the earth. The flux-density limit, $F[\theta(\beta)]$, as seen from the satellite, can be derived by a simple transformation of coordinates; i.e.,

$$\theta = \arccos \left(\frac{\sin \beta}{r} \right) \tag{19}$$

$$r = \frac{r_2}{r_1}$$

where r_2 and r_1 are the radius of the earth and the synchronous orbit, respectively. The total power radiated from a satellite can be evaluated by integration:

$$P = 2\pi \int_0^{\sin^{-1}(r_2/r_1)} R(\beta) F[\theta(\beta)] \sin \beta \, d\beta \quad (20)$$

where $R(\beta) = r_1 \cos \beta [1 - \sqrt{1 - \sec^2 \beta (1 - r^2)}]$. (21)

The total radiated power for a hypothetical satellite antenna having a pattern which illuminates only the earth with precisely the flux-density specified in the limits is given in Table 1. This table shows that it is not practicable to radiate the full permissible flux-density, since the total radiated power is excessive and since no realistic antenna can produce the exact beam shape.

In order to benefit from the increased limit at high elevation angles, the satellite has to be equipped with an earth-oriented antenna with a narrow beam. For narrow beams which are pointing away from the subsatellite point, and which are thus closer to the earth's edge, illumination of the earth's surface with a smaller power flux-density than is permissible at the beam center is required so that emission arriving at all angles of elevation satisfy the limits. Figure 11 shows the power flux-density limit as seen from the satellite, as well as the loci of usable power flux-density at antenna beam centers for various beamwidths. It can be observed that a satellite antenna of less than one-half-degree beamwidth can closely follow the rapid change of power flux-density as the edge of the earth is approached; however, broader satellite beams will be restricted by the power flux-density limit at the edge rather than at the beam center. The satellite antenna main beam pattern used to derive the loci plotted in Figure 11 was assumed to be

$$10 \log \frac{G_s(\beta_k)}{G_s(0)} = -3 \left(\frac{\beta_k}{\beta_0/2} \right)^2 \quad (22)$$

TABLE 1. RADIATED SATELLITE POWER

Flux-Density Limit, dB (W/m ² · 4 kHz)	P, W/4 kHz	P _T , W/500 MHz
$F(\theta) = -152 + \frac{\theta}{15}$ [eq. (9)]	0.119	14,875
$F(\theta)$ given in Ref. [11]	2.35	293,750

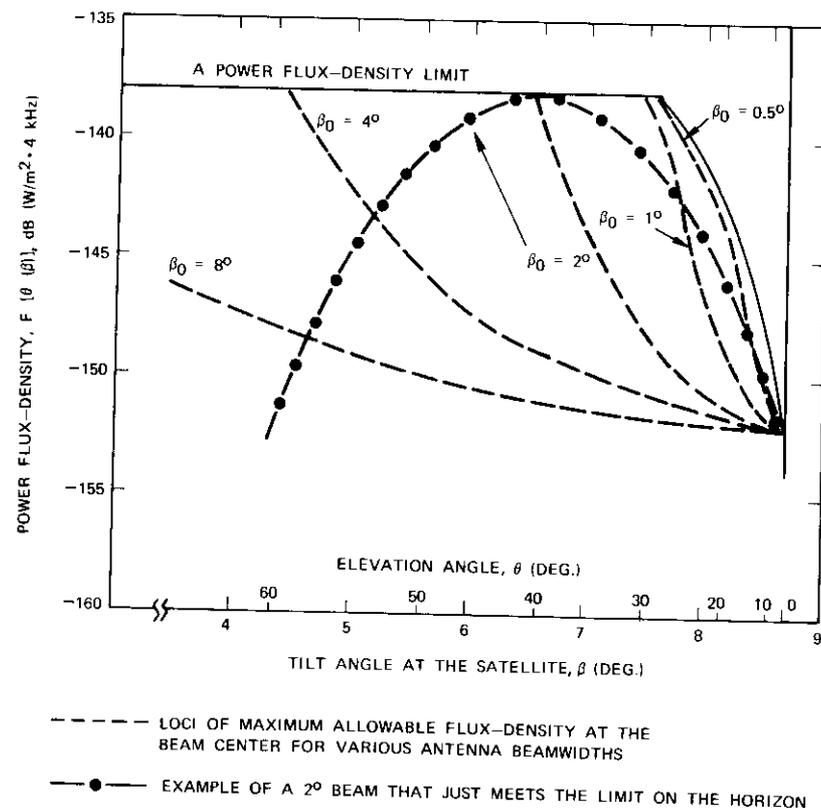


Figure 11. A Power Flux-Density Limit and Its Utilization by Geostationary Satellites

Frequency Dependence of Permissible Power Flux-Density at the Surface of the Earth from Communications Satellites

Fading Considerations

The noise budget of a communication system, a measure of the degradation of the baseband signal, can be decomposed in three parts with reference to the fading of the received carrier level. These three parts are first, fade-dependent noise, such as thermal noise; secondly, fade-independent noise, such as equipment, intermodulation, and quantization noise; and thirdly, partially fade-dependent interference noise.

In the presently shared 4- and 6-GHz frequency bands, multipath fading of the radio-relay carrier is the dominant mode of fading. At

these frequencies the received level of the interference carrier can be considered to be uncorrelated and the ratio of interference noise to thermal noise can be considered to be constant. The correlation existing between the attenuation in the signal path and the interference path determines the fade dependence of the interference noise. When the wanted and interfering carriers follow the same path, i.e., for on-beam interference, and the fading is caused by atmospheric attenuation rather than multipath transmission, a high level of correlation should exist between the received levels of the two carriers.

Since rain attenuation increases considerably and becomes the prime fade mechanism at higher operating frequencies, a very strong correlation can be expected between the signal attenuation and the attenuation of interference entering the main beam. The interference carrier levels entering the main beam of the terrestrial radio system can therefore be considerably larger than the RF thermal noise without significantly altering the system's operating margins.

Frequency Dependence of Radio-Relay Sensitivity to an Interfering Power Flux-Density

For the worst case of uncorrelated interference, the degradation in performance suffered by an "average" radio relay due to the presence of cochannel satellite interference is directly proportional to the isotropic interference transfer factor, S_i , defined previously. In order to estimate a reasonable frequency dependence for the power flux-density limit, it is necessary to determine S_i for equipment in representative frequency bands. Table 2 shows the bands selected for this study and the present U.S. terrestrial service allocations.

A survey of the technical characteristic of equipment operating in these bands was made and the corresponding isotropic interference transfer factors, S_i , are plotted in Figure 12. It can be seen from equation (2) that S_i will increase at least as rapidly as the square of the frequency of operation. Since the radio-relay system noise temperature also increases with frequency, S_i tends to increase as the cube of frequency.

The on-beam interference factor, S_m , computed by dividing S_i by the beam-center gain of the radio-relay antenna, is plotted as a function of operating frequency in Figure 13. It should be noticed that, for reasons of pointing accuracy, tolerance of the antenna surface, and other mechanical considerations, the antenna beamwidth (and, hence, the main beam gain) rather than the antenna aperture is maintained independent of frequency. Because the antenna gain is more or less constant with frequency, S_m also tends to increase as the cube of

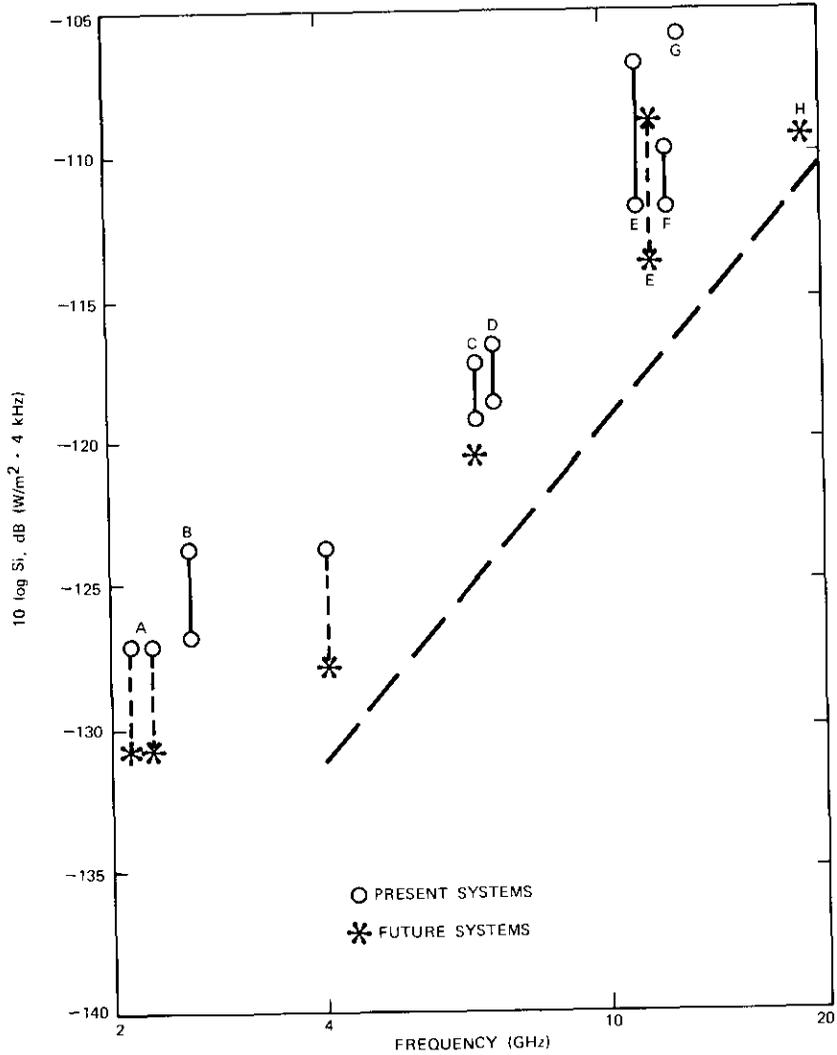
TABLE 2. FREQUENCY BANDS USED IN STUDY

Frequency Bands Considered, MHz	Present U.S. Allocations, MHz	Terrestrial Service
A 2,150-2,200	2,160-2,180 2,180-2,200	Common Carrier-F Private MW-F
B 2,500-2,550	2,500-2,690	Instructional TV
C 5,925-6,425	5,925-6,425	Common Carrier-F
D 6,625-7,125	6,575-6,875 6,875-7,125	Private MW-F TV Aux, Broadcast
E 10,700-11,700	10,700-11,700	Common Carrier-F
F 11,700-12,200	11,700-12,200	Common Carrier-M
G 12,200-12,700	12,200-12,700	Private MW-F
H 17,700-21,200	17,700-19,700 19,700-21,200	Common Carrier M-F None

frequency. On the logarithmic scales used in Figures 12 and 13, a cubic variation is translated as a straight line of slope $30 \log f$, as shown by the dashed line.

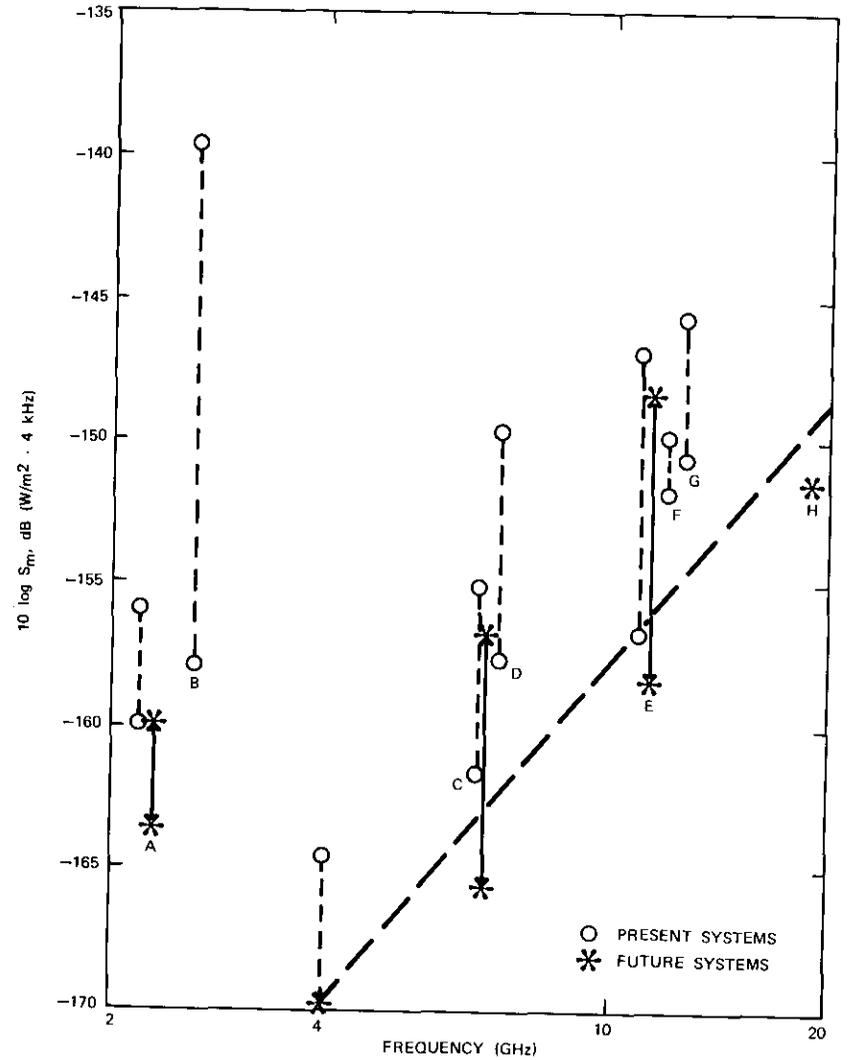
The radio-relay system degradation, measured by the ratio of interference noise to thermal noise, remains unchanged as long as the ratio of power flux-density to interference factor is not altered. Hence, it follows that the ratio of interference noise to thermal noise will remain constant if the frequency dependence of the permissible power flux-density at different frequency bands is similar to that of S_i or S_m .

Figure 14 shows the frequency dependence of the power flux-density at low elevation angles. Also shown as a reference is a straight line proportional to the cube of the frequency of operation and starting at -152 dB ($W/m^2 \cdot 4$ kHz) at 4 GHz. The conclusions of the SJM, Geneva, 1971 [8], have been included. It can be observed that the permissible power flux-density is lower than that suggested by this study for frequencies between 4 and 15.4 GHz. Above that frequency range, the SJM limit is specified with reference to 1 MHz. The value shown in Figure 14 corresponds to the hypothetical case of uniform spreading of the energy in every MHz of bandwidth. The exact resulting power flux-density in the worst 4 kHz, corresponding to the SJM limit of -115 dB (W/m^2 MHz), depends on each particular emission and will always be higher than the -139 -dB ($W/m^2 \cdot 4$ kHz) value shown in Figure 14.



NOTE: LETTERS A TO H IDENTIFY FREQUENCY BANDS FROM TABLE 1

Figure 12. Isotropic Interference Transfer Factors of Terrestrial Systems vs. Frequency



NOTE: LETTERS A TO H IDENTIFY FREQUENCY BANDS FROM TABLE 1.

Figure 13. On-Beam Interference Transfer Factors of Terrestrial Systems vs. Frequency

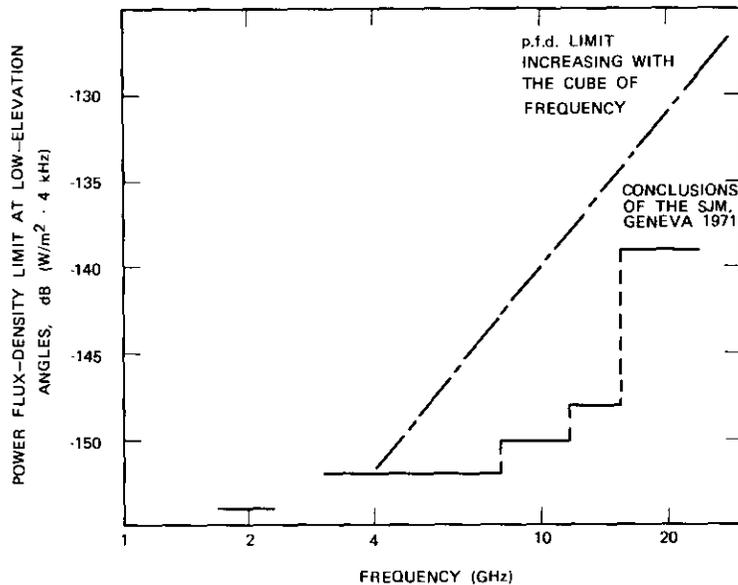


Figure 14. Examples of Possible Frequency Dependence for Power Flux-Density Limit

Power Flux-Density Requirements of Satellite Systems

The power flux-density required to operate a communications satellite system increases with frequency for many of the same reasons that the radio-relay system can accept a larger power flux-density without increases in interference. These reasons are:

a. Mechanical considerations cause antenna gain limitations leading to constant gain or beamwidth as a function of frequency. The reduction in effective aperture requires a corresponding increase in power flux-density, i.e., with the square of the frequency.

b. The receiving system temperature increase is caused partially by equipment characteristics and partially by the antenna noise temperature increases during periods of rainfall and cloud cover. The latter is proportional to the ambient temperature of the atmosphere and its

attenuation. Since the attenuation increases with frequency, so does the antenna noise temperature.

c. Atmospheric attenuation caused by rain and cloud absorption requires larger operating margins. As a result, the power flux-density requirements of satellite systems increase at a rate between the square and the cube of the frequency of operation.

TERRESTRIAL RADIO SYSTEM INTERFERENCE TO SATELLITES IN SHARED FREQUENCY BANDS

Terrestrial radio systems cause a different interfering power flux-density at every point of the geostationary satellite orbit. This effect occurs because of the nonuniform distribution of these systems on the surface of the earth and also because direct on-beam exposures from the transmitting antennas of the terrestrial systems may or may not be present.

For this interference mode, no correlation can be considered between fading of the desired signal received by the satellite and fading of the interfering signals of the terrestrial radio systems. The total interference into a satellite receiver should therefore be limited to a small portion of the thermal noise to avoid significantly modifying the operating margin.

There is a significant difference between the properties of the nearly isotropic interference aggregate caused by the sidelobes of the terrestrial antennas and the properties of interference caused by on-beam exposures. The effect of the sidelobe interference can be considered as an increase in the apparent temperature of the earth as seen from the satellite, whereas, at the horizon, isolated on-beam exposures may be many orders of magnitude larger. To keep this increase of the earth's apparent temperature within tolerable limits, the maximum power delivered into the antennas of terrestrial transmitters must be limited. On the other hand, to control the interference from on-beam exposures, a limitation of the maximum e.i.r.p. of terrestrial transmitters toward the geostationary satellite orbit is desirable.

Interference Caused by the Sidelobes of Terrestrial Radio Systems

At the input of a satellite receiver, the total interference power caused by the terrestrial radio system population sharing the same frequency band is given by

$$I = \sum_{M} P_{E}(\lambda) G(\lambda) K(\lambda)$$

$$l(\beta_k, f) = \frac{(\lambda/4\pi)^2}{R^2(\beta_k)} \quad (24)$$

The resulting ratio of interference to up-link thermal noise is therefore

$$\frac{I}{N} = \frac{\sum_{k=1}^M P_k G_r(\alpha_k) G_s(\beta_k) l(\beta_k, f)}{kTB} \quad (25)$$

The expected value of the interference-to-noise ratio can be derived by taking the average values of the gains, the power, and the path loss:

$$E\left(\frac{I}{N}\right) = \frac{M \bar{P} \bar{G}_r \bar{G}_s \bar{l}(f)}{kTB} \quad (26)$$

where the bar above a symbol indicates its average value. In order to control the sidelobe interference from radio relays into satellite receivers, the power from the average visible transmitter, \bar{P} , should be limited to

$$\bar{P} \leq \frac{kTB}{M \bar{G}_r \bar{G}_s \bar{l}(f)} \left(\frac{I}{N}\right)_d \quad (27)$$

where $(I/N)_d$ is the design value of interference-to-noise ratio.

The receiving antenna gain at the satellites does not strongly affect the permissible radio transmitter power, since a high-gain receiving antenna will receive interference from a proportionally smaller area at the surface of the earth. Therefore, for an earth-coverage receiving antenna, if

$$T = 1500^\circ\text{K}$$

$$10 \log \bar{G}_s = 18 \text{ dB}$$

$$10 \log \bar{G}_r = 0 \text{ dB}$$

then the average radio-relay transmitter power, \bar{P} , causing the design value of interference-to-thermal noise ratio, $(I/N)_d$, is

$$10 \log \bar{P} \leq 40.2 + 20 \log f + 10 \log B - 10 \log M + 10 \log (I/N)_d \quad (28)$$

where f is in GHz and B is in MHz.

In the presently shared 6-GHz band, the limit of permissible terrestrial transmitter power set by the CCIR is 13 dBW [3]. Introducing into equation (28) a total of 10,000 transmitters which do in fact have the maximum permissible power results in an interference which more

smaller than 2.5 MHz. With increasing frequency, the average hop length of terrestrial systems is expected to be smaller due to propagation considerations and, therefore, the total number of visible transmitters can safely be assumed to increase with frequency, perhaps more than compensating the $20 \log f$ term in equation (28).

In frequency bands above about 12 GHz, digital modulation methods are expected to assume increasing importance [7]. Therefore, the transmitter power will no longer be concentrated in a narrow frequency band, but will be spread over large bandwidths. As a consequence, it becomes more desirable to specify the power limitations of terrestrial systems with reference to a bandwidth unit for such frequency bands. Such a bandwidth unit could be 20 MHz, which is probably the smallest bandwidth to be utilized at these frequency bands by terrestrial radio relays. As an example, a limitation of 10 dB(W/20 MHz) would result in a negligible (-8 -dB) ratio of interference to thermal noise at 15 GHz with a total of one million terrestrial radio transmitters.

Interference Caused by a Single On-Beam Exposure from a Terrestrial Radio System

The number of interfering signals resulting in on-beam exposures of the satellite is expected to be quite small. As a matter of fact, new radio routes can be planned to avoid this interference situation entirely. The corresponding ratio of interference noise to thermal noise using the same definitions is given by

$$\frac{I}{N} = \frac{P_k G_r(0) G_s(\beta_k) l(\beta_k, f)}{kTB \Delta l(\theta_k, f)} \quad (29)$$

$$\Delta l(\theta_k, f) = \text{antilog}_{10} \frac{\Delta L(\theta_k, f)}{10} \quad (30)$$

where ΔL is the excess path loss due to atmospheric absorption. If $T = 1500^\circ\text{K}$,

$$10 \log \left(\frac{I}{N}\right) = -26.8 + (\text{e.i.r.p.})_r + 10 \log A - 10 \log B - \Delta L(\theta_k, f) \quad (31)$$

$$(\text{e.i.r.p.})_r = 10 \log P_k G_r(0) \quad (32)$$

where B is in MHz, and A is in square meters.

Figure 15 shows I/N versus the e.i.r.p. per MHz toward a geostationary satellite of a radio-relay transmitter. The effective area, A , of the satellite receiving antenna has been used as a parameter. Effective areas of the order of 1 m^2 can reasonably be expected in future

satellites. (The INTELSAT IV spot-beam antenna is 50 inches in diameter.)

The atmospheric attenuation, $\Delta L(\theta, f)$, increases with frequency and decreases with elevation angle. At low frequencies, with a radio-relay antenna pointing a few degrees above the local horizon, ΔL becomes negligible. A geostationary satellite located within the main beam of such a station will receive the most interference.

From Figure 15 it becomes clear that an e.i.r.p. density of about 27 dB(W/MHz), or equivalently, 40 dB(W/20 MHz), would double the up-link thermal noise in the absence of attenuation above free space for a satellite using a 1-m² receiving antenna. Such a level is considered tolerable since there will generally be some attenuation and since the satellite receiving antenna gain toward that particular location will not be the full main beam gain.

SUMMARY AND CONCLUSIONS

This paper has presented general considerations related to frequency sharing between communications satellite systems and terrestrial radio systems. In particular, the sharing constraints required to prevent harmful interference between the space segment and the terrestrial systems have been discussed.

The limitations on the permissible power flux-density at the surface of the earth from geostationary communication satellites was investigated. It was concluded that such a limit should be a function of both the angle of arrival above the horizon and the frequency of transmission. Arguments have been presented which indicate that the dependence on the angle of arrival above the horizon can be quite significant. To reflect the decreased sensitivity of radio-relay systems to interference at higher operating frequencies, it should be possible to escalate the power flux-density radiated by geostationary satellites at a rate between $20 \log f$ and $30 \log f$. This is fortunate, since the power flux-density requirements of communication satellite systems are also expected to increase at a comparable rate.

It was further shown that interference into the satellite receiver requires a limitation on both the transmitter power and the e.i.r.p. of the radio-relay station. In view of the probable use of only digital modulation techniques for radio-relay systems operating above 15 GHz, it may be permissible to specify the power and e.i.r.p. limitations with reference to a bandwidth unit. Consideration of interference into narrow-beam satellite receiving antennas indicates that the e.i.r.p. limitation of terrestrial radio systems toward the geostationary satellite orbit should

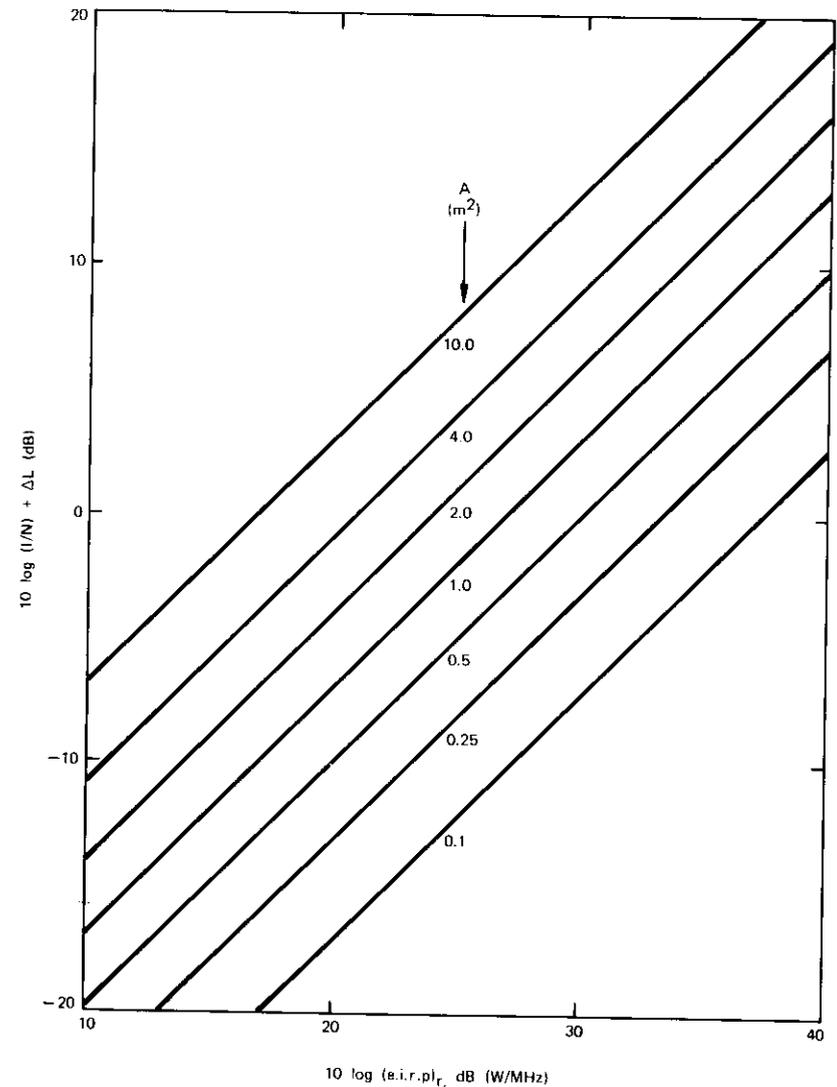


Figure 15. Interference-to-Thermal-Noise Ratio at the Satellite Receiver as a Function of the e.i.r.p. of the Terrestrial Transmitter Toward the Satellite

clusion was derived with respect to permissible transmitter power when the potentially increasing density of radio-relay transmitters on the surface of the earth was considered.

The required sharing constraints related to interference, caused as well as received by the space segment of communication satellite systems, appear to impose no serious limitation on the future development of either communication satellite or terrestrial radio systems in shared bands. The need for exclusive bands for terrestrial systems and satellite systems arises mainly from consideration of interference with the earth segment of communication satellite systems.

LIST OF PRINCIPAL SYMBOLS

A	Effective aperture of antenna
B	Unit of bandwidth for satellite systems
b	Unit of bandwidth (4 kHz for <i>FDM/FM</i> systems)
$F(\theta)$	Power flux-density, in W/m^2 , per unit bandwidth arriving at an elevation angle θ
f	Frequency of transmission
$G_r(\alpha)$	Radio-relay antenna gain at an angle α offset from the direction of pointing of the main beam
$G_r(0)$	Radio-relay antenna gain in the center of the main beam
$G_s(\beta)$	Satellite antenna gain at an angle β offset from the direction of pointing of the main beam
$G_s(0)$	Satellite antenna gain in the center of the main beam
<i>hrc</i>	Hypothetical reference circuit
I	Interfering carrier level
k	Boltzmann's constant
M	Number of radio-relay transmitters in the main beam of the satellite antenna
m	Number of radio-relay stations in a hypothetical reference circuit
N	RF thermal noise in the significant bandwidth, B
N_i	Baseband noise resulting from interference

N_{th}	Design value for the thermal noise per hop
n	Number of geostationary satellites visible from a radio-relay station
R	Path length between the radio-relay station and the geostationary satellite
r	r_2/r_1
r_1	Radius of geostationary satellite orbit
r_2	Radius of the earth
S_i	Isotropic interference transfer factor; power flux-density per unit bandwidth which results in $N_i = N_{th}$ for an isotropic antenna
S_m	On-beam interference transfer factor [$S_i/G(0)$]
T	Equivalent receiving system temperature of the system being interfered with
α	Angle between the directions of the main beam of the radio-relay antenna and the interfering geostationary satellite
α_0	Half-power beamwidth of the radio-relay antenna
β	Tilt angle, in satellite-centered coordinates, between the directions to the subsatellite point and the radio-relay station
β_k	Angle between the direction of the main beam of the satellite antenna and the radio-relay station, k
β_0	Half-power beamwidth of the satellite antenna
β'	Angle between the directions of the main beam of the satellite antenna and the subsatellite point
γ	$\gamma_s - \gamma_r$
γ_c	Critical azimuth
γ_r	Azimuth of the main beam of the radio-relay antenna
γ_s	Azimuth of the geostationary satellite
δ	$\delta_s - \delta_r$
δ_r	Longitude of the radio-relay station
δ_s	Longitude of the geostationary satellite
ϵ	Elevation angle of the radio-relay antenna beam
θ	Elevation angle of the geostationary satellite
λ	Wavelength of transmission
ϕ	Latitude of the radio-relay station

REFERENCES

- [1] R. G. Gould, "Protection of the Stationary Satellite Orbit," *Telecommunication Journal*, Vol. 34, No. 8, August 15, 1967, pp. 307-312.
- [2] International Radio Consultative Committee (C.C.I.R.), "Communication-satellite Systems and Line-of-sight Radio-relay Systems Sharing the Same Frequency Bands: Maximum Allowable Values of Power Flux-density at the Surface of the Earth Produced by Communication Satellites," Recommendation 358-1, *XIIth Plenary Assembly, New Delhi, 1970*, Vol. IV, Part 1, p. 288, Geneva: International Telecommunication Union, 1970.
- [3] International Radio Consultative Committee (C.C.I.R.), "Line-of-sight Radio-relay Systems Sharing the Same Frequency Bands as the Space Station Receivers of Active Communication-satellite Systems: Maximum Equivalent Isotropically Radiated Power of Line-of-sight Radio-relay System Transmitters," Recommendation 406-2, *XIIth Plenary Assembly, New Delhi, 1970*, Vol. IV, Part 1, pp. 289-290, Geneva: International Telecommunication Union, 1970.
- [4] International Radio Consultative Committee (C.C.I.R.), "Communication-satellite Systems and Line-of-sight Radio-relay Systems Sharing the Same Frequency Bands: Maximum Allowable Values of Interference in a Telephone Channel of a Radio-relay System," Recommendation 357-1, *XIIth Plenary Assembly, New Delhi, 1970*, Vol. IV, Part 1, pp. 286-287, Geneva: International Telecommunication Union, 1970.
- [5] International Radio Consultative Committee (C.C.I.R.), "Hypothetical Reference Circuits: Definitions of Terms and Documentation," Recommendation 390-1, *XIIth Plenary Assembly, New Delhi, 1970*, Vol. IV, Part 1, pp. 34-35, Geneva: International Telecommunication Union, 1970.
- [6] International Radio Consultative Committee (C.C.I.R.), "Exposures of the Antennae of Radio-relay Systems to Emissions from Communication-satellites," Report 393, *Documents of the XIth Plenary Assembly, Oslo, 1966*, Vol. IV, Part 2, pp. 435-443, Geneva: International Telecommunication Union, 1967.
- [7] International Radio Consultative Committee (C.C.I.R.), "Report of the Special Joint Meeting, Geneva, 3 February-3 March 1971," Part I, p. 16, Geneva: International Telecommunication Union, 1971.
- [8] A. S. May and M. J. Pagonis, "Model for Computation of Interference to Radio-relay Systems from Geostationary Satellites," *Bell System Technical Journal*, Vol. 50, No. 1, Jan. 1971, pp. 81-102.
- [9] International Radio Consultative Committee (C.C.I.R.), "Communication-satellite Systems Sharing the Same Frequency Bands as Line-of-sight Radio-relay Systems: Detailed Analysis of Interference Relations Between Stationary Communication Satellites and a Long East-west Radio-relay System," American submission, C.C.I.R. Study Groups, Period 1963-1966, Document IV/237-E, IX/229-E, March 18, 1966.

- [10] International Radio Consultative Committee (C.C.I.R.), "Power Flux Density at the Surface of the Earth from Communication Satellites in the Band 1-10 GHz," Canadian submission, C.C.I.R. Special Joint Meeting, Geneva, 1971, Document M/93-E, Nov. 18, 1970.
- [11] International Radio Consultative Committee (C.C.I.R.), "Power Flux-density at the Surface of the Earth from Communication Satellites in the Band 1 to 10 GHz," Report 387-1, *XIIth Plenary Assembly, New Delhi, 1970*, Vol. IV, Part 1, pp. 300-309, Geneva: International Telecommunication Union, 1970.

ACKNOWLEDGMENT

The authors wish to express their gratitude to Mr. Emeric Podraczky for his guidance and contributions to this work.



Dr. Nand Kishore M. Chitre was born in Calcutta, India. He received the B.Sc. and M.Sc. degrees in physics from the Banaras Hindu University in 1953 and 1955 and the M. Tech. degree in Electronics from the Indian Institute of Technology, Kharagpur, in 1956. He studied at the electrical engineering department of the Imperial College of Science and Technology, London, and was awarded the Ph.D. degree in physics from the University of London in 1962. Dr. Chitre joined COMSAT in December, 1968. He is Manager of the Systems Simulation Branch in the Systems Laboratory of COMSAT Laboratories.

Jorge C. Fuenzalida was born in Santiago, Chile. After finishing engineering school, he emigrated to the U. S. A. in 1965. He received the M.S.E.E. degree from Columbia University in 1966, and subsequently was a Ph.D. candidate in systems engineering at the University of Pennsylvania. Mr. Fuenzalida is a member of the technical staff in the Systems Laboratory at COMSAT Laboratories, working on interference and non-linear intermodulation in FM systems, on satellite capacity optimization problems, on frequency sharing between satellites and radio-relay systems and on orbit and spectrum sharing. He has been active in the Study Group IV of the C.C.I.R.



Index: automatic control, communication satellites, computers, ground support equipment, multi-channel communication, power measurement, radiometers

A COMPUTER-CONTROLLED SATELLITE SIGNAL MONITORING SYSTEM

E. E. STEINBRECHER AND L. F. GRAY

ABSTRACT

A computer-controlled monitoring system to determine automatically the power, center frequency, and rms deviation of 50 or more carriers radiated from a multicarrier communications satellite is described. The measurements are made without degradation of telemetry reception or command functions in an earth station that is approximately 10 dB less sensitive than the standard type of international earth station. The frequency-shift radiometer principle is used and precipitation-induced noise temperature variations are cancelled. The system monitor has been in operation at Andover, Maine, since May 1970.

INTRODUCTION

To achieve the desired noise performance in each telephone or television channel at a given earth station, sufficient carrier power must be radiated by a communications satellite. On the other hand, the total power available for all carriers is limited by the satellite size; to obtain the optimum efficiency and, thereby, the greatest number of available carriers, the power in each carrier should not exceed the value required to provide an adequate margin.

The satellite transponders respond proportionately to the signals from the earth stations. Hence, each earth station illuminating the satellite must adjust its radiated power so that the corresponding down-link power is correct. While the power generated for each carrier at each earth station can be measured quite precisely, the overall losses and gains, including the 6-GHz earth station and satellite antenna characteristics, and satellite look angle are less precisely known. It has been found in practice that, for each carrier, the desired equivalent isotropic

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

radiated power (e.i.r.p.) from the satellite is best adjusted by an actual down-link measurement.

As each earth station in the INTELSAT system is normally instrumented to receive only carriers directed to it, in 1967, the Interim Communications Satellite Committee (ICSC) decided that special instrumentation should be installed to measure the down-link power of all carriers at appropriate monitoring stations. At that time, it was further agreed that center frequencies and rms deviations should be measured. It was subsequently agreed that system monitoring functions should be combined with tracking, telemetry, and command (TT&C) functions at four locations, namely, Andover (Maine), Paumalu (Hawaii), Fucino (Italy), and Carnarvon (Australia). At a still later date, it was agreed that the monitoring equipment should additionally have the capability of being connected to the communications antennas at the four locations so that telephone channel noise measurements could be made under approximately the same noise conditions enjoyed by the communications stations that would normally receive each carrier.

When the satellite transponders are fully loaded, the operating point is in the nonlinear region of the transponder amplifier transfer characteristic. In this region, the intermodulation products will increase rapidly if the input power is increased. It is therefore necessary to determine the transponder operating point in terms of the total power output within 0.5 dB so that the availability of additional channel capacity can be ascertained. It is also necessary to determine the power of each individual carrier to within 0.5 dB.

Satellite antenna gains have been increasing as the art progresses, and consequently the satellite channel capacity is becoming limited by available bandwidth rather than by available power. It is therefore important that the center frequency and deviation values are also kept within specified limits to ensure conservation of bandwidth. This makes it necessary to determine center frequency to within 50 kHz and deviation to within 0.5 dB.

A monitoring system designed to perform the functions described previously has been in operation at the Andover, Maine, TT&C station since May 1970. The equipment is capable of determining absolute values of satellite e.i.r.p. to within 0.3 dB and relative values to within 0.1 dB. The rms deviation of each carrier can be measured to within 0.5 dB and the center frequency of the modulated carriers can be measured to within 50 kHz. Continuous monitoring of any desired carrier is possible for investigative purposes, but simultaneous continuous monitoring of all carriers was not considered to be justified. The carriers are therefore measured sequentially and in the test system

35 carriers from one satellite are measured, and the power, center frequency, and rms deviation for all carriers are printed out in approximately 10 minutes.

The system to be described was developed for use at TT&C stations having a G/T of approximately 32 dB at 4 GHz. The operating G/T at the communications stations receiving each carrier varies between 40.7 and 43 dB. Consequently, for each carrier, the carrier-to-noise ratio in the Carson's Rule bandwidth is as much as 11 dB less at the monitoring station than at the communications station. The equipment was designed to operate with these maximum errors at carrier-to-noise values in the Carson's Rule bandwidth down to 0 dB. At lower values, the errors increase and the center-frequency and rms deviation errors increase more rapidly than the power-measurement errors.

Monitoring Methods

The monitoring methods used will be described briefly in the following paragraphs and in detail in the next section. In that section, equation (5) will show that the satellite beam-center e.i.r.p. is given by

$$P_c = 2\pi d^2 A_e C B_i \frac{P_s}{P_L} \frac{P_L}{P_r}, \text{ in watts}$$

- where
- d = range from measuring station to satellite, in meters
 - A_e = aspect-angle correction factor
 - C = flux density from a selected celestial source, in W/Hz/m²
 - B_i = Carson's Rule bandwidth of carrier being measured
 - P_s = power received from satellite at antenna reference point
 - P_L = power introduced by a reference noise lamp at antenna reference point
 - P_r = power received from celestial source at antenna reference point.

The system is calibrated against a celestial source whose flux has been carefully measured by radio astronomers [1]. The ratio P_L/P_r is established for each operating frequency. This ratio remains constant as long as the antenna gain, noise lamp output, and the coupling factor of the directional coupler that connects the noise lamp to the low-noise receiver input remain constant. It is expected that, for any given monitoring station, the primary calibration need not be repeated more often than every six months.

For satellites in stable synchronous orbits of low eccentricity, the range, d , from a given satellite to a given TT&C station will be constant within narrow limits. The flux, C , is the value existing at the time of the primary calibration, and the bandwidth, B_i , will not change unless the telephone channel capacity of the carrier is changed. It is therefore convenient to combine d , C , B_i , and the lamp-star ratio (P_s/P_r) into a single calibration factor for each frequency. The basic procedure for measuring each carrier is to measure the ratio P_s/P_L . If the satellite orbital plane has some inclination to the earth's equatorial plane or if the spin axis is misaligned, the look angle from a given monitoring station to the satellite antenna will change with time of day, and the value of A_e must be constantly determined. The satellite e.i.r.p. is then established by taking the product of the calibration factor (P_s/P_L) and A_e .

The measurement of P_s/P_L involves the separation of the values of P_s and P_L from the background noise whose power may equal or exceed the value of P_s or P_L alone. The background noise is a function of atmospheric conditions and the noise generated in the low-noise amplifier; therefore, it must be evaluated at each measurement. The actual power measurements are performed with instruments that follow approximately 100 dB of amplification after the antenna reference point; this amount of gain is subject to some fluctuation.

The radiometer, which is the classical device used by radio astronomers to measure power that is low in comparison with the background noise, is designed to avoid the effects of receiver gain fluctuations. The frequency-shift radiometer principle is used in the monitor under discussion to establish the ratio P_s/P_L . The procedure is illustrated in the amplitude-frequency display at the bottom of Figure 1. Power measurements are made in three frequency bands for each carrier. One measurement is made at the carrier frequency and the other two are made at unoccupied slots above and below each carrier.

Briefly, the procedure is as follows. The power measurements at the carrier frequency include both the carrier power and the background noise power. At this frequency, the background noise cannot be separately measured without turning the antenna away from the satellite, but this is not permitted because one requirement of the system is uninterrupted telemetry reception. The background noise at the carrier frequency is estimated by measuring the background noise in the test slots above and below the carrier frequency and taking the average. This value is then corrected for the ratio of bandwidths between the carrier and test slots and is subtracted from the carrier measurement to obtain the carrier power value alone. The lamp power value is obtained by switching on the noise lamp, again taking measurements in the test

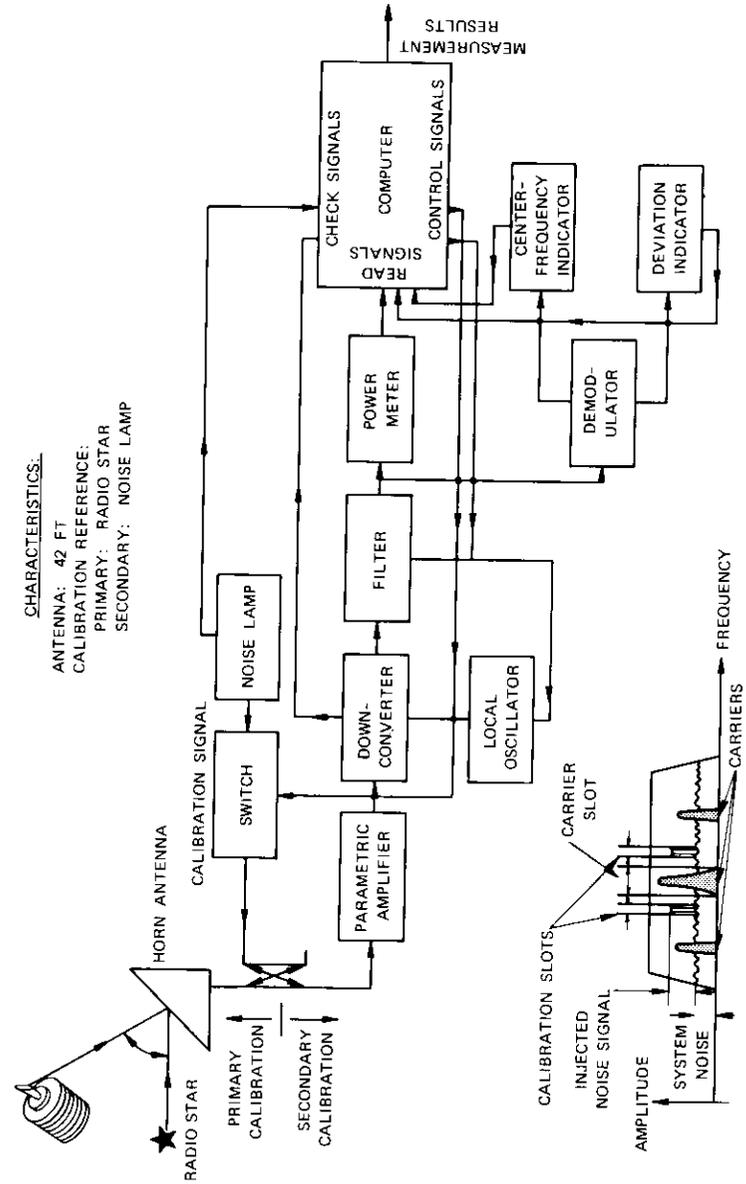


Figure 1. Simplified Block Diagram

slots, and subtracting the background noise power. The ratio of P_s/P_L can then be determined.

Conventional switched radiometers are switched at high speeds in the 10- to 1000-Hz range; high-gain audio amplifiers are used to measure small differences in input signals. In the system being described, the switching rate is slowed (in the range of 1 to 2 Hz), but is fast enough to avoid errors due to receiver gain or noise temperature variations. Instead of high-gain audio amplifiers, a thermoelectric power meter and a digital voltmeter are used to obtain high resolution. Therefore, the combination of the switched local oscillator, switched noise lamp, power meter, and associated digital voltmeter shown in Figure 1 constitutes a frequency-shift radiometer.

The center-frequency and rms deviation are determined directly from a calibrated demodulator. When the monitoring system is connected to a standard communications antenna to obtain normal values of carrier-to-noise ratio, the demodulator is connected to appropriate baseband filters that are slightly above the operating baseband to permit a determination of telephone channel noise.

Control of the local oscillator frequencies, filter selection, and necessary computations are carried out by a small process-control computer. Figure 1 is a simplified block diagram of the system. The computer output appears on a page printer or, alternatively, punched paper tape is produced to permit transmission of the data through the telemetry transmission facilities to the Operations Center at COMSAT-INTELSAT Headquarters in Washington, D.C.

Summary of Results

The result of a year of operation of the power-measuring system indicates that the experimental error does not exceed 0.3 dB if the range to the spacecraft, the spacecraft transmitting antenna beam patterns, and the satellite viewing angles are accurately known. An error analysis to support this figure is given in Appendix A. The center-frequency and rms deviation errors can be directly determined by comparison with known test signals. The experimental errors for these measurements are estimated to be 50 kHz and 0.5 dB, respectively.

The power-measurement system was found to be useful for bringing new earth stations into the system, as well as for maintaining power values radiated by existing stations. It was also used to confirm the feasibility of modifying the INTELSAT III transmission parameters to use more efficiently the available power and bandwidth of the satellite.

It was found that heavy precipitation conditions which increased the

system temperature substantially did not introduce errors other than those due to attenuation of the signal. For average rainfall conditions, this attenuation would not exceed 0.5 dB. The operating equipment installed at Andover, Maine, is shown in Figure 2.

MEASUREMENT TECHNIQUE

Measurement References

The microwave power received from a distant source can be referenced to the flux measured from celestial sources such as Cassiopeia A [1], [2] by radio astronomers. For routine measurements of received power, it is not convenient to refer constantly to a celestial source. Further, at some locations, a suitable celestial source is visible for only a portion of each day. A suitable secondary standard can be calibrated in terms of the celestial source flux and used as a reference. For this purpose, an argon-vapor discharge lamp, which has a stable output over long periods of time, was used. The noise temperature stability of an argon lamp is better than 0.001 dB/°K [3], [4], [5].

In the case of center-frequency measurements, adequate accuracy is obtained by referring the frequency of a calibration signal to a frequency counter, which may, in turn, be compared with standard frequency radio stations. Similarly, the frequency modulation or deviation may be established by modulating a calibration signal with a known modulating frequency and observing the carrier disappearance on a spectrum analyzer. The measuring device for both center-frequency and deviation measurements is a frequency discriminator of either the phase-locked-loop or conventional variety.

Power Measurement

The relationship between the satellite beam-center e.i.r.p., P_c , and the received signal at the antenna reference point is

$$P_c = \frac{P_s L_p A_c}{G_{ac}} \quad (1)$$

where P_s = received power from satellite at earth station antenna flange, in watts

L_p = path loss from satellite to antenna = $(4\pi d)^2/\lambda^2$

d = range to satellite, in meters

λ = wavelength, in meters

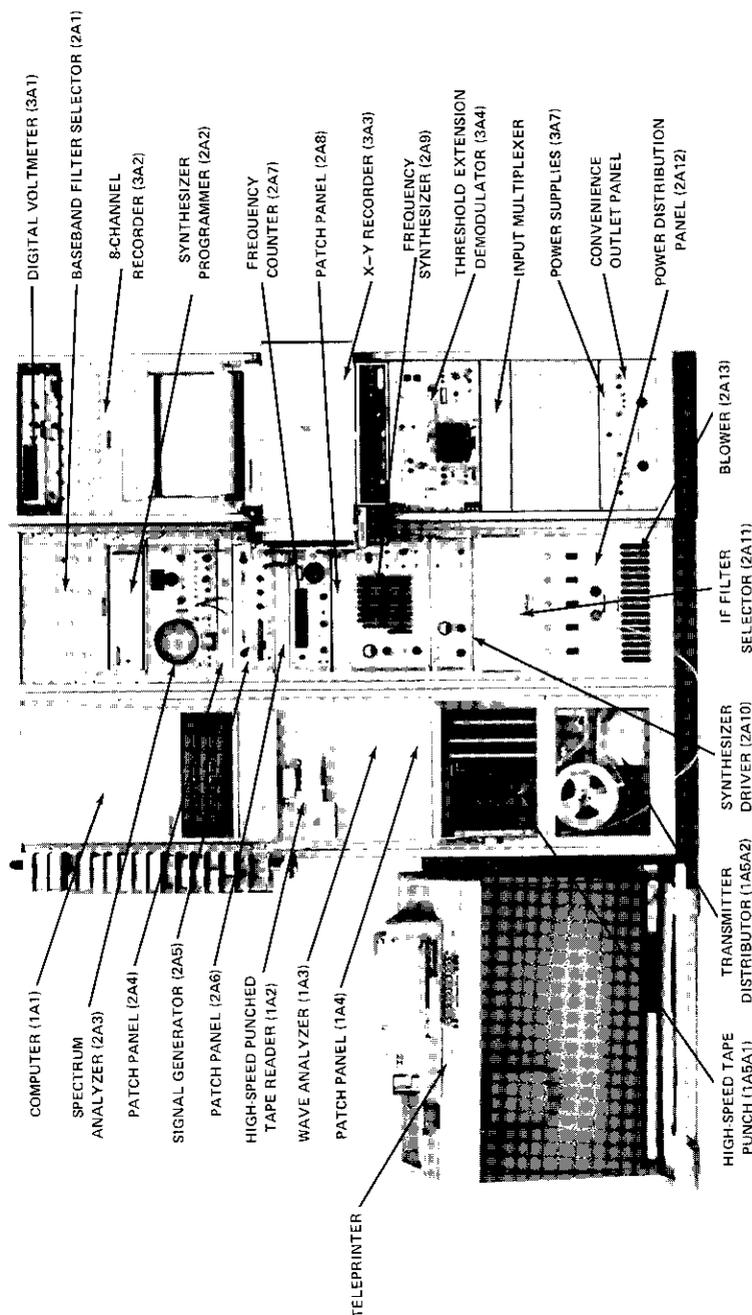


Figure 2. Satellite System Monitor (Andover, Maine, Installation)

A_e = ratio between P_e and e.i.r.p., in the direction to the monitoring station

G_{ac} = antenna gain at measured frequency.

The various parameters are illustrated in Figure 3.

Now, if the received power from a celestial source is used as a calibration reference, the power received at the antenna flange in the bandwidth of interest from such a source is

$$P_r = \frac{G_{ac} C B_i n}{G_m} \quad (2)$$

where G_m = gain of one square meter, given by $4\pi/\lambda^2$, at frequency of interest

C = flux from celestial source, in W/Hz/m², corrected as necessary for the angular extent of the source

n = polarization correction factor (0.5 in the case of a randomly polarized source such as Cassiopeia A)

B_i = bandwidth, in Hz, in which P_r is measured.

Both the path loss and flux from the celestial source should be corrected for atmospheric attenuation. This attenuation varies rapidly with antenna elevation angle, but, if the celestial source calibration is performed at the same elevation that is used for viewing the satellite, the attenuation applies approximately equally to both measurements and may be assumed to cancel under clear weather conditions. Equation (2) may be rewritten as

$$G_{ac} = \frac{8\pi P_r}{C B_i \lambda^2} \quad (3)$$

Substituting this value for G_{ac} in equation (1) gives the satellite beam-center e.i.r.p.:

$$P_e = 2\pi d^2 A_e C B_i \frac{P_s}{P_r} \quad (4)$$

Since it is not practical to make continual direct comparisons between P_s and P_r , a noise lamp is introduced as a secondary reference. For each operating frequency, the power ratio, P_l/P_r , where P_l is the power introduced by the noise lamp in the bandwidth, B_i , at the antenna reference point, is established.

Equation (4) may be rewritten so that the satellite beam-center e.i.r.p. is given by

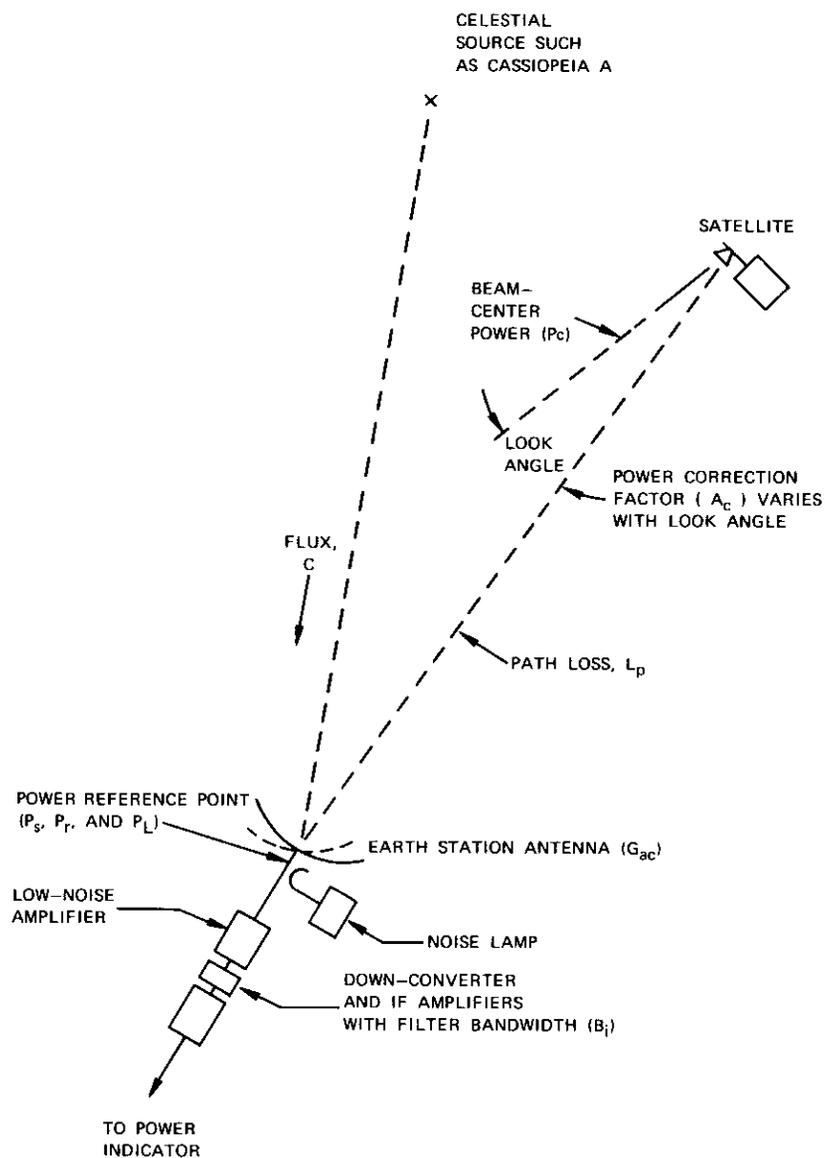


Figure 3. Power-Measurement Parameters

$$P_c = 2\pi d^2 A_r C B_i \frac{P_s P_L}{P_L P_r} \quad (5)$$

The range, d , is determined from satellite orbital calculations. The celestial source flux, C , varies with frequency, and, in the case of Cassiopeia A, decreases slowly with time in a known fashion and must be corrected accordingly. The bandwidth, B_i , which corresponds to the Carson's Rule bandwidth for each carrier, is measured. The ratio P_L/P_r is determined during the primary calibration. The secondary calibration effectively establishes system gain and noise temperature for each frequency. Then the ratio between the received power from the satellite and the lamp noise power, or $P_s P_L$, is determined. The methods of determining these ratios are discussed in the following sections.

Primary Calibration

To obtain accurate power ratios, it is essential that all amplifiers between the antenna and the detecting instrument are operated in the linear portions of their transfer characteristics to avoid signal compression.

Approximately 100 dB of amplification is necessary between the antenna flange and receiver output to operate suitable power-indicating instruments. It is important that this value of gain remain constant between the times of the lamp and star measurements. The primary calibration measurements are made with the same bandwidth and at the same frequency, which means that the system configuration does not change between measurements. Since the lamp noise power and lamp-star power ratio are constant over a long interval of time for a given frequency and bandwidth, the effects of gain variation of the amplifiers and inaccuracies in antenna orientation toward the celestial source may be reduced by taking the mean value of 10 or more measurements. A sample primary calibration is given in Appendix B.

The following three measurements are taken on each frequency for the primary calibration:

$$P_1 = GP_n$$

$$P_{1s} = G(P_n + P_r)$$

$$P_1' = G(P_n + P_L)$$

where G = receiving system gain between antenna reference point and power indicator (see Figure 3)

P_n = background noise power in system bandwidth at reference point

P_r = power from celestial source in system bandwidth at reference point

P_L = noise power from lamp in system bandwidth at reference point

P_1 = power indicator reading with antenna off celestial source and lamp off

P_{1s} = power indicator reading with antenna pointed toward celestial source and lamp off

P_1' = power indicator reading with antenna pointed away from celestial source and lamp on.

From the preceding equations it follows that

$$P_1' - P_1 = G(P_L + P_n - P_n) = GP_L$$

$$P_{1s} - P_1 = G(P_r + P_n - P_n) = GP_r$$

If the gain, G , remains constant during the measurements, the lamp-star ratio is given by

$$\frac{P_L}{P_r} = \frac{P_1' - P_1}{P_{1s} - P_1} \quad (6)$$

Neither the absolute values of the system gain nor the lamp noise power needs to be known to determine the lamp-star ratio.

It may be of interest to determine the receiving system characteristics, such as the antenna figure of merit (G/T) or the actual antenna gain and system noise temperature. The determination of the absolute system temperature requires a knowledge of the noise lamp effective temperature, while the antenna gain value depends upon a knowledge of both the flux from the celestial source and the noise lamp effective temperature.

The effective temperature of the lamp can be accurately determined by comparing the lamp noise with noise from a room temperature load and a calibrated cold load [4]. An argon lamp, operating at 9700°K, coupled to the reference point through a 20-dB directional coupler in an environment of 290°K, will produce an effective temperature of 97°K + 2.9°K or 99.9°K.

If measurements are made in the same bandwidth, the ratio of system noise power to lamp noise power will equal the ratio of the effective temperatures. The power appearing at the antenna reference point will be $kT_{si}B_i$ as a result of the system noise, and $kT_L B_i$ as a result of the lamp noise, where

k = Boltzmann's constant = 1.380×10^{-23} W/Hz/°K

T_{si} = system noise temperature

T_L = effective lamp temperature.

It follows that the temperature ratio may be expressed as

$$\frac{T_{si}}{T_L} = \frac{P_n}{P_L} = \frac{P_1}{P_1' - P_1}$$

If the absolute value of T_L is known, the system temperature is given by

$$T_{si} = \frac{P_1 T_L}{P_1' - P_1} \quad (7)$$

The flux density, C , and the corrections to C for angular extent and atmospheric attenuation for various celestial sources are available from astronomical measurements [1], [2]. Noting that $P_L/B_i = kT_L$ and combining equations (3) and (6), one finds that the antenna gain is given by

$$G_{ac} = \frac{8\pi k T_L (P_{1s} - P_1)}{C \lambda^2 (P_1' - P_1)} \quad (8)$$

The antenna figure of merit may be determined without reference to a noise lamp. It is obtained by combining equations (7) and (8) to give

$$\frac{G_{ac}}{T_{si}} = \frac{8\pi k (P_{1s} - P_1)}{C \lambda^2 P_1} \quad (9)$$

It should be noted that the value of C that is used for evaluating equations (8) or (9) must include the correction factor for atmospheric attenuation. This correction is unnecessary when establishing the primary calibration for equation (5) because it will affect the celestial source power and satellite power equally under clear weather conditions, since they are at the same elevation.

Secondary Calibration

The secondary calibration is performed each time a satellite carrier power is measured; the result is the ratio of P_s/P_L . A radiometric method is used to reduce the effect of receiver gain fluctuations and system noise temperature changes.

Radiometers

Radiometers are used in a variety of configurations. In the switched total-power radiometer described by Dicke [6], [7], the receiver input

is rapidly switched between the antenna and a reference load. Methods of continuous automatic null-balancing in this type of radiometer have been described by Machin, Ryle, and Vonberg [8] and also by Goggin [9]. In establishing the null, these radiometers perform a subtraction function.

A radiometer may also perform a ratio function; such a unit was described by Ohm and Snell [10] for a space communications receiver application. In this system, the receiver was left continuously connected to the antenna and the reference noise was switched on and off. The reference noise was coupled through a directional coupler; this arrangement permitted the receiver to remain in a low-noise configuration. An advantage of this type of radiometer is that tracking and telemetry signals are not interrupted because the antenna is never disconnected. The ratio can be calibrated in terms of satellite e.i.r.p. and is independent of receiver gain changes, but is not independent of system temperature changes.

Frequency-Shift Radiometer

Simply by switching the received frequency, radiometers for spectral line measurements in radio astronomy use the antenna and the system noise as the reference load at the off-line frequency [11], [12]. This type of instrument may be called a frequency-shift radiometer. In the application to be described, receiver gain fluctuations are avoided by taking ratios, and system temperature variation effects are avoided by performing a subtraction.

As a result of design analyses and early field trials by COMSAT, it has been concluded that the frequency-shift type of radiometer best meets the requirements for satellite power measurements. In this application, the receiver is first switched to an unoccupied "guard" frequency between carriers to measure the system noise. Then the lamp is switched on and a second power measurement is made. If the gain and system noise values do not change between these two readings (performed on the same frequency), a simple subtraction yields a value for the lamp noise power alone.

The receiver local oscillator is then switched so that the received power on the desired carrier frequency is measured. It is possible at this point to again switch on the lamp and measure the carrier plus lamp plus system noise power in an attempt to establish the gain at the carrier frequency. However, this procedure leads to error because the satellite power output may change as a result of spin ripple,*

* Spin ripple occurs with INTELSAT II-type satellites which have omnidirectional antennas spinning about a north-south axis.

distant station tracking errors, or power amplifier output fluctuations at the distant station; therefore, it is not used.

The magnitude of the error is illustrated by the following example. If the lamp noise power is 10 percent of the satellite carrier power, and the carrier power decreases by 10 percent between the times when the lamp is on and off, the observed output will be the same with the lamp on and off, and the apparent carrier-to-lamp ratio will become infinite. In practice, it has been found that the error is small if the gain and system noise at the carrier frequency are inferred from measurements at frequencies just above and just below the carrier frequency.

For the ratio-type frequency-shift radiometer, the ratio between the receiver output power indications for the on-carrier (lamp off) and off-carrier (lamp on) conditions is given by

$$R_{sLn} = \frac{G_1(P_s + P_{n1})}{G_2(P_L + P_{n2})} \quad (10)$$

where R_{sLn} = satellite-lamp ratio with system noise included

G_1 = gain at carrier frequency

G_2 = gain at guard frequency

P_{n1} = system noise power at carrier frequency

P_{n2} = system noise power at guard frequency

If the output power values, G_1P_{n1} and G_2P_{n2} , can be established, these can be subtracted from the numerator and denominator of equation (10) to obtain the ratio

$$R_{sL} = \frac{G_1(P_s + P_{n1} - P_{n1})}{G_2(P_L + P_{n2} - P_{n2})} \quad (11)$$

which reduces to

$$R_{sL} = \frac{G_1P_s}{G_2P_L}$$

where R_{sL} is the satellite-to-lamp power ratio. The ratio G_1/G_2 is easily measured, P_s/P_L can be found, and the evaluation of equation (5) can be completed. The output power value, G_2P_{n2} , is measured by simply switching off the lamp. The output power value, G_1P_{n1} , can be determined by moving the antenna away from the satellite, but this is not an acceptable procedure because tracking and telemetry data will be lost. Instead, the value for G_1P_{n1} is established by taking the average of system noise power measurements above and below the carrier frequency. It can be seen that the ratio determined from equation (11) will not change if P_{n1} and P_{n2} increase as a result of precipitation

Because the satellite signal is approximately sinusoidal and the lamp signal is noise, it is essential that the power detector at the receiver output is a true power reading device independent of the signal waveform. Bolometers or thermoelectric-type power meters are suitable, but require some settling time. For this reason, the radiometer switching is much slower than in the Dicke-type radiometer and the receiver is set on each frequency for about 1.5 seconds. Receiver gain fluctuations and system noise changes occur at a much slower rate than 1.5 seconds, and consequently the error introduced is negligible.

The P_s/P_L ratio that is determined by the preceding method is linearly related to increases in P_s , which means that the output power indicator can be interfaced with a computer without introducing nonlinear calibrations.

The lamp and system noise measurements are made in the unoccupied frequency range between carriers. This frequency range is necessarily narrower than the bandwidth occupied by the carrier. Since the value of P_{n_1} is being inferred from the observation of P_{n_2} and since the value of P_L in equation (5) is obtained in the carrier bandwidth, B_c , it is necessary to increase the power readings taken in the guard band by the ratio between the bandwidths which are used for the measurements.

The transmission of a number of carriers through the satellite transponders generates a certain amount of intermodulation power which cannot be separated from carrier-power measurements. If no intermodulation power appears in the guard bands, the values of P_{n_1} that are inferred from observations of the noise power in the guard bands will not include the intermodulation power; therefore, the calculated values of P_c will include the intermodulation power. The true useful value of P_c can be determined by subtracting the intermodulation power, which can be computed from a knowledge of the carrier power values, the satellite frequency assignments, and the transponder amplifier phase and amplitude characteristics.

System Temperature and Power Computation

The following measurements are made to determine the ratio P_s/P_L :

$$\begin{aligned} V_1 &= G_2 P_{n_2} \\ &= \text{system noise at carrier frequency} - \Delta f \\ V_2 &= G_2 P_L + G_2 P_{n_2} \\ &= \text{system noise plus lamp noise at carrier frequency} - \Delta f \\ V_3 &= G_3 P_{n_3} \end{aligned}$$

$$\begin{aligned} &= \text{system noise at carrier frequency} + \Delta f \\ V_4 &= G_3 P_L + G_3 P_{n_3} \\ &= \text{system noise plus lamp noise at carrier frequency} + \Delta f \\ V_5 &= G_1 P_s + G_1 P_{n_1} \\ &= \text{carrier power plus system noise} \end{aligned}$$

Now, if $G_1 = (G_2 + G_3)/2$ and $P_{n_1} = (P_{n_2} + P_{n_3})/2$, then it can be shown that the desired satellite-lamp ratio is given by

$$\frac{P_s}{P_L} = \frac{2V_5}{V_2 - V_1 + V_4 - V_3} - \frac{1}{2} \left[\frac{V_1}{V_2 - V_1} + \frac{V_3}{V_4 - V_3} \right]. \quad (12)$$

The satellite beam-center e.i.r.p. is then determined by combining equations (5), (6), and (12) to obtain

$$\begin{aligned} P_c &= 2\pi d^2 A_c C B_i \left[\frac{P_1' - P_1}{P_{1s} - P_1} \right] \left[\frac{2V_5}{V_2 - V_1 + V_4 - V_3} \right. \\ &\quad \left. - \frac{1}{2} \left(\frac{V_1}{V_2 - V_1} + \frac{V_3}{V_4 - V_3} \right) \right]. \quad (13) \end{aligned}$$

It is not necessary to determine the absolute values of antenna gain, noise lamp temperature, or system temperature to find P_c .

A useful test to ascertain correct operation of the monitoring system equipment is to determine the noise temperature. An excessive noise temperature is an indication of malfunction of the low-noise receiver. If the lamp has been extinguished, noise temperature will appear to be infinite because the denominator of equation (7) will become zero. The exact value is not required for this measurement, and it is adequate to assume that the lamp noise temperature is 100°K if a 20-dB coupler is used. From equation (7), the system temperature is

$$T_{si} = \frac{T_L}{2} \left(\frac{V_1}{V_2 - V_1} + \frac{V_3}{V_4 - V_3} \right). \quad (14)$$

It is convenient to first compute the system temperature for each frequency by using equation (14). If the system temperature exceeds some arbitrary value such as 350°K, the process is discontinued and the computer prints out an error message. Since the system temperature given by equation (14) is the second term inside the bracket in equation (13) multiplied by T_L , it is convenient to store this value in the computer for later use in evaluating P_c . To take care of the T_L multiplication, equation (13) is then rewritten so that the satellite e.i.r.p. is given by

$$P_c = \frac{2\pi d^2 A_c C B_i}{T_L} \left[\frac{P_1' - P_1}{P_{1s} - P_1} \right] \left[\frac{2V_5 T_L}{V_2 - V_1 + V_4 - V_3} - T_{si} \right]. \quad (15)$$

For convenience, the primary calibration and other factors that do not vary between measurements on any given satellite are combined in a calibration factor, M . This factor assumes a different value for each carrier and includes the bandwidth, B_i , which varies with channel capacity. The equation that is used by the computer for the calculation of each carrier e.i.r.p. is

$$P_c = M A_c \left[\frac{2V_5 T_L}{V_2 - V_1 + V_4 - V_3} - T_{si} \right] \quad (16)$$

where A_c is the aspect-angle correction factor which must be included in each computation because it varies with time of day and day of the year.

SYSTEM DESCRIPTION

Block Diagram

Figures 4a, 4b, and 4c are more detailed block diagrams of the RF, IF, and computer sections, respectively. The calibration signal from the noise lamp is connected through a switch to a directional coupler that is between the antenna and the parametric amplifier. Under normal conditions, for stations that are receiving telemetry, the signal-to-noise ratio in the telemetry channel is in the range of 25 to 30 dB; there is no degradation of error rate in the received telemetry if the signal-to-noise ratio is decreased by 3 dB. Likewise, the tracking error is insignificant if the signal-to-noise ratio in the autotrack receiver is decreased by 3 dB. If the broad-band calibration signal from the noise lamp is made equal to the background noise, the signal-to-noise ratios are decreased by 3 dB. The large margins of beacon signal power from the satellite are provided for the case when the satellite is in its initial transfer orbit and lower satellite antenna gains may be available. During such initial periods, the monitoring system is switched off.

As seen in the block diagram, the parametric amplifier output signals are down-converted to an IF of 70 MHz, amplified and filtered, and connected to a power meter and demodulator. The process-control computer controls the noise lamp switch, operates the various switches to set up the proper filters, and also selects the appropriate local oscillator frequency. In addition, the measured data are collected and used to compute the carrier power. The results are displayed on a page printer or in the form of punched 5-level tape to record the data on a strip chart.

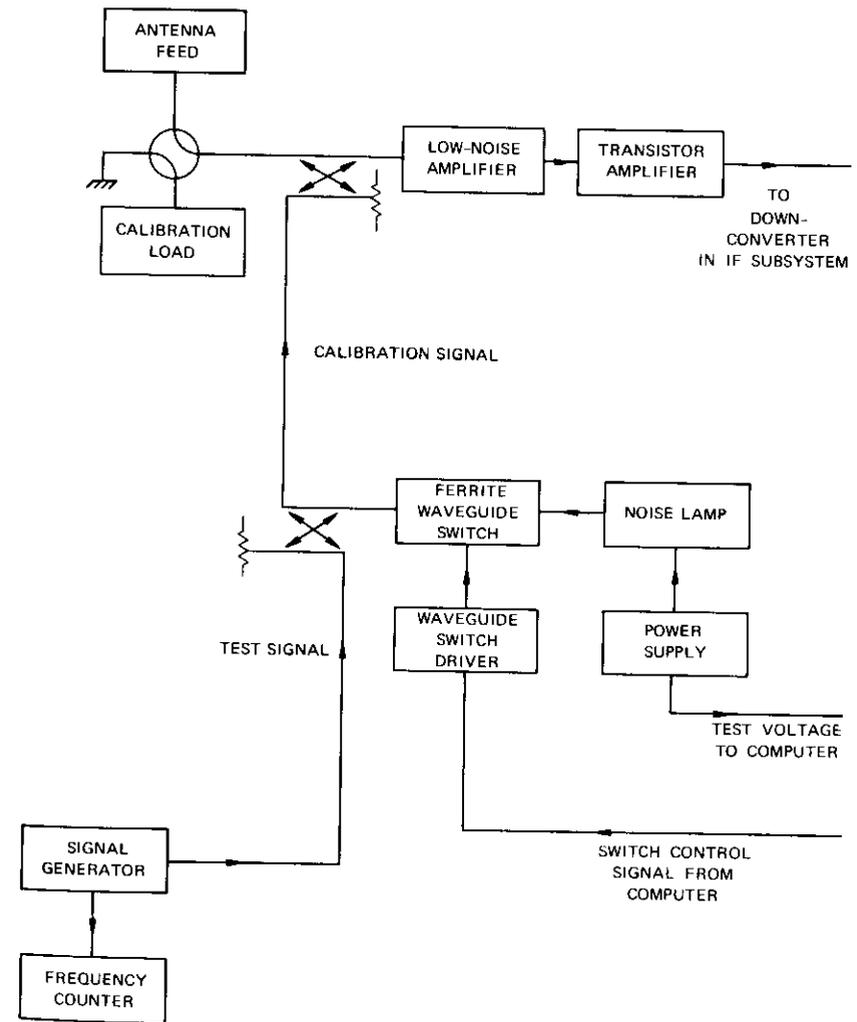


Figure 4a. Satellite System Monitor Block Diagram, RF Subsystem

The receiver carrier frequency or the calibration frequencies are selected by the computer, which adjusts the synthesizer-controlled local oscillator frequency. The output power is measured on a thermoelectric type of power meter which is interfaced with the computer through a digital voltmeter.

The filtered intermediate-frequency signal is also connected to a demodulator; for center-frequency measurements, the digital voltmeter is

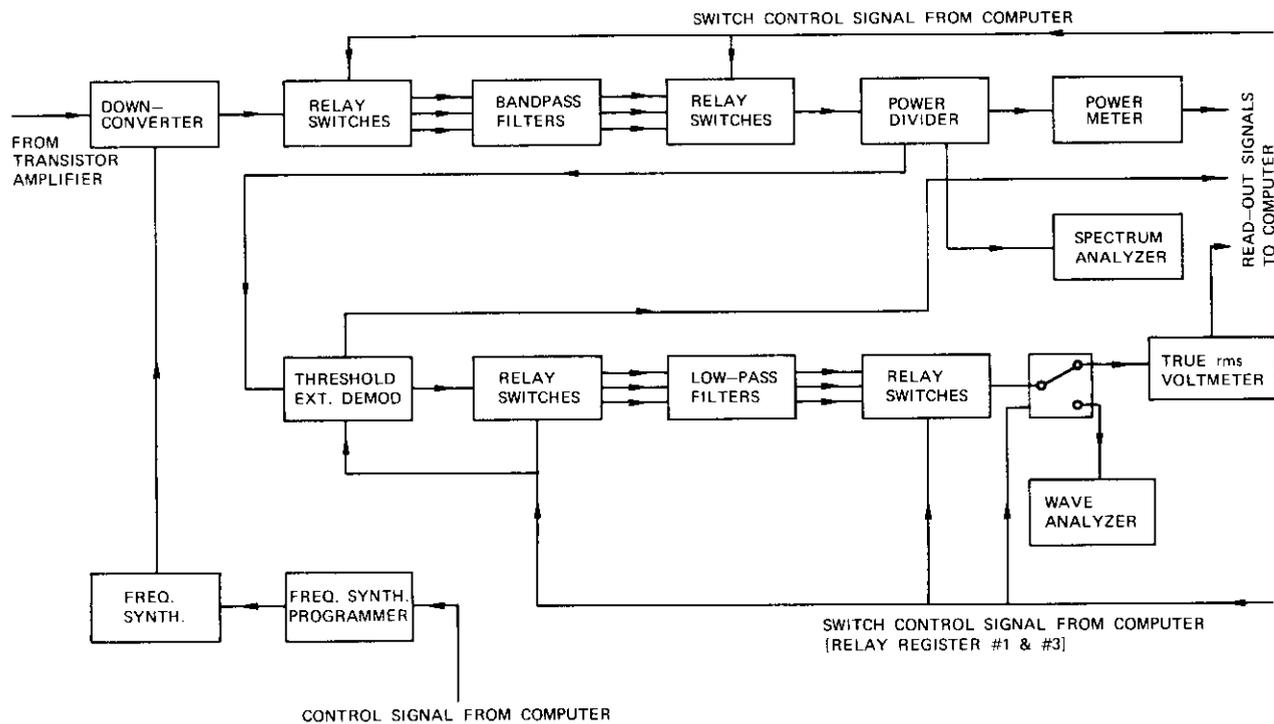


Figure 4b. Satellite System Monitor Block Diagram, IF and Baseband Subsystem

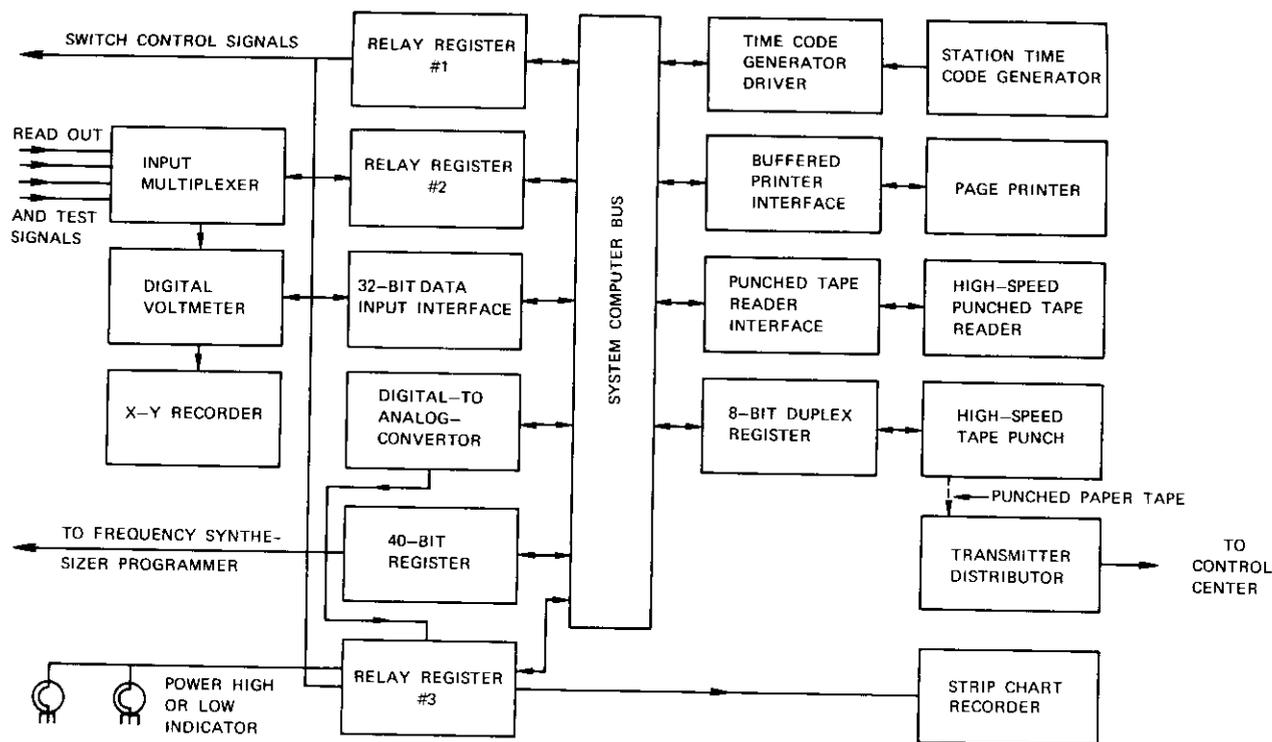


Figure 4c. Satellite System Monitor Block Diagram, Computer Subsystem

connected to the discriminator. The computer selects an appropriate baseband filter for connection to the demodulator output and for deviation measurements; the baseband filter output is connected to a true rms voltmeter. For interface with the computer, the digital voltmeter is reconnected to the rms voltmeter output. The computer solves equation (16) for output power and applies the necessary correction factors to the center-frequency and deviation measurements, so that the printout can be interpreted directly in W (e.i.r.p. at beam center), kHz (off frequency, and MHz (rms deviation).

Computer Software

Punched paper tape programs are fed into the computer with a high-speed optical reader. (The program for two sets of 35 carriers uses approximately 7500 words of memory.) This program consists of an executive program and nine peripheral subroutines. The executive program consists of four sections: initialization, measurements and computations, data output, and a critical voltage scan, as described in the following paragraphs.

Initialization

During the initialization section, which reads the absolute object program into the computer, the operating limits for the program are established, computer memory is set aside to store system data, system constants are initialized, and the relay registers are cleared. Next, the operator messages are printed out and the data tape is read into the computer memory. The data tape contains the local oscillator frequencies, the carrier power maximum and minimum limits, the calibration factor for each carrier, the filter selection data, the maximum and minimum critical-voltage limits, and the aspect-angle correction data. The operator may select the operating mode by means of a series of 15 operator-option sense switches.

Measurement and Computation

The measurement and computation section obtains the output power, center-frequency, and deviation measurements. The system noise and lamp noise measurements are performed only once for each of the guard frequencies, but, since the satellite carriers may be varying, the carrier measurements are repeated five times and an average is taken. In this section, the computer also tests for the sense switch positions and selects the required operating mode. The operating options are shown in Table 1.

TABLE 1. OPERATIONAL OPTIONS

Switch Number	Function
0	Normal mode of operation
1	Off-satellite test using noise lamp as simulator
2	Single-carrier mode
3	Modified printout when using lamp simulation of carriers
4	Operation to prevent stopping for high noise temperature
5	New carrier assignments for 8-channel recorder
6	Baseband monitoring of a single carrier
7	X-Y recording of carrier plus noise of a single carrier
8	Frequency deviation monitoring of a single carrier
9	Center-frequency monitoring of a single carrier
10	Digital voltmeter readings printed
11	High-speed punch inhibited
12	Aspect-angle correction factor calculated using time (when switch is down, a constant factor is inserted).
13	Page printer inhibited
14	Selection of single carrier to be operated. After one carrier has been printed out, the machine will stop and wait for another to be selected, but if 14 is turned down and 2 is left up, the single carrier last selected will be continuously repeated.
15	Orderly shut-down

Data Output

The data output section provides for three types of data output. In the first type, the computed power data are displayed on a strip chart recorder. In addition, a single carrier may be recorded continuously, but, in this case, the value recorded is proportional to V_n , which corresponds to the carrier power plus the system noise power. The second type is the page printer display which prints the computer power, center frequency and deviations, along with carrier numbers and frequencies. The third type which may be selected is the high-speed punch output, which generates 5-level tape for retransmission over teletype lines. In this case, the computed values, carrier numbers and frequencies, together with a header giving the date, time, satellite number, and station number in coded form, are punched. Figure 5 shows a sample tape printout as it would be received at the Operations Center.

Critical Voltage Scan

The computer program performs a system self-test by checking a few system voltages. These are derived from the noise lamp current

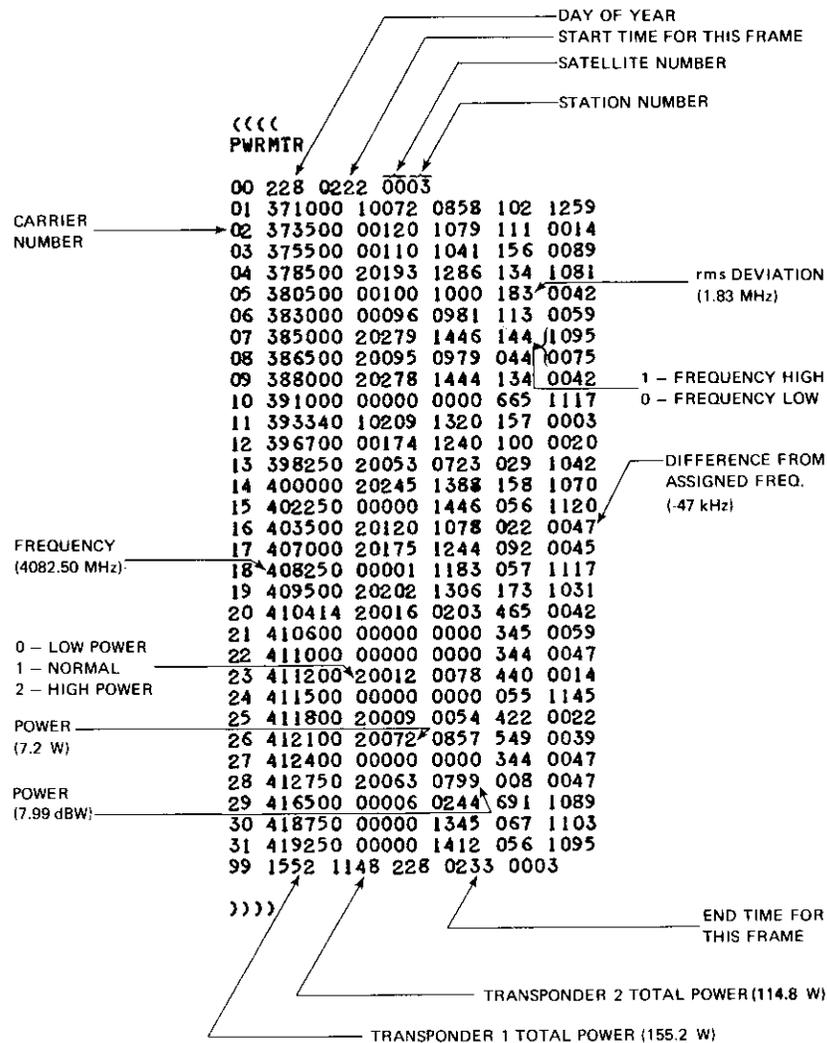


Figure 5. Printout at Control Center

and six test points in the down-converter section and compared with maximum and minimum preset limits. If the limits are exceeded, a high- or low-limit error message is printed on the page printer and the program waits approximately 10 seconds before testing the same voltage again. The computer continues to test the same voltage until it is corrected to within the assigned limits.

After a malfunction which has caused a critical voltage to be out of limits has been corrected, the computer continues to monitor the remainder of the critical voltages. When this process is completed, the computer proceeds to the beginning of a frame of data and measures power, center frequency, and deviation, unless one of the sense switches has been switched to call for a different mode.

SYSTEM TEST AND CALIBRATION

Subsystem Tests

Filter Bandwidth

The noise bandwidth and the response of the intermediate frequency bandpass filters were obtained using the computer and peripheral hardware. These data yielded the values for B_i .

Bandwidth Compensation

As mentioned earlier, noise measurements at the carrier and guard frequencies must be referred to the same bandwidth by taking the bandwidth ratio into account. A convenient method of doing this is to inject a flat noise signal into the system and to adjust the filter attenuators for the same output indication on the power meter.

Down-Converter Performance

The down-converter can be switched in a few microseconds to any frequency in a 500-MHz range. This is accomplished by using a frequency synthesizer capable of rapid switching. The down-converter must be tested for this rapid mobility. Also, the converter must be operated in the linear portion of the transfer characteristic to avoid errors because of signal compression or intermodulation product generation. The noise figure requirements are minimal because adequate preamplification gain is provided.

Interconnecting Transmission Lines

The waveguide and cable connection between the parametric amplifier and the down-converter, which is located a few hundred feet away in an equipment building, must be well matched and terminated to avoid amplitude variations over the operating range. The power-measurement procedure compensates for a smooth gain slope, but closely spaced ripples which will be encountered in long mismatched lines will cause measurement errors. A scanning program, similar to that used for measuring the filter bandwidths, is used to test for overall gain variations. The signal is a noise source, and the output at the end of the

system is measured in a 1-MHz bandwidth at 1-MHz intervals across the band. Mismatches are corrected or isolators are inserted as necessary so that amplitude variations will not cause errors exceeding 0.1 dB in the power measurement.

Deviation Calibration

The rms deviation is measured by observing the demodulator baseband output with a true rms voltmeter. An appropriate baseband filter is connected to the demodulator output to attenuate any noise components at frequencies above the operating baseband. The demodulator is calibrated by the Bessel-zero method.

Center-Frequency Calibration

The demodulator is also used to determine the offset of the center frequency of the modulated carriers. A test signal referenced to a frequency counter is used to calibrate the center-frequency indication.

System Tests

Primary Calibration

The primary calibration is performed using a celestial object for which data are available. The detailed procedure and data for one such calibration are given in Appendix B.

Secondary Calibration Check

As outlined previously, the measurement technique is based upon the assumption that the gain and the system noise at the carrier frequency are equal to the arithmetic mean of the values obtained in guard-band slots at frequencies just above and just below the occupied bandwidth edges. To test the secondary calibration procedure, a program is used which switches the noise lamp on when the system is adjusted to the carrier frequencies. With the antenna aimed at the cold sky, the noise lamp simulates the satellite signal. Since the noise lamp power is equivalent to some value of satellite power, values are calculated for each carrier frequency from the primary calibration. The computer is made to print out the apparent values with the noise lamp on, and the difference between the measured and calculated values represents the error.

Determination of Aspect-Angle Correction Factor

The satellite diurnal aspect-angle variation is assumed to be a sine

function having a period of one sidereal day. The aspect-angle correction factor is given by

$$A_c = U + F_N \sin \frac{2\pi}{24} \left(T - \frac{24D}{365} - F \right) \quad (17)$$

where U = average value of A_c

F_N = amplitude of sinusoidal variation

T = time of day in hours

D = number of day in year

F = time of positive zero axis crossing on January 1.

A time code generator feeds Universal Time into the computer and the program determines the aspect angle for the particular day and hour of measurement. It is important that F_N and F are updated each time that the satellite is maneuvered because the aspect angle will change at these times. The factor U depends only on the latitude of the monitoring station and longitude difference between the monitoring station and the subsatellite point for the particular satellite being monitored.

SATELLITE e.i.r.p. DATA

There are two methods of estimating the experimental error in the power-measuring system. One is to compare the observed data with satellite prelaunch and postlaunch data taken by other means. The second method is to make an estimate based upon the expected errors in determining the factors used in the computations. An error analysis using the latter method is given in Appendix A.

With the INTELSAT III type of satellite, the beacon power is a function of the total input to the satellite. The beacon power drops as the input is increased. For a given value of output power in the communication carriers, there is a certain specific value of beacon power. Prelaunch data are available for single-carrier transfer characteristics at three frequencies in each transponder. No data are available for specific multicarrier frequency plans, but estimates can be made for the multicarrier case. For example, the total multicarrier power will be 1.3 to 1.5 dB less than the single-carrier power at saturation. Also, the beacon suppression will be 2.0 to 2.5 dB less at saturation.

The transfer characteristics for the three single-carrier tests for INTELSAT III, F7, lower transponder, are shown in Figure 6. The dashed curves indicate the estimated multi-carrier characteristics.

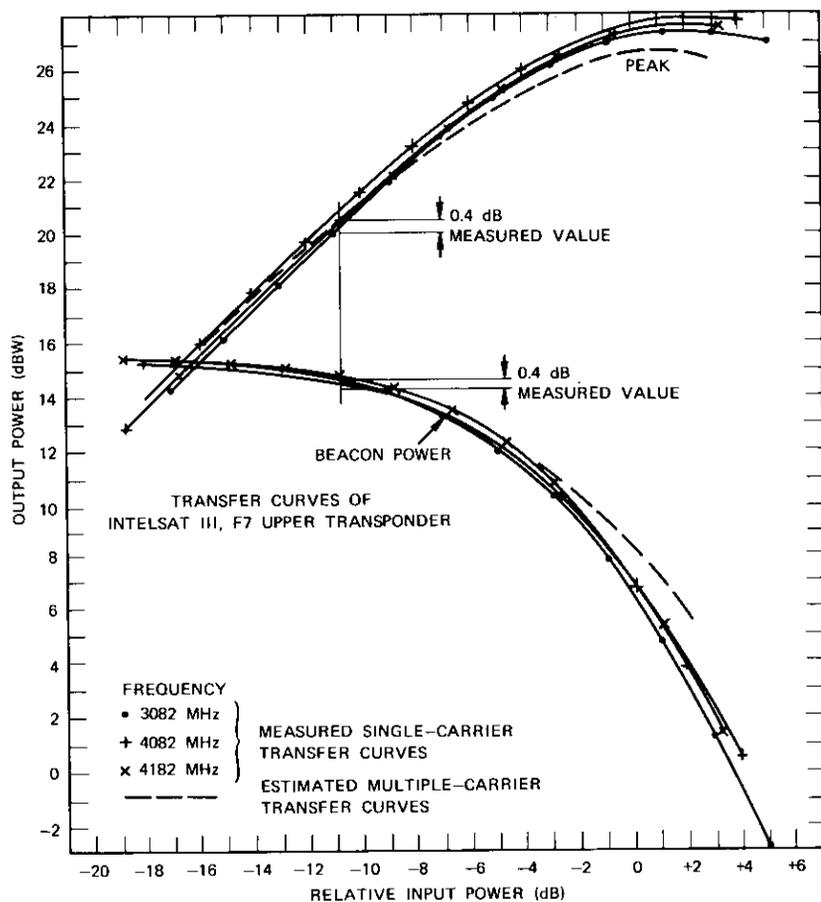


Figure 6. INTELSAT III, F7 Upper Transponder Data

For the test illustrated, the total measured communication power (not including beacon power) corresponded to a relative input power of -11 dB, while the measured beacon power corresponded to an input power of -9.5 dB. The input power can, of course, have only one value at a time, and a vertical line is drawn at a location where both the beacon power and total power increase by the same amount. The error determined by this method was found to be about 0.4 dB. This is clearly an inexact method because of the transfer characteristic variation with frequency. This method also assumes identical performance under prelaunch and postlaunch conditions. This test was repeated for both transponders in F6 and F7; the rms error for all four transponders was -0.64 dB.

In another test, the monitor data and postlaunch data taken with a different earth station antenna observing INTELSAT III, F7 were compared. The saturation power for the single-carrier transfer characteristic was within 0.3 dB and the two sets of transfer characteristics matched very closely.

FIELD EXPERIENCE

The computer-controlled system monitor has been in operation at Andover, Maine, since May 1970. It has been found to be useful for bringing new earth stations into the system and for modifying the transmission parameters to most efficiently use the available power and bandwidth of the satellite. Satellite maneuvers such as repositioning or reorientation can change the range or aspect angle, and carrier power adjustments may be necessary.

Computer programs which permit the power data derived from the monitor to be stored and then plotted on a daily or monthly basis have been written for the COMSAT IBM 360/65 computer. Figure 7 is a plot of the 3805-MHz carrier from INTELSAT III, F7 for August 16, 1970. Figure 8 is a plot of the 3834.2-MHz low-band beacon

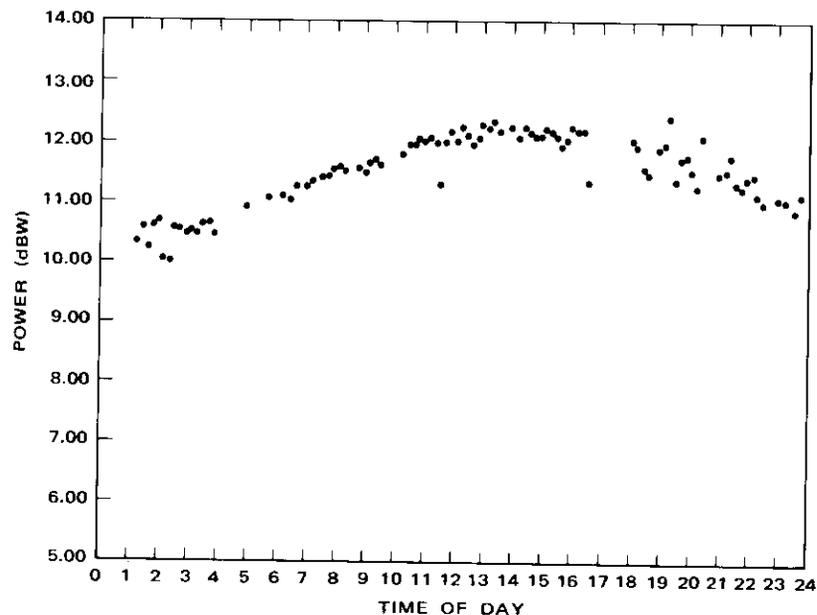


Figure 7. Carrier Variation Over One 24-Hour Period Without Aspect-Angle Correction

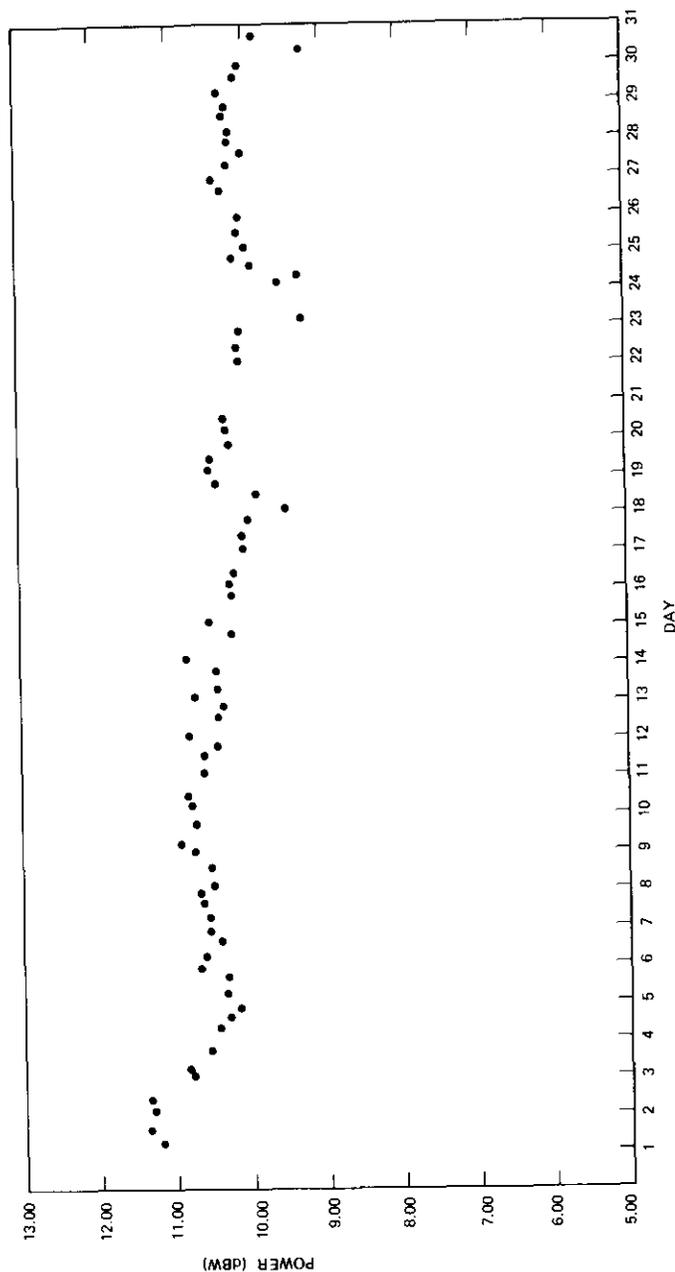


Figure 8. Beacon Variation Over Month of October 1970 Without Aspect-Angle Correction

from INTELSAT III, F6 for the month of October 1970. In both cases, a constant aspect-angle correction was used. It can be seen that the aspect-angle variation for F6 was small, but was about 2 dB for F7.

The test in which the noise lamp was used to simulate received carriers was repeated at intervals over a period of several weeks; the average error for this period was 0.12 dB.

REFERENCES

- [1] A. A. Penzias and R. W. Wilson, "Measurement of the Flux Density of Cas A at 4080 Mc/s," *The Astrophysical Journal*, Vol. 142, No. 3, Oct. 1, 1965, pp. 1149-1155.
- [2] R. W. Kreutel, Jr. and A. O. Pacholder, "The Measurement of Gain and Noise Temperature of a Satellite Communications Earth Station," *Microwave Journal*, Vol. 12, No. 10, Oct. 1969, pp. 61-66.
- [3] W. W. Mumford, "A Broad-band Microwave Noise Source," *Bell System Technical Journal*, Vol. 28, No. 4, Oct. 1949, pp. 608-618.
- [4] A. J. Estlin *et al.*, "Absolute Measurement of Temperatures of Microwave Noise Sources," *IRE Transactions on Instrumentation*, I-9, No. 2, Sept. 1960, pp. 209-213.
- [5] P. Koskos, personal correspondence, Feb. 26, 1970.
- [6] R. H. Dicke, "The Measurement of Thermal Radiation at Microwave Frequencies," *Review of Scientific Instruments*, Vol. 17, No. 7, July 1946, pp. 268-275.
- [7] H. P. Taylor, "The Radiometer Equation," *Microwave Journal*, Vol. 10, No. 6, May 1967, pp. 39-42.
- [8] K. E. Machin, M. Ryle and D. D. Vonberg, "The Design of an Equipment for Measuring Small Radio-frequency Noise Powers," *Proceedings of the IEE (U.K.)*, Vol. 99, Part III, No. 59, May 1952, pp. 127-134.
- [9] W. B. Goggins, Jr., "A Microwave Feedback Radiometer," *IEEE Transactions on Aerospace and Electronic Systems*, AES-3, No. 1, Jan. 1967, pp. 83-90.
- [10] E. A. Ohm and W. W. Snell, "A Radiometer for a Space Communications Receiver," *Bell System Technical Journal*, Vol. 42, No. 5, Sept. 1963, pp. 2047-2080.
- [11] M. E. Tuiri, "Radio Astronomy Receivers," *IEEE Transactions on Military Electronics*, MIL-8, Nos. 3-4, July-Oct. 1964, pp. 264-272.
- [12] I. S. Shklovsky, *Cosmic Radio Waves*, Cambridge, Mass.: Harvard University Press, 1960, pp. 211-216.

APPENDIX A. ERROR ANALYSIS

The experimental error that will occur in any observation of the satellite equivalent isotropic radiated power is a function of many parameters. There will be measurement errors and also errors in the assumed constants, such

as the range to the satellite and the value taken for the flux from celestial sources. These are independent errors and it is suggested that their average value can be determined by taking the square root of the sum of the squares of the mean variation of each independent variable. It is acknowledged that this is not a precise treatment because some errors are random and others are biased, but, for simplification, all errors are treated as independent random variables.

Derivation

The relationship which is used for determination of the satellite e.i.r.p. (P_c) is given in equation (16). When equation (16) is combined with equation (14), the relationship for P_c becomes

$$P_c = MA_c T_L \left[\frac{2V_5}{V_2 - V_1 + V_4 - V_3} - \frac{1}{2} \left(\frac{V_1}{V_2 - V_1} + \frac{V_3}{V_4 - V_3} \right) \right]. \quad (A1)$$

As discussed previously, the experimental error is given by

$$\begin{aligned} \Delta P_c = & \left[\left(\frac{\partial P_c}{\partial V_1} \Delta V_1 \right)^2 + \left(\frac{\partial P_c}{\partial V_2} \Delta V_2 \right)^2 + \left(\frac{\partial P_c}{\partial V_3} \Delta V_3 \right)^2 + \left(\frac{\partial P_c}{\partial V_4} \Delta V_4 \right)^2 \right. \\ & + \left(\frac{\partial P_c}{\partial V_5} \Delta V_5 \right)^2 + \left(\frac{\partial P_c}{\partial M} \Delta M \right)^2 + \left(\frac{\partial P_c}{\partial A_c} \Delta A_c \right)^2 \\ & \left. + \left(\frac{\partial P_c}{\partial T_L} \Delta T_L \right)^2 \right]^{1/2}. \end{aligned} \quad (A2)$$

The partial derivatives of equation (A1) are as follows:

$$\frac{\partial P_c}{\partial V_1} = A_c M T_L \left[- \frac{2V_5}{(V_2 - V_1 + V_4 - V_3)^2} - \frac{V_1}{2(V_2 - V_1)^2} - \frac{1}{2(V_2 - V_1)} \right] \quad (A3)$$

$$\frac{\partial P_c}{\partial V_2} = A_c M T_L \left[- \frac{2V_5}{(V_2 - V_1 + V_4 - V_3)^2} + \frac{V_1}{2(V_2 - V_1)^2} \right] \quad (A4)$$

$$\frac{\partial P_c}{\partial V_3} = A_c M T_L \left[- \frac{2V_5}{(V_2 - V_1 + V_4 - V_3)^2} + \frac{V_3}{2(V_4 - V_3)^2} - \frac{1}{2(V_4 - V_3)} \right] \quad (A5)$$

$$\frac{\partial P_c}{\partial V_4} = A_c M T_L \left[- \frac{2V_5}{(V_2 - V_1 + V_4 - V_3)^2} + \frac{V_3}{2(V_4 - V_3)^2} \right] \quad (A6)$$

$$\frac{\partial P_c}{\partial V_5} = A_c M T_L \left[\frac{2}{(V_2 - V_1 + V_4 - V_3)} \right] \quad (A7)$$

$$\frac{\partial P_c}{\partial M} = A_c T_L \left[\frac{2V_5}{(V_2 - V_1 + V_4 - V_3)} - \frac{V_1}{2(V_2 - V_1)} - \frac{V_3}{2(V_4 - V_3)} \right] \quad (A8)$$

$$\frac{\partial P_c}{\partial A_c} = M T_L \left[\frac{2V_5}{(V_2 - V_1 + V_4 - V_3)} - \frac{V_1}{2(V_2 - V_1)} - \frac{V_3}{2(V_4 - V_3)} \right] \quad (A9)$$

$$\frac{\partial P_c}{\partial T_L} = A_c M \left[\frac{2V_5}{(V_2 - V_1 + V_4 - V_3)} - \frac{V_1}{2(V_2 - V_1)} - \frac{V_3}{2(V_4 - V_3)} \right]. \quad (A10)$$

The calibration factor, M , includes other variables and needs to be analyzed separately. From equation (15), M can be shown to be

$$M = \frac{2\pi d^2 C B_i}{T_L} R \quad (A11)$$

where

$$R = \frac{P_i - P_1}{P_{1s} - P_1}. \quad (A12)$$

It is now possible to calculate ΔM in the same way as ΔP_c ; i.e.,

$$\begin{aligned} \Delta M = & \left[\left(\frac{\partial M}{\partial R} \Delta R \right)^2 + \left(\frac{\partial M}{\partial d} \Delta d \right)^2 + \left(\frac{\partial M}{\partial B_i} \Delta B_i \right)^2 + \left(\frac{\partial M}{\partial C} \Delta C \right)^2 \right. \\ & \left. + \left(\frac{\partial M}{\partial T_L} \Delta T_L \right)^2 \right]^{1/2}. \end{aligned} \quad (A13)$$

The partial derivatives are

$$\frac{\partial M}{\partial R} = \frac{2\pi C B_i d^2}{T_L} \quad (A14)$$

$$\frac{\partial M}{\partial d} = \frac{4\pi C B_i d R}{T_L} \quad (A15)$$

$$\frac{\partial M}{\partial B_i} = \frac{2\pi C d^2 R}{T_L} \quad (A16)$$

$$\frac{\partial M}{\partial C} = \frac{2\pi B_i d^2 R}{T_L} \quad (A17)$$

$$\frac{\partial M}{\partial T_L} = \frac{2\pi C B_i d^2 R}{T_L^2} \quad (A18)$$

Numerical Evaluation

To obtain values for the partial derivatives and variations, a series of measurements was made on a single carrier. One of the available options was to have the computer print out the observed values for each condition of measurement. From these data the standard deviation was obtained, and this was used for the delta terms in each case. For the constants, probable errors were estimated in each case.

For the aspect-angle factor, A_c , the uncertainty was taken as ± 0.25 dB. The maximum change in T_L between primary and secondary calibrations was taken as ± 0.1 dB. The absolute value of T_L is not important because it appears in both the numerator and denominator of equation (A1).

The filter bandwidth, B_i , was found by graphical integration methods with a probability of error of ± 0.012 dB. The range uncertainty was $\pm 0.75 \times 10^5$ meters, which is the difference between apogee and perigee and is not taken into account at each computation.

The nominal value of the carrier that was used for the error analysis was 18 watts and the experimental error derived as discussed earlier was 2.39 watts, which is 0.27 dB. The data were taken with a carrier from INTELSAT III, F6 and hence the accuracies stated reflect actual operational conditions with normal fluctuation of the carrier included.

APPENDIX B. PRIMARY CALIBRATION

A special computer program was used to execute the measurements on P_1 , P'_1 , and P_{1S} during the celestial source calibration. It instructed the operator when to aim the antenna at the cold sky and when to initiate a programmed track on the celestial source. All frequencies that had a bandwidth of approximately 10 MHz were measured with a common IF filter, and the exact center frequency was used for each carrier that was to be measured. A complete frequency run was made with the antenna pointing toward the celestial source, and then the antenna was moved off the source for the lamp-on and lamp-off measurements. The latter were performed successively for each frequency. The process was repeated about 15 times as quickly as possible, so that the measurements would be made at an elevation which was as close as possible to the elevation that would subsequently be used for satellite measurements.

The data were then averaged to get a mean value for each frequency for the three measurements. The lamp-off data were divided into lamp-on and celestial source data. The resulting data were free of variations because of the frequency response of the parametric amplifier and the following amplifiers. The lamp-on and celestial source data now represented the values which appeared at the antenna reference point. Because it was unlikely that the antenna feed and lamp coupler would have sharp discontinuities in a gain versus frequency plot, it was considered legitimate to perform one smoothing on the data.

The procedure used for data smoothing is equivalent to bisecting the third

line of a Pascal's triangle formed around each data point and the points on either side [B1]. Each corrected point is simply derived from $(a_1 + 2a_2 + a_3)/4$, where a_2 is the point to be corrected and a_1 and a_3 are the data points at the next frequency above and below the frequency corresponding to a_2 . The end points cannot be corrected with this method and these points are not changed. A sample primary calibration is given in Table B-1. The original data are given in the first three columns and the normalized data are shown in columns A and C. The result of the smoothing is shown in columns B and D. The lamp-star ratio is then given by $(D-1)/(B-1)$, and is determined for each frequency. To proceed with determination of the calibration factor, M , for each frequency, the value of the flux from the celestial source must be determined for each frequency.

In the case of Cassiopeia A, the flux is decreasing by 1.2 percent per year and the frequency ratio index is 0.76. The value of C for each frequency is determined from

$$C = \frac{4.08}{f} \times 1.047 \times 10^{-23} \times 0.989^D \quad (\text{B1})$$

where f = carrier frequency, in GHz

D = time differential between January 1, 1968, and the date of measurements, in years.

An intermediate calibration factor, M' , is then determined from

$$M' = \frac{2\pi d^2 C}{T_L} R \quad (\text{B2})$$

where R = lamp-star ratio determined previously

d = distance to satellite, in meters

T_L = noise lamp temperature, usually taken to be 100°K

C = star flux, in W/Hz/m², as determined from equation (B1).

Finally, the calibration factor, M , is determined for each frequency by multiplying M' by the bandwidth that will be used for each carrier; i.e.,

$$M = M'B_s \quad (\text{B3})$$

REFERENCE

- [B1] P. H. Dillinger, "Data Smoothing," *The Electronic Engineer*, December 1969, p. 61.

TABLE B-1. PRIMARY CALIBRATION DATA

Frequency	Average of 12 Data Runs			<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>M'</i>
	P_{S+N}	P_N	P_{L+N}	P_{S+N}/P_N	Smoothed	P_{L-N}/P_N	Smoothed	$\times 10^{-8}$
3720	108.9	87.05	173.6	1.250	1.250	1.995	1.995	0.4221
3735	115.1	89.73	184.9	1.283	1.269	2.061	2.025	0.4054
3755	125.6	100.8	201.6	1.273	1.285	2.021	2.050	0.3922
3770	125.0	98.1	198.3	1.295	1.295	2.076	2.074	0.3830
3785	134.8	102.1	217.3	1.319	1.305	2.128	2.100	0.3780
3797.5	134.6	103.0	218.1	1.306	1.311	2.117	2.120	0.3759
3805	127.3	96.4	207.5	1.319	1.314	2.151	2.132	0.3778
3812.5	135.4	101.5	221.3	1.333	1.315	2.181	2.142	0.3783
3817.5	150.8	114.4	246.1	1.317	1.316	2.150	2.148	0.3790
3840	166.7	125.8	273.9	1.323	1.323	2.176	2.174	0.3779
3865	169.2	127.2	279.2	1.330	1.337	2.195	2.198	0.3678
3880	138.8	102.3	226.1	1.356	1.346	2.211	2.202	0.3582
3900	140.6	104.0	231.7	1.351	1.349	2.228	2.222	0.3602
3915	129.7	96.8	216.7	1.339	1.348	2.228	2.238	0.3635
3925	106.7	79.1	177.2	1.348	1.348	2.178	2.252	0.3671
3934	104.5	76.8	177.3	1.358	1.349	2.307	2.265	0.3700
3967	86.6	64.5	146.6	1.342	1.339	2.273	2.296	0.3880
3975	89.9	68.2	152.4	1.319	1.335	2.235	2.305	0.3938
3985	94.6	71.4	162.2	1.331	1.334	2.262	2.317	0.3982
4000	73.1	54.4	128.2	1.343	1.335	2.357	2.333	0.4018
4020	85.2	64.0	148.7	1.331	1.331	2.323	2.332	0.4024
4037.5	91.1	69.1	158.9	1.318	1.327	2.300	2.330	0.4067
4050	90.0	66.9	155.1	1.339	1.332	2.290	2.328	0.3992
4070	85.7	63.6	146.6	1.347	1.340	2.273	2.327	0.3875
4082.5	68.4	50.2	116.9	1.362	1.342	2.330	2.327	0.3857
4095	77.1	59.2	131.3	1.324	1.337	2.283	2.330	0.3906
4115	88.1	66.3	157.5	1.328	1.330	2.375	2.345	0.4022
4135	81.5	61.1	143.8	1.332	1.329	2.352	2.355	0.4048
4165	73.0	54.9	129.7	1.329	1.326	2.362	2.345	0.4042
4187.5	69.8	52.5	122.8	1.329	1.312	2.338	2.305	0.4091
4192.5	69.2	53.5	122.4	1.292	1.292	2.285	2.285	0.4081

ACKNOWLEDGMENTS

The authors wish to acknowledge the assistance of Mr. David Reiser for the experimental work and testing which he performed on this project and Mr. Louis Pollack for his encouragement and suggestions. The support of the Technical Subcommittee of the ICSC and the assistance of Mr. Harry Gross of the INTELSAT Systems Management Office are also gratefully acknowledged.



Ernst E. Steinbrecher received the degree of Diplom Ingenieur from the Institute of Technology in Vienna, Austria. He then worked in a research and development institute of the German Air Force as an engineer in the microwave field. From 1946 to 1953 he headed his own laboratory in Vienna, working in the field of communications techniques. From 1953 to 1966 he held positions at the U.S. Army Signal Corps Engineering Laboratories and several other organizations, where he developed microwave components and explored new means of signal processing. He joined COMSAT in 1966 and is presently a member of the technical staff of the RF Transmission Laboratory.

L. F. Gray was educated at the University of British Columbia where he received a B.A.Sc. (E.E.) degree in 1938. He was with Canadian Marconi until 1943 and then served in the Royal Canadian Navy. From 1945 to 1951 he was with Federal Telephone and Radio and Air Associates, Inc. From 1951 to 1964 he was at ITT Federal Laboratories, where he was concerned with equipment development for television broadcasting and for troposcatter and space communication systems. At present he is manager of the Earth Terminals Branch of the RF Transmission Laboratory at COMSAT Laboratories.



Index: proton irradiation, radiation damage, solar cells

A MODEL FOR NONPENETRATING PROTON DAMAGE TO SILICON SOLAR CELLS

R. A. ARNDT AND L. H. WESTERLUND

ABSTRACT

Those energetic protons that are contained in a satellite radiation environment and are stopped in a solar cell produce non-uniform damage in the cell. As a result, the usual expression, $(1/L^2) = (1/L_0^2) + K_L \phi$, relating the proton fluence, ϕ , to L_0 and L , the undamaged and damaged diffusion lengths in the base region, respectively, by means of a damage coefficient, K_L , is not applicable. Expressions for short-circuit current and open-circuit voltage are developed assuming that the damage results in three uniform layers: a, a layer of moderate damage over most of the proton range, b, a layer of heavier damage occurring at the end of the proton range, and c, a layer sustaining no damage beyond the proton range. Results are presented showing the dependence of short-circuit current and open-circuit voltage on proton energy and fluence.

One conclusion drawn from this model is that a unique value of K_L does not exist for protons below a given energy. A second conclusion is that, of all the protons stopped in a solar cell, the low-energy protons are the most effective in changing the electrical properties because the damage is produced near the junction.

INTRODUCTION

The radiation that a satellite encounters in space is capable of damaging the electrical components of the satellite. This radiation consists mainly of electrons and protons; both of these produce permanent damage in the form of displaced lattice atoms, lattice vacancies, and associated complex defects.

Since solar cells are on the outside of a satellite, they are more susceptible to proton damage than to electron damage. This is because the heavy mass of the proton, relative to that of the electron, results in a shorter range in materials and a greater ability to create displacements. It is possible to protect the solar cells against protons having energies of about 10 MeV or less by means of a transparent cover, but this would require a prohibitively thick cover to shield a cell against all of the protons in the space environment. In order to design effec-

tive covers and radiation-resistant solar cells that minimize both the weight and the proton damage, one needs to know the nature of the proton damage and its effect on the electrical characteristics of the cell.

The present work considers the effects of proton irradiation on only the electrical output of the cell. A major loss in the electrical output of solar cells results from the reduction in short-circuit current which is caused by a radiation-induced reduction of the minority carrier diffusion length in the base region of the cell. (The radiation has little effect on the minority carrier diffusion length in the diffused surface layer, since this is already very short in an undamaged cell.) Relationships have been developed previously which connect the short-circuit current, base region minority carrier diffusion, and radiation fluence when the damage is uniform throughout the base (see, for example, Reference [1]). This case arises in irradiation with particles, such as electrons or high-energy protons, that penetrate the cell. When protons have insufficient energy to penetrate the cell, the damage is not uniform; models have been proposed by Wysocki *et al.* [2] and by Tada [3] to describe the effects of such nonuniform damage.

Although there is straggling (the distribution in range of individual protons) in the range of a beam of monoenergetic protons, the amount of straggling is usually small compared to the range. This means that the protons can be considered to have an abrupt stopping point, and therefore, a cell will suffer no damage in the region beyond the range. The earlier models [2], [3] for low-energy proton damage, which are similar to each other, are based on this feature. For example, according to Tada [3], the base region of a proton irradiated solar cell consists of a uniformly damaged layer up to the range of the protons and an undamaged layer beyond this range. He then computes the open-circuit voltage and short-circuit current as a function of proton energy and fluence and compares his results with those of experiments [4], [5]. Both models fit the experimental data reasonably well for protons of energy down to about 0.5 MeV.

Very near the end of its range, a proton loses nearly all of its remaining energy in elastic collisions that result in displaced atoms. In the present work, a more realistic model is developed which takes into account the comparatively heavy damage that occurs near the end of the proton range for monoenergetic particles. This model, which results in a better fit to experimental results, assumes that the base region of a solar cell is divided into three layers: a. a moderately damaged layer formed by the protons over most of their range; b. a more heavily damaged, thin layer formed near the end of the range; and c. an un-

damaged layer beyond the proton range. Theory relating the short-circuit current, open-circuit voltage, and the fluence and energy of the protons is presented in the following section.

THEORY

Displacement Production

Displaced lattice atoms are produced by elastic collisions between an incident proton and the atoms of the target material which, in the present case, is silicon. For protons of energy less than 10 MeV, these elastic collisions are best described by Rutherford scattering theory (see Appendix A). Using the results of Rutherford theory, one can determine the defect production in a solar cell due to protons of various energies. For a displacement energy of 20 eV, values of N_d/ϕ , the density of displacements per incident proton per square centimeter, are listed in Table A-1 and are plotted in Figure 1 as a function of depth in the cell. The figure represents the maximum displacement density to be expected, since any loss of displaced atoms through recombination with vacancies is not considered. The dashed lines were used to determine the thickness of layer *b*. It can be seen from Figure 1 that, for thicknesses of solar cells commonly in use (200-300 μm), the displacement density resulting from a 10-MeV proton irradiation is very nearly uniform.

Uniform Damage

It has been shown experimentally [6] that, for uniform damage,

$$\frac{1}{L^2} = \frac{1}{L_0^2} + K_t \phi \quad (1)$$

where L_0 and L are the diffusion lengths before and after irradiation, respectively; K_t is the damage coefficient; and ϕ is the fluence. The connection between the diffusion length and the measured short-circuit current and open-circuit voltage will be set forth in the following paragraphs.

The steady-state minority carrier distributions and current-voltage illumination characteristics for p-n junctions were derived long ago [7]. A slightly generalized form of the argument will be presented here to establish a basis for the treatment of the more complicated model of the damaged base region.

The time rate of change of excess minority carrier density, Δn , generated in a p-type base region by monochromatic light absorbed according to Beer's Law, is

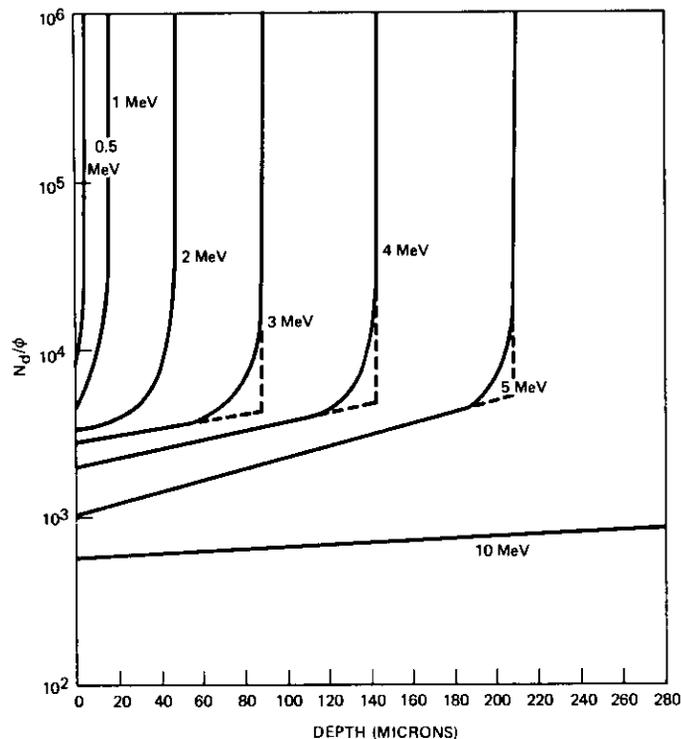


Figure 1. Displacement Density Due to Proton Irradiation as a Function of Depth in a Silicon Solar Cell

$$\frac{\partial(\Delta n)}{\partial t} = D_n \frac{\partial^2(\Delta n)}{\partial x^2} - \frac{\Delta n}{\tau_n} + Q \exp(-\alpha x). \quad (2)$$

In equation (2), D_n is the diffusion coefficient of electrons, x is the distance, τ_n is the low-level electron lifetime, α is the absorption constant, and Q is the incident photon flux neglecting reflection losses. The steady-state solution to equation (2) is

$$\Delta n = n_1 \exp\left(\frac{x}{L_n}\right) + n_2 \exp\left(-\frac{x}{L_n}\right) + \frac{Q\alpha L_n^2}{D_n(1 - \alpha^2 L_n^2)} \exp(-\alpha x) \quad (3)$$

where L_n represents $\sqrt{D_n \tau_n}$. The coefficients n_1 and n_2 are evaluated from the following conditions. At $x = j$,

$$\Delta n = n_0 [\exp(qV/kT) - 1], \quad (4a)$$

at $x = w$,

$$\Delta n = 0, \quad (4b)$$

where j is the depth of the junction below the surface and w is the cell thickness.

The junction itself is assumed to have no thickness, and equation (4b) follows from the assumed ohmic contact over the back surface. In (4a), n_0 is the thermal equilibrium concentration of minority carriers, q is the electronic charge, V is the junction voltage, k is Boltzmann's constant, and T is the absolute temperature. The expressions for n_1 and n_2 are

$$n_1 = \frac{n_0 \left[\exp\left(\frac{qV}{kT}\right) - 1 \right]}{\exp\left(\frac{j}{L_n}\right) - \exp\left(\frac{2w-j}{L_n}\right)} - \frac{\frac{Q\alpha L_n^2}{D_n(1 - \alpha^2 L_n^2)} \left[\exp(-\alpha j) - \exp\left(\frac{w-j - \alpha w L_n}{L_n}\right) \right]}{\exp\left(\frac{j}{L_n}\right) - \exp\left(\frac{2w-j}{L_n}\right)} \quad (5a)$$

and

$$n_2 = -\frac{\exp\left(\frac{2w}{L_n}\right)}{\exp\left(\frac{j}{L_n}\right) - \exp\left(\frac{2w-j}{L_n}\right)} \left\{ n_0 \left[\exp\left(\frac{qV}{kT}\right) - 1 \right] - \frac{Q\alpha L_n^2}{D(1 - \alpha^2 L_n^2)} \left[\exp(-\alpha j) - \exp\left(\frac{w-j - \alpha w L_n}{L_n}\right) \right] \right\} - \frac{Q\alpha L_n^2}{D(1 - \alpha^2 L_n^2)} \exp\left(\frac{w - \alpha w L_n}{L_n}\right). \quad (5b)$$

An equation similar to equation (2) must be solved for the front, n-type region; the steady-state solution is

$$\Delta p = p_1 \exp\left(\frac{x}{L_p}\right) + p_2 \exp\left(-\frac{x}{L_p}\right) + \frac{Q\alpha L_p}{D_p(1 - \alpha^2 L_p^2)} \exp(-\alpha x). \quad (6)$$

The coefficients p_1 and p_2 are obtained from the following boundary conditions. At $x = 0$,

$$D \frac{\partial(\Delta p)}{\partial x} = s(\Delta p) \quad (7a)$$

at $x = j$,

$$\Delta p = p_0 \left[\exp\left(\frac{qV}{kT}\right) - 1 \right]. \quad (7b)$$

These coefficients are given by

$$P_1 = \frac{\frac{D_p}{L_p + s}}{\left(\frac{D_p}{L_p - s}\right) \exp\left(-\frac{j}{L_p}\right) + \left(\frac{D_p}{L_p + s}\right) \exp\left(\frac{j}{L_p}\right)} \cdot \left\{ P_0 \left[\exp\left(\frac{qV}{kT}\right) - 1 \right] - \frac{Q\alpha L_p}{D_p(1 - \alpha^2 L_p^2)} \right. \\ \cdot \left. \left[\exp(-\alpha j) + \frac{\alpha D_p - s}{D_p} \exp\left(-\frac{j}{L_p}\right) \right] \right\} \\ + \frac{Q\alpha L_p(\alpha D_p + s)}{D_p(1 - \alpha^2 L_p^2)} \left(\frac{D_p}{L_p - s}\right) \quad (8a)$$

and

$$P_2 = \frac{\frac{D_p}{L_p - s}}{\left(\frac{D_p}{L_p - s}\right) \exp\left(-\frac{j}{L_p}\right) + \left(\frac{D_p}{L_p + s}\right) \exp\left(\frac{j}{L_p}\right)} \cdot \left\{ P_0 \left[\exp\left(\frac{qV}{kT}\right) - 1 \right] - \frac{Q\alpha L_p}{D_p(1 - \alpha^2 L_p^2)} \right. \\ \cdot \left. \left[\exp(-\alpha j) + \frac{\alpha D_p - s}{D_p} \exp\left(-\frac{j}{L_p}\right) \right] \right\} \quad (8b)$$

where s is the surface recombination velocity.

Equation (7a) must be applied at the front face of the solar cell, since only a very small portion of this face is provided with an ohmic contact. If one neglects that portion covered by the ohmic contact, the only carrier flow out of the cell at the front face is that particle diffusion current which disappears via surface recombination. The total current density in the cell (neglecting recombination in the junction) is given by

$$J = J_n + J_p = -q \left\{ D_n \frac{\partial(\Delta n)}{\partial x} - D_p \frac{\partial(\Delta p)}{\partial x} \right\} \quad x = j \quad (9)$$

where q , the electron charge, does not take into account the sign of the carrier, and the short-circuit current, J_l , is obtained by setting V equal to 0 in the expressions for Δn and Δp .

Since the resulting expression for J_l is unwieldy, the value of J_l as a function of wavelength was computed numerically. The value for Q was taken to be $10^{17} \text{ cm}^{-2} \text{ sec}^{-1}$ and assumed to be independent of wavelength. This is a reasonably good representation of the photon flux from the sun at wavelengths between $0.35 \mu\text{m}$ and $1 \mu\text{m}$. Optical absorption constants were taken from the work of Dash and Newman [8]. Other values employed in the computation were

$$L_n = 0.01 \text{ cm} \\ s = 10^5 \text{ cm sec}^{-1} \\ j = 0.00005 \text{ cm} \\ L_p = 0.0002 \text{ cm} \\ w = 0.02 \text{ cm}$$

While the values for L_n , j , and w are typical of present-day cells, those for s and L_p are not well known and represent rough estimates.

The results of the computation are displayed in Figure 2. It may be noted that, for wavelengths between $0.95 \mu\text{m}$ and the threshold, all of the short-circuit current is generated in the base region. In this case,

$$J_l = J_{ln} = -\frac{qQ\alpha L_n}{1 - \alpha^2 L_n^2} \exp(-\alpha j) \cdot \left\{ + \frac{\exp\left(\frac{j}{L_n}\right) \left[\exp\left(\frac{w}{L_n}\right) - \exp(-\alpha w) \right]}{\sinh\left(\frac{w}{L_n}\right)} + 1 - \alpha L_n \right\} \quad (10)$$

For very low values of α , since j/L_n is small, J_l reduces to

$$J_l = qQ\alpha L_n \left[\frac{1 - \cosh\left(\frac{w}{L_n}\right)}{\sinh\left(\frac{w}{L_n}\right)} \right] \quad (11)$$

where J_l is the short-circuit current that one would get from "uniform" generation of carriers produced by long wavelength light near the photoconductive threshold.

The open-circuit voltage, V_0 , is obtained by solving equation (9) for V under the condition that J equals 0. This results in the following expression:

$$V_0 = \frac{kT}{q} \ln \left(1 + \frac{J_l}{J_s} \right) \quad (12)$$

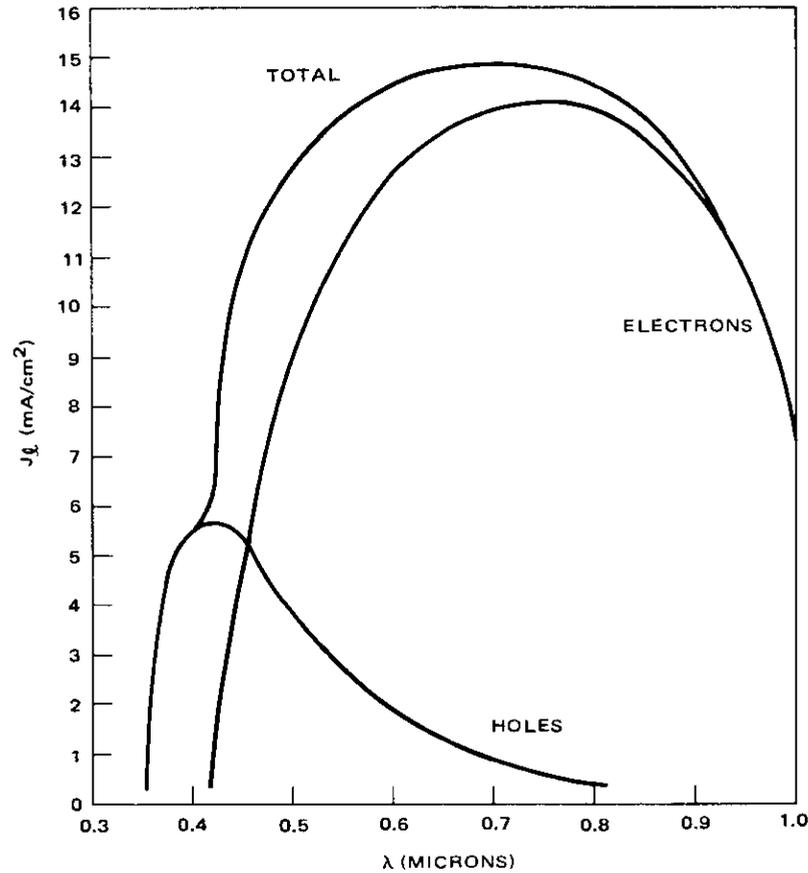


Figure 2. Short-Circuit Current vs. Wavelength of Light (λ) for Holes in Front Diffused Layer and Electrons in Base Region

In equation (12), J_s , the reverse saturation current in the dark, is the sum of all the coefficients of the $[\exp(qV/kT) - 1]$ terms and is given by

$$J_s = J_{sn} + J_{pn}$$

$$= -q \left\{ n_0 \frac{D_n}{L_n} \left[\frac{\exp\left(\frac{j}{L_n}\right) + \exp\left(\frac{2w-j}{L_n}\right)}{\exp\left(\frac{j}{L_n}\right) - \exp\left(\frac{2w-j}{L_n}\right)} \right] \right.$$

$$- \frac{D_p p_0}{L_p} \left[\frac{\left(\frac{D_p}{L_p + s}\right) \exp\left(\frac{j}{L_p}\right) - \left(\frac{D_p}{L_p - s}\right) \exp\left(-\frac{j}{L_p}\right)}{\left(\frac{D_p}{L_p + s}\right) \exp\left(\frac{j}{L_p}\right) + \left(\frac{D_p}{L_p - s}\right) \exp\left(-\frac{j}{L_p}\right)} \right] \right\} \quad (13)$$

All of the necessary relations for determining the short-circuit current and the open-circuit voltage (generated by light in the neighborhood of the threshold) as a function of proton fluence for uniform damage are now at hand. A value of L_n can be determined for a given value of ϕ from equation (1) and known values of the damage coefficient. This value of L_n is then put into equations (10), (12), and (13), to determine the values of J_i , J_s , and V_o . The results are displayed in Figure 3, where the values of J_i and V_o , normalized to the undamaged values, are shown as a function of ϕ . The undamaged value of L_n was taken to be 0.01 cm; the K_L value used was 2.8×10^{-7} , the value appropriate to 10-Mev protons; and the cell thickness was 0.02 cm. The value of the absorption coefficient was taken small enough to result in essentially uniform generation of carriers. It can be seen that the short-circuit current is much more sensitive to damage produced in the base region than is the open-circuit voltage.

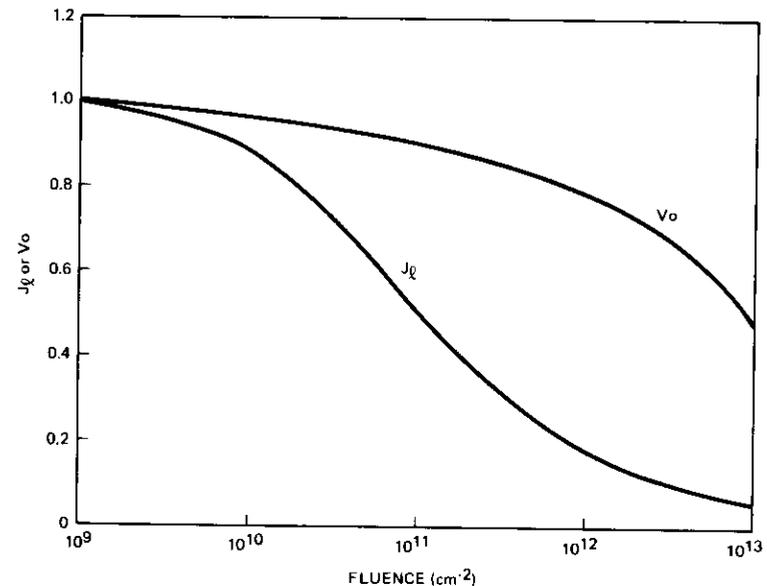


Figure 3. Short-Circuit Current and Open-Circuit Voltage, Normalized to Undamaged Values, vs. Fluence of 10-MeV Protons

Nonuniform Damage

As pointed out in the introduction, a model for low-energy proton damage to silicon solar cells should take into account the heavy damage occurring at the end of range when the proton is stopped in the cell. The proposed model considers only changes in the base region and the damage is assumed to yield three layers, the first two of which are uniformly damaged. In terms of diffusion lengths, the damage profile would be as shown in Figure 4, where, for simplicity, the subscript n has been dropped from the base layer diffusion length.

The expressions for the short-circuit current and open-circuit voltage are obtained similarly to those for uniform damage. A solution of the form of equation (3) is applied to each of the three layers in the base, and a solution of the form of equation (6) is applied to the front diffused layer. The overall solution for the base region is obtained by

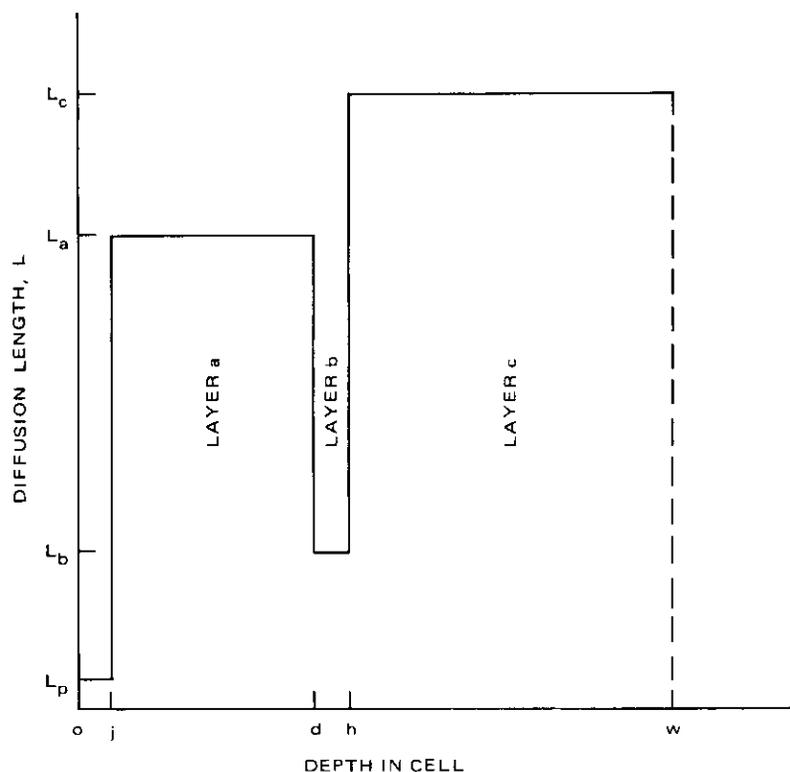


Figure 4. Schematic Representation of Damage in Terms of Minority Carrier Diffusion Length for the 3-Layer Base Region Model

matching the solutions in the three layers according to the following continuity conditions on the excess minority carrier concentration and on its gradient. At $x = d$,

$$\Delta n(a) = \Delta n(b) \quad (14a)$$

at $x = h$,

$$\Delta n(b) = \Delta n(c) \quad (14b)$$

at $x = d$,

$$\frac{\partial[\Delta n(a)]}{\partial x} = \frac{\partial[\Delta n(b)]}{\partial x} \quad (14c)$$

at $x = h$,

$$\frac{\partial[\Delta n(b)]}{\partial x} = \frac{\partial[\Delta n(c)]}{\partial x} \quad (14d)$$

The expressions for short-circuit current and open-circuit voltage in the case of a 3-layer base region are very complicated, and only graphical results will be presented. The diffusion lengths used in obtaining the results were computed from equation (1), where the K_L values were obtained by using Rutherford theory to extrapolate from the experimental value of K_L for 10-MeV protons. Values of K_L for protons of various energies are given in Table A-1. With reference to Figures 1 and 4, the value of h was chosen to be equal to the range of the incident proton, and the value of d was chosen as the intersection of the extrapolated straight line portions of each curve of Figure 1, as shown by the dashed lines on the 3-, 4-, and 5-MeV curves. In most cases, this method yields a value of $h - d$ equal to 0.0002 cm.

In Figure 5, which shows the open-circuit voltage normalized to the undamaged value as a function of fluence for various proton energies, the numbers on the curves represent values of proton energy. The parameters used in obtaining these curves were $D_n = 35 \text{ cm}^2 \text{ sec}^{-1}$, $T = 300^\circ\text{K}$, $p_0 = 7.5 \times 10^{17} \text{ cm}^{-3}$, $n_0 = 6 \times 10^{14} \text{ cm}^{-3}$, $w = 0.02 \text{ cm}$; the light was a wavelength near the long wavelength threshold with an intensity of 20 mW/cm^2 . A gradual drop in the open-circuit voltage with increasing fluence is seen in the figure.

More interesting results are seen in the short-circuit current. The short-circuit current as a function of fluence of 1- and 2-MeV protons is shown in Figures 6 and 7. These curves were computed using the values of the parameters previously given ($L_n = 0.015 \text{ cm}$, $w = 0.05 \text{ cm}$), and the light input was a 2800°K spectrum with an intensity of 70 mW/cm^2 . The uppermost curve in each figure resulted from setting the K_L in layer b equal to K_L in layer a ; i.e., the upper curve represents the lower limit of the damage in layer b . The lowermost curve in each figure was obtained by setting K_L in layer b equal to 5×10^{-4} , which

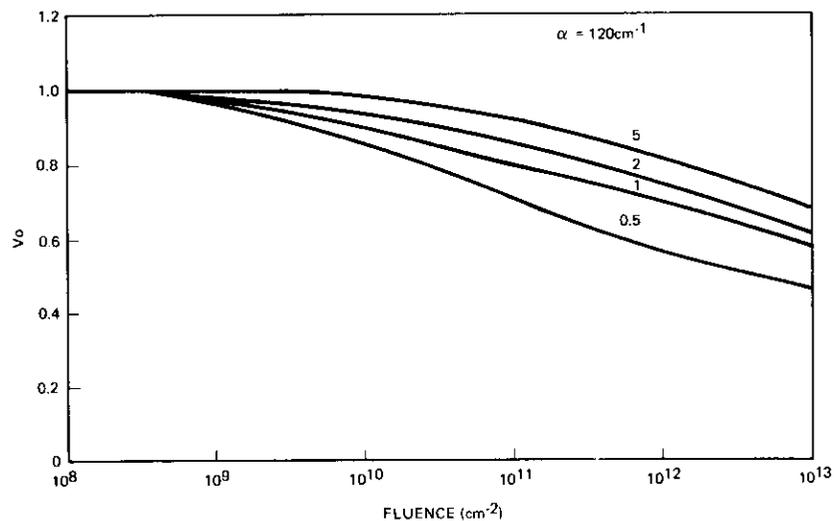


Figure 5. Normalized Open-Circuit Voltage vs. Proton Fluence for $\alpha = 120 \text{ cm}^{-1}$

represents the upper limit of the damage in this layer. It is possible to determine this upper limit because the energy remaining in the proton beam is insufficient to produce additional displacements. Therefore, in Figure 1, the value of 10^6 for N_d/ϕ is the upper limit on displacement density.

The vertical bars in Figures 6 and 7 are the experimental results of Carter and Downing [4]. The middle curves of Figures 6 and 7 show the result of fitting the experimental data by choosing a value for K_L in layer b . For the 1-MeV protons, the best fit was obtained with a value of K_L of 1.45×10^{-5} in layer b . For 2-MeV protons, the best fit to the experimental data was obtained with a K_L value of 3×10^{-5} in layer b .

DISCUSSION

From a comparison of the lowest curve in Figures 6 and 7 with the experimental data, it is obvious that the damage produced at the end of the proton range is not nearly as great as one might have expected from Figure 1. On the other hand, the K_L values for layer b which give the best fit to experimental data for 1- and 2-MeV protons are 7 and 25 times greater than the K_L values for layer a , respectively. That these values of K_L in layer b are about an order of magnitude greater than those in layer a is an indication that the damage at the end of range must be taken into account. The fact that the K_L values in layer b

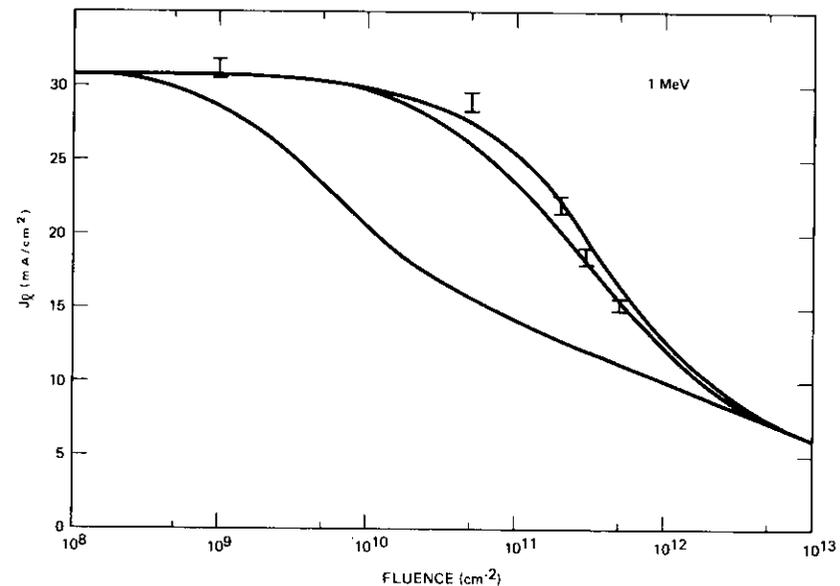


Figure 6. Short-Circuit Current vs. Fluence for 1-MeV Protons

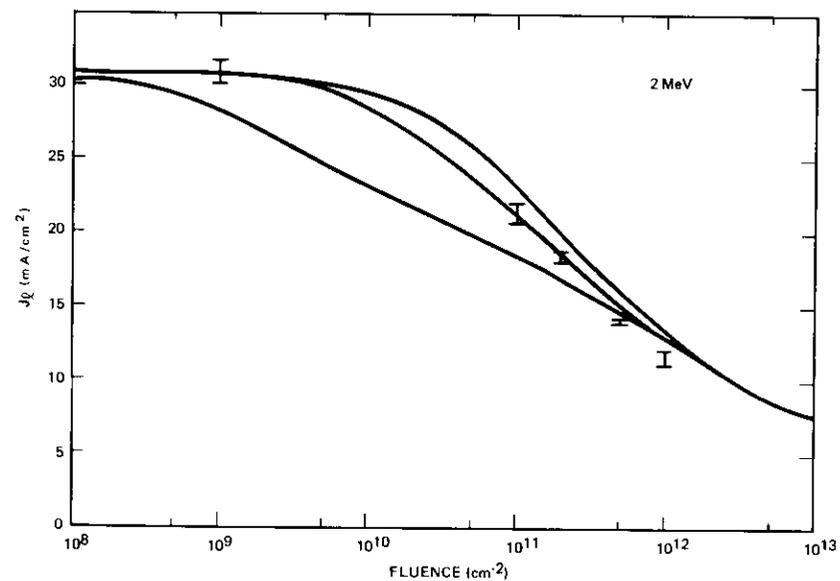


Figure 7. Short-Circuit Current vs. Fluence for 2-MeV Protons

that give the best fit are not the maximum values is probably attributable to recombination of vacancy-interstitial pairs. This "annealing" is likely to be greater here than in the case of defects produced at higher energies because of the closer spacing of the defects.

Another effect of including layer *b* can be seen in Figure 8, where the excess minority carrier distribution is shown in the base of a 0.02-cm thick cell at a fluence of $5 \times 10^{11} \text{ cm}^{-2}$ of 1-MeV protons. In Figure 8, the excess minority carrier density was generated by light near the photoconductive threshold of silicon. The minimum in the curve is due to the carrier "sink" that was produced by the relatively heavy damage at the end of the proton range. No such minimum would occur in the carrier distribution for a 2-layer base.

In Figure 7, an effect which might account for the lack of an even better match to experimental result is the very high displacement density that occurs at high fluence. For example, at a fluence of 10^{12} cm^{-2} , the maximum displacement density produced by 2-MeV protons would be of the order of 10^{15} cm^{-3} . This displacement density is about the same order as the doping level in the base, and one might expect extensive undoping as a result. This could lead to the creation of a junction, a complication not considered in the present work.

It should be noted in Figures 6 and 7 that the curves come together at both low and high fluences. This, coupled with the observation in the numerical computations that, for values of L_b below a certain value, the short-circuit current becomes independent of the diffusion length in the undamaged region, suggests the possibility of an analytic discussion. As will be shown in the following paragraphs, it turns out that a 2-layer base region suffices as a model for this discussion and further simplification can be achieved by assuming "uniform" generation of carriers. Only the short-circuit will be computed. The solution to the diffusion equation is found as before, and equations (14a) and (14c) are applied at the boundary between the two layers to evaluate the coefficients. Solving for the short-circuit current in the base, one obtains

$$J = qQ\alpha L_1 \left[\frac{(A + B) \frac{L_2}{L_1} \exp\left(\frac{d-w}{L_2}\right)}{B \frac{L_1}{L_2} \sinh \frac{d}{L_1} + A \cosh \frac{d}{L_1}} + \frac{B \frac{L_1^2 - L_2^2}{L_1 L_2} - B \frac{L_1}{L_2} \cosh \frac{d}{L_1} - A \sinh \frac{d}{L_1}}{B \frac{L_1}{L_2} \sinh \frac{d}{L_1} + A \cosh \frac{d}{L_1}} \right] \quad (15)$$

where

$$A = \exp\left(\frac{d}{L_2}\right) - \exp\left(\frac{d-2w}{L_2}\right)$$

$$B = \exp\left(\frac{d}{L_2}\right) + \exp\left(\frac{d-2w}{L_2}\right)$$

and the subscripts 1 and 2 refer to the two layers.

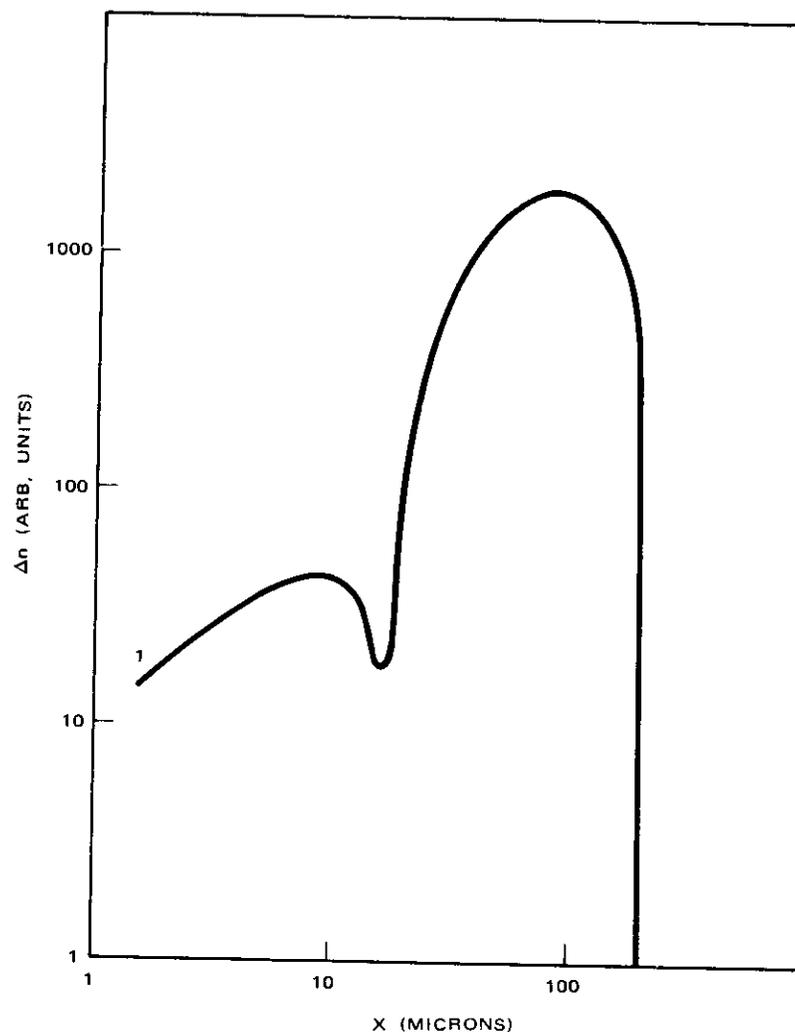


Figure 8. Excess Minority Carrier Distribution in a 0.02-cm Cell for a $5 \times 10^{11} \text{ cm}^{-2}$ Fluence of 1-MeV Protons

Diffusion length profiles are shown in Figure 9 for three proton fluences. At the onset of damage (solid line), the diffusion length in the first two layers is nearly the same as in the undamaged layer and no change in short-circuit current results. The dashed line represents the results of increased, but still low fluence, when the damage increases. This is the case in which, as mentioned previously, the value of the diffusion length in layer *c* has very little effect on the short-circuit current. Finally, the dot-dash line represents the results of high fluence. Because of the difference in the K_L values for layers *a* and *b*, one obtains a limiting case of $L_1 \gg L_2$ in equation (15) where *a* has become 1 and *b* has become 2. Applying this condition to equation (15), after simplification, one obtains

$$J_{in} = qQ\alpha L_1 \left(\frac{1 - \cosh \frac{d}{L_1}}{\sinh \frac{d}{L_1}} \right) \quad (16)$$

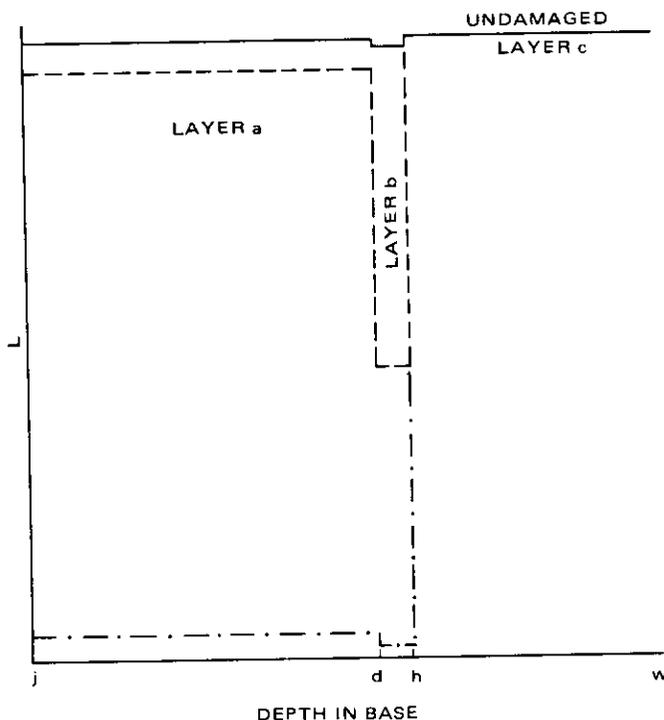


Figure 9. Diffusion Length vs. Depth in the Base Region

This is exactly the same equation as equation (11), except that w has been replaced by d . Hence, for sufficient damage in the layer beyond the junction, the effective thickness of the cell becomes reduced to the proton range.

The divergence at low fluence between the uppermost and lowermost curves of Figures 6 and 7 may be accounted for by a change in effective cell thickness. For a light spectrum of 2800°K, however, the differences in short-circuit current are not as great as for wavelengths near the photoconductive threshold. For very high fluence, the case shown by the dot-dash line in Figure 9 is obtained. This is a limiting case of equation (15) in which $L_1 \ll L_2$, where 1 refers to layers *a* and *b* combined and 2 refers to layer *c*. Applying this condition to equation (15), one obtains

$$J_{in} = qQ\alpha L_1 \left(\frac{\frac{-L_2}{L_1} - \sinh \frac{d}{L_1}}{\cosh \frac{d}{L_1}} \right) \quad (17)$$

For sufficiently large values of d/L_1 , the sinh term in equation (17) dominates and large changes in L_2 are relatively ineffective. If d/L_1 is greater than about 5, the whole term in brackets reduces to -1 , with J_{in} approaching zero as L_1 approaches zero. Thus, the form of Figures 6 and 7 at high fluences may be readily understood.

CONCLUSIONS

The model presented here accounts for two important features of nonpenetrating radiation damage in solar cells, namely the nonuniform damage that is produced and the heavily damaged region occurring at the end of the proton range.

It can be concluded from this model that one should not invoke equation (1) to determine a value of K_L for nonpenetrating proton damage. This is obvious from Figure 1, which shows that such damage is nonuniform. Also, one should not attempt to apply equation (1) by substituting an effective value of K_L , since such a damage coefficient will be fluence dependent as well as energy dependent. Another precaution must be taken when attempting to simulate the proton space spectrum corresponding to synchronous satellite altitudes. Since the fluence of low-energy protons at synchronous altitudes is much greater than that of the higher-energy protons and since the presence of a

heavily damaged region tends to mask the underlying volume, one needs to take great care in simulating the low-energy portion of the spectrum.

The damage resulting from very low proton energies (<0.5 Mev) is highly nonuniform and, therefore, cannot be treated by the present model. Future work might take into account this nonuniform damage by treating a 2-layer base region model in which the diffusion length in the first layer is allowed to vary continuously with distance. For a better understanding of nonpenetrating proton damage, it will also be necessary to correlate displacement defects with carrier-recombination defects and study the high-fluence regime where possible undoping might occur.

REFERENCES

- [1] F. Larin, *Radiation Effects in Semiconductor Devices*, New York: Wiley, 1968, p. 184.
- [2] J. J. Wysocki, J. Loferski and E. Davison, "Low-energy Proton Bombardment of Silicon Cells," *Proceedings of the Fourth Photovoltaic Specialists Conference*, Vol. I: Radiation Effects on Solar Cells and Photovoltaic Devices, NASA Lewis Research Center, Cleveland, Ohio, June 1964, pp. A-3-1—A-3-13 (AD444-113).
- [3] H. Y. Tada, "A Theoretical Model for Low Energy Proton Irradiated Silicon Solar Cells," *Proceedings of the Fifth Photovoltaic Specialists Conference*, Vol. II, NASA Goddard Space Flight Center, Greenbelt, Md., Jan. 1966, pp. D-8-1—D-8-48 (AD477-120).
- [4] J. R. Carter and R. G. Downing, "Charged Particle Radiation Damage in Semiconductors, XI: Effects of Low Energy Protons and High Energy Electrons in Silicon," Report No. 4161-6012-RU-000, Space Technology Laboratories, May 1965 (N65-23290).
- [5] J. M. Denney and R. G. Downing, "Charged Particle Radiation Damage in Semiconductors, IX: Proton Radiation Damage in Silicon Solar Cells," Report No. 8653-6026-KU-000, Space Technology Laboratories, Aug. 1963 (N65-18050).
- [6] K. D. Smith *et al.*, "The Solar Cells and Their Mounting," *Bell System Technical Journal*, Vol. 42, No. 4, Part 3, July 1963, pp. 1765-1816.
- [7] E. S. Rittner, "Electron Processes in Photoconductors," *Photoconductivity Conference*, ed. by R. G. Breckinridge, B. R. Russell and E. E. Hahn, New York: Wiley and London: Chapman and Hall, 1956, pp. 215-268.
- [8] W. C. Dash and R. Newman, "Intrinsic Optical Absorption in Single-crystal Germanium and Silicon at 77°k and 300°k," *Physical Review*, Vol. 99, No. 4, Aug. 15, 1955, pp. 1151-1155.

APPENDIX A ELASTIC COLLISIONS AND DISPLACEMENTS FROM RUTHERFORD SCATTERING THEORY [A1]

Elastic collisions are of the Rutherford type when the incident proton energy, E , is greater than some critical energy E_A , which is defined by

$$E_A = 2E_R \frac{(M_1 + M_2)}{M_2} Z_1 Z_2 \sqrt{Z_1^{2/3} + Z_2^{2/3}} \quad (\text{A1})$$

where E_R is the Rydberg energy ($E_R = 13.6$ eV), M is the atomic mass, Z is the atomic number, and the subscripts 1 and 2 refer to the incident and struck atoms, respectively. For protons incident on silicon, $E_A = 151$ eV.

The maximum energy, T_m , that can be transferred in a Rutherford collision is given by

$$T_m = \frac{4M_1 M_2}{(M_1 + M_2)^2} E \quad (\text{A2})$$

and, for the present case, $T_m = 0.133 E$. The cross section σ_d , for displacing an atom when E_d represents the least energy required to be imparted to the struck atom is given by

$$\sigma_d = 16\pi a_0^2 Z_1^2 Z_2^2 \frac{M_1^2}{(M_1 + M_2)^2} \frac{E_R^2}{T_m^2} \left(\frac{T_m}{E_d} - 1 \right) \quad (\text{A3})$$

where a_0 is the Bohr radius of hydrogen. Values of σ_d for $E_d = 20$ eV are listed in Table A-1. The number of primary displaced atoms per unit volume, n_p , can now be calculated from

$$n_p = \phi n_0 \sigma_d \quad (\text{A4})$$

where ϕ is the bombarding fluence, and n_0 is the atom density of the target. Values of n_p are listed in Table A-1.

The total number of displacements per unit volume is given by

$$N_d = \phi n_0 \sigma_d \bar{\nu} \quad (\text{A5})$$

where $\bar{\nu}$, the mean number of displaced atoms produced by both the incident particle and by the struck atoms of the lattice, is given by

$$\bar{\nu} = \frac{G}{2E_i} \quad (\text{A6})$$

In equation (A6), G is either the average energy transferred, \bar{T} , or an ionization threshold energy E_i , whichever is smaller. Expressions for \bar{T} and E_i are

ENERGY, E (MeV)	RANGE IN SILICON (MICRONS)	Tm (eV)	σ_d (cm ⁻²)	$\frac{n_p}{\phi}$ (cm ⁻¹)	T (eV)	$\bar{\nu}$	$\frac{N_d}{\phi}$ (cm ⁻¹)	K _L
0.001		1.33 x 10 ²	1.74 x 10 ⁻¹⁷	8.70 x 10 ⁵	49	1.22	1.06 x 10 ⁶	5.00 x 10 ⁻⁴
0.002		2.66 x 10 ²	9.40 x 10 ⁻¹⁸	4.70 x 10 ⁵	56	1.40	6.58 x 10 ⁵	3.26 x 10 ⁻⁴
0.005		6.65 x 10 ²	4.07 x 10 ⁻¹⁸	2.04 x 10 ⁵	73	1.82	3.71 x 10 ⁵	1.79 x 10 ⁻⁴
0.01		1.33 x 10 ³	2.05 x 10 ⁻¹⁸	1.03 x 10 ⁵	85	2.13	2.19 x 10 ⁵	1.02 x 10 ⁻⁴
0.02		2.66 x 10 ³	1.02 x 10 ⁻¹⁸	5.10 x 10 ⁴	98	2.45	1.25 x 10 ⁵	5.90 x 10 ⁻⁵
0.05	0.77	6.65 x 10 ³	4.07 x 10 ⁻¹⁹	2.04 x 10 ⁴	116	2.90	5.92 x 10 ⁴	2.76 x 10 ⁻⁵
0.1	1.22	1.33 x 10 ⁴	2.05 x 10 ⁻¹⁹	1.03 x 10 ⁴	130	3.25	3.35 x 10 ⁴	1.61 x 10 ⁻⁵
0.2	2.26	2.66 x 10 ⁴	1.02 x 10 ⁻¹⁹	5.10 x 10 ³	144	3.60	1.84 x 10 ⁴	8.55 x 10 ⁻⁶
0.5	6.44	6.65 x 10 ⁴	4.07 x 10 ⁻²⁰	2.04 x 10 ³	162	4.05	8.26 x 10 ³	3.85 x 10 ⁻⁶
1	16.5	1.33 x 10 ⁵	2.05 x 10 ⁻²⁰	1.03 x 10 ³	176	4.40	4.53 x 10 ³	2.07 x 10 ⁻⁶
2	47	2.66 x 10 ⁵	1.02 x 10 ⁻²⁰	5.15 x 10 ²	190	4.75	2.45 x 10 ³	1.16 x 10 ⁻⁶
3	90	3.99 x 10 ⁵	6.84 x 10 ⁻²¹	3.42 x 10 ²	198	4.95	1.69 x 10 ³	7.97 x 10 ⁻⁷
4	145	5.32 x 10 ⁵	5.13 x 10 ⁻²¹	2.56 x 10 ²	203	5.08	1.30 x 10 ³	6.13 x 10 ⁻⁷
5	210	6.65 x 10 ⁵	4.07 x 10 ⁻²¹	2.04 x 10 ²	208	5.20	1.06 x 10 ³	5.00 x 10 ⁻⁷
10	693	1.33 x 10 ⁶	1.02 x 10 ⁻²¹	1.03 x 10 ²	222	5.55	5.72 x 10 ²	2.80 x 10 ⁻⁷

Table A-1. Displacement Densities and Damage Coefficients Computed from Rutherford Scattering Theory

where m is the electronic mass.

Values of $\bar{\nu}$ are listed in Table A-1, and, for protons, E_i is 1 keV. It is now possible to calculate N_d ; values for this parameter are in Table A-1.

REFERENCE

[1] G. J. Dienes and G. H. Vineyard, *Radiation Effects in Solids*. New York: Interscience Publishers, 1951, Chapter 2.

ACKNOWLEDGEMENTS

The authors thank Dr. E. S. Rittner and Dr. R. W. Rostrom for many helpful discussions. The assistance of C. S. Dudney in performing the computer calculations is also greatly appreciated.



Dr. L. H. Westerlund is currently a member of the technical staff of the Systems Lab at COMSAT Laboratories. The work leading to this paper was performed while he was a member of the Space Physics Branch of the Physics Lab. Before joining COMSAT, Dr. Westerlund was a research associate at Rice University, where he earned a B.A. in Physics (1964) and a Ph.D. in Space Science (1968).

Dr. Richard A. Arndt has been a member of the technical staff of the Physics Lab since coming to COMSAT Laboratories two years ago. After receiving his Ph.D. from Northwestern University in 1961, he was a staff scientist at the Brookhaven National Laboratory for eight years.



ANALYSIS OF ADJACENT CHANNEL INTERFERENCE IN A MULTICARRIER FM COMMUNICATIONS SYSTEM

MARVIN WACHS

ABSTRACT

In a bandwidth-limited system, the total telephone channel capacity is directly related to the bandwidth required per carrier. This study analyzes the adjacent channel interference effects to determine the minimum bandwidth in a multichannel, telephony, frequency modulation, transmission system. Results indicate that carrier spacings slightly greater than Carson's Rule bandwidth are usable in high-quality systems. Comparison indicates close agreement between theory and practice.

INTRODUCTION AND BACKGROUND

Unlike an amplitude-modulated carrier, a wide-band angle-modulated carrier occupies a theoretically infinite transmission bandwidth. In a typical satellite communications system, many frequency-modulated carriers, separated in frequency, are relayed via a common path. The use of proper filtering in such a system permits minimum carrier separation, and, hence, maximum channel capacity. The specification of the filtering function is a compromise between interference from the adjacent carriers and in-band distortion. An analysis of these degradations will be presented.

Two distinct noise mechanisms are investigated. In a system employing the proper filtering and adjacent carrier frequency separation, the amplitude of the desired carrier (at the input to the demodulator) will be larger than the interference for a great majority of the time. During these periods the noise in the desired channel can be obtained in closed form by convolution of the spectra of the desired carriers with the interfering ones. Infrequently, however, the interfering amplitude will become greater than that of the desired carrier. During these intervals an impulsive noise appears which is not easily examined in the classical frequency domain.

The major portion of this study is devoted to analyzing this latter noise mechanism.

The degradation due to a small adjacent carrier has been examined in the pertinent literature. The representation of interference as a small phasor rotating about the head of a large (desired) carrier phasor is a common starting point for most approaches to the problem. The results for simple sinusoidal modulation may be derived by the straightforward application of trigonometry [1],[2]. While "Monte Carlo" analysis might be used to extend the approach to noise modulation [3], the computational complexity imposes a practical limitation. The lack of provision for the inclusion of RF filtering is a further shortcoming of this simplified approach.

Analysis for frequency-division-multiplexed baseband modulation has also appeared. A rigorous mathematical expansion [4],[5] yields a closed-form equation for the calculation of the signal-to-distortion ratio due to a small modulated interfering carrier. As the interference increases, however, the expression takes the form of an extremely complex series expansion. The experimental data presented indicate less than satisfactory agreement for large interference cases.

A less rigorous approach [6] utilizing quasi-static methods has also appeared. However, similar to the preceding analysis, this analysis is inadequate for the case of carrier-to-interference ratios of less than unity and RF filtering.

The purpose of the present analysis is to provide a practical means of predicting interference effects as experienced in present generation communications satellites. Though only two carriers are considered, the analysis is equally applicable to the n-carrier case, since the modulations of the individual carriers are uncorrelated.

The effect of thermal noise is not considered here. For systems operating above the threshold, it appears that interference and thermal noise are independent to a first-order approximation and can be added on a power basis.

DESCRIPTION OF FM COMMUNICATIONS SYSTEM

The frequency-modulated communications link under consideration is illustrated in Figure 1a. The individual voice frequency channels are single-sideband, frequency-division-multiplexed to form the composite baseband. As the number of multiplexed channels increases, the baseband amplitude statistics approach a Gaussian distribution and may be

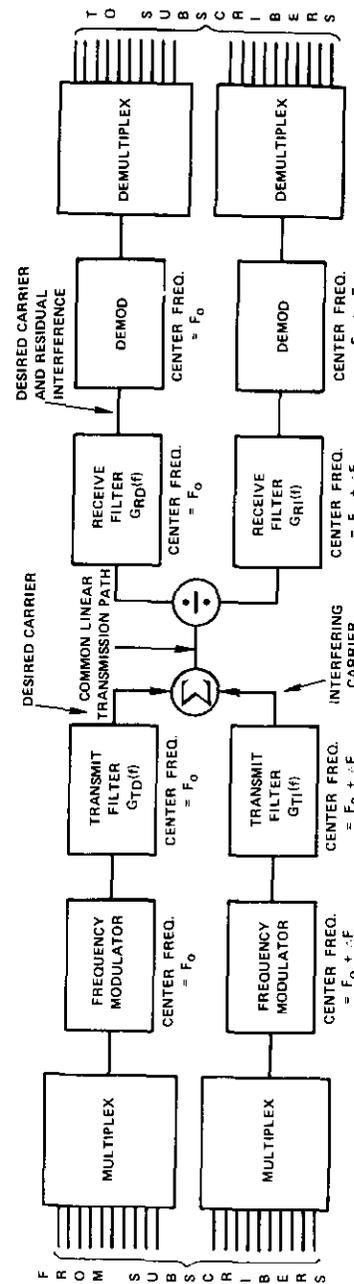


Figure 1a. Typical Communications Link

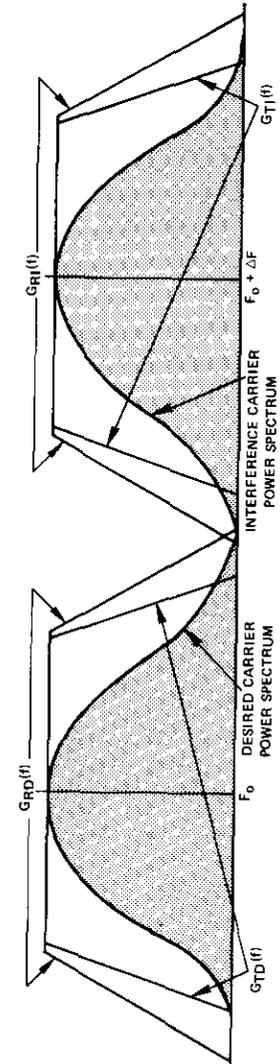


Figure 1b. Schematic Frequency Plane

simulated by band-limited thermal noise. The modulation index is controlled by adjusting the rms baseband voltage applied to the modulator and is defined as the full-load deviation (rms) divided by the maximum baseband frequency. Both the modulator and demodulator have linear voltage frequency characteristics. As schematically diagrammed in Figure 1b, the transmit and receive bandpass filters* are centered on the carrier frequency. The two sets of filters shown (i.e., transmit and receive for both carriers) are independently required. While the transmit filter affords protection to the adjacent carrier, the receive filter provides protection to the received carrier from the adjacent interference. Neither filter is expendable in an efficiently designed system. The output of the demodulator is the composite baseband. The baseband is synchronously demultiplexed, and finally each channel is individually filtered by a bandpass filter (0.3-3 kHz for telephony).

IN-BAND NOISE

The bandpass filtering required for the control of adjacent channel noise degrades the transmitted carrier. For passive filtering, the degradation is due to two causes: group delay and truncation of signal sidebands. Group-delay distortion is due to nonlinearities in the filter passband phase response. This effect can be reduced by inclusion of an additional network to equalize (linearize) the phase in the bandwidth of significant signal components and will be neglected in this discussion. Noise due to truncation, on the other hand, results from a combination of amplitude and phase effects. At frequencies near the corners and on the skirt of the filter amplitude response, the amplitude and phase of the sidebands are seriously perturbed by the filter.

As a first approximation, assume a rectangular bandpass filter, perfectly phase-equalized (group delay constant) in the passband. The passband width is small compared with the center frequency. The sidebands falling outside the passband are infinitely attenuated, while those within are unperturbed. The results of a Monte Carlo analysis [3] may be applied to obtain quantitative results. An algorithm was prepared to compute the sidebands of an FM carrier. A multichannel telephony baseband was simulated by 10 equal sine waves of random phase spaced equally over the baseband bandwidth. The modulation was recomputed from the resulting truncated spectrum and compared with the original

* All filters are group-delay-equalized within the equal-ripple bandwidth and are assumed to have symmetrical amplitude response about their center frequencies. This is valid when the carrier frequency is much greater than the bandwidth.

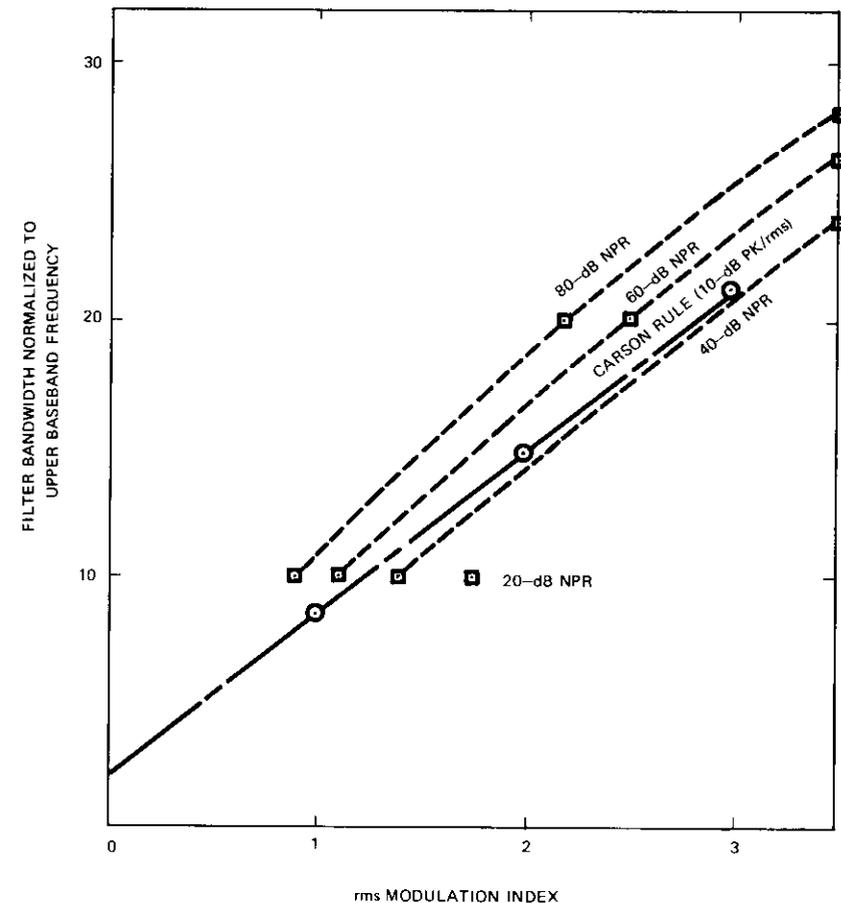


Figure 2. Truncation Noise as a Function of Filter Bandwidth and Modulation Index

modulating waveform. The signal-to-distortion ratio thus calculated is presented as a function of β and filter bandwidth in Figure 2. This simplified analysis leads to the conclusion that filters having a passband equal to Carson's Rule bandwidth* are usable for high-quality transmission.

* $CRBW = 2(f_m + P \cdot \Delta f_{rms})$, where f_m is the highest baseband frequency, Δf_{rms} is the root-mean-square deviation, and P is a peak factor relating the peak-to-rms deviation (assumed to be 10 dB).

IMPULSE NOISE

To examine the nature of the impulse mechanism, the carrier is represented by a phasor $V_D(t)$ with an amplitude A_D and phase $\theta_D(t)$; similarly, the interference is a phasor $V_I(t)$ of amplitude A_I and angle $\theta_I(t)$. The sum of the two, $V_S(t)$, is illustrated in Figure 3.

If the reference axis is normalized to the instantaneous frequency of the desired carrier, $V_I(t)$ will rotate about the head of $V_D(t)$ at the instantaneous difference frequency. When $A_I \ll A_D$, $V_S(t)$ will approximate $V_D(t)$. As A_I approaches A_D , however, $V_S(t)$ will deviate from $V_D(t)$ and may even encircle the origin [for $A_I > A_D$]. Such an encirclement appears as a click in the demodulated baseband. Two possible phase excursions are illustrated in Figure 4. The demodulation of these waveforms is given in Figure 5.

Thus, the total energy (i.e., the energy demodulated in the baseband) for the encirclement ($A_I > A_D$) is 2π ; however, when $A_D > A_I$, the energy approaches zero. The rotational frequency of the vector during these periods is greater than the highest baseband frequency. The baseband output at any channel therefore is essentially the response of the channel filter to a unit impulse.

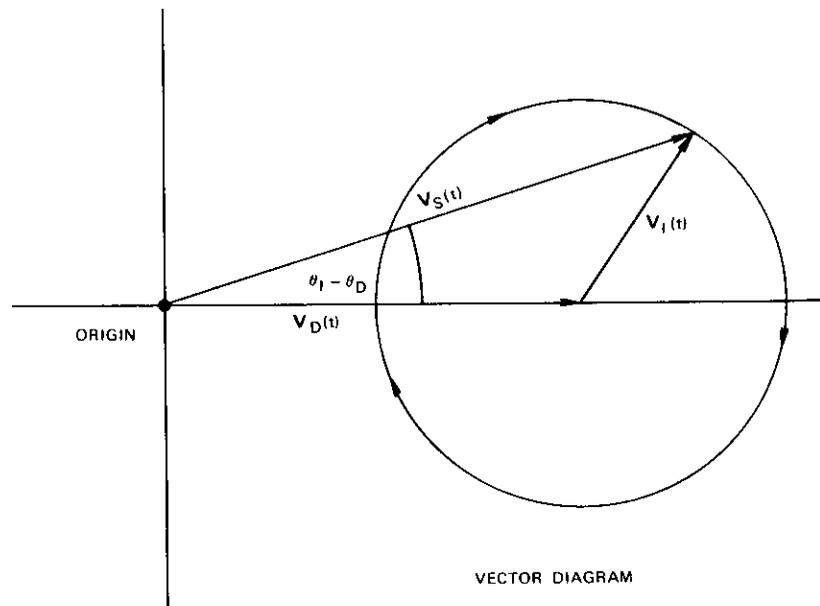


Figure 3. Diagram Illustrating Phasor Definitions

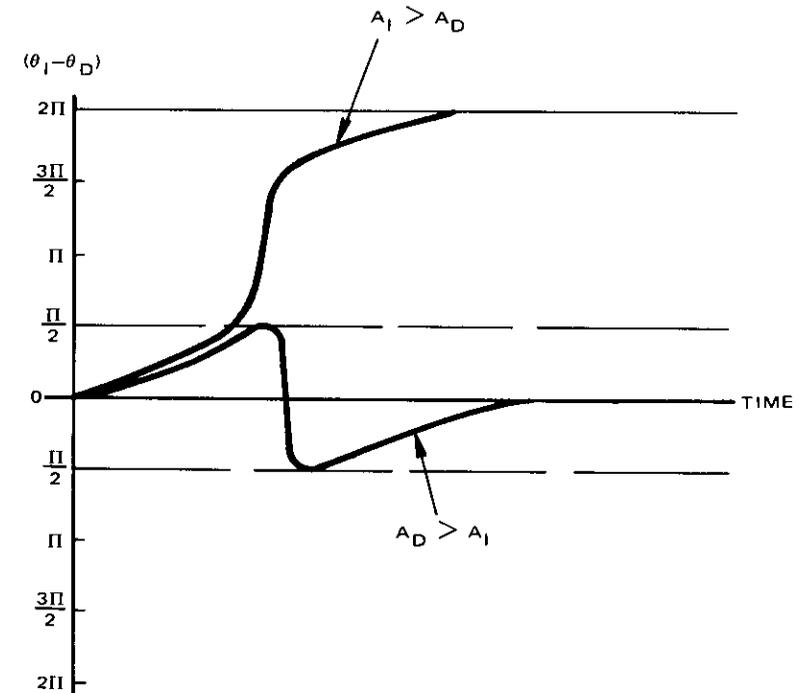


Figure 4. Typical Phase Excursion

The necessary and sufficient conditions [7] for the formation of a click may be stated as

- a. $A_I > A_D$ [amplitude condition], and
- b. $\theta_D(t) + \pi - \epsilon < \theta_I(t) < \theta_D(t) + \pi + \epsilon$ [phase condition].

These conditions implicitly assume that

$$\frac{d}{dt} [\theta_I(t) - \theta_D(t)] \neq 0$$

simultaneously with conditions a and b.

Quasi-Static Approach

Many analyses of frequency modulation have been performed in the frequency domain;* a less rigorous though useful approach is attractive

* The frequency domain representation of a frequency-modulated carrier consists of constant-amplitude sidebands at the modulating frequency(ies) and its higher-order terms.

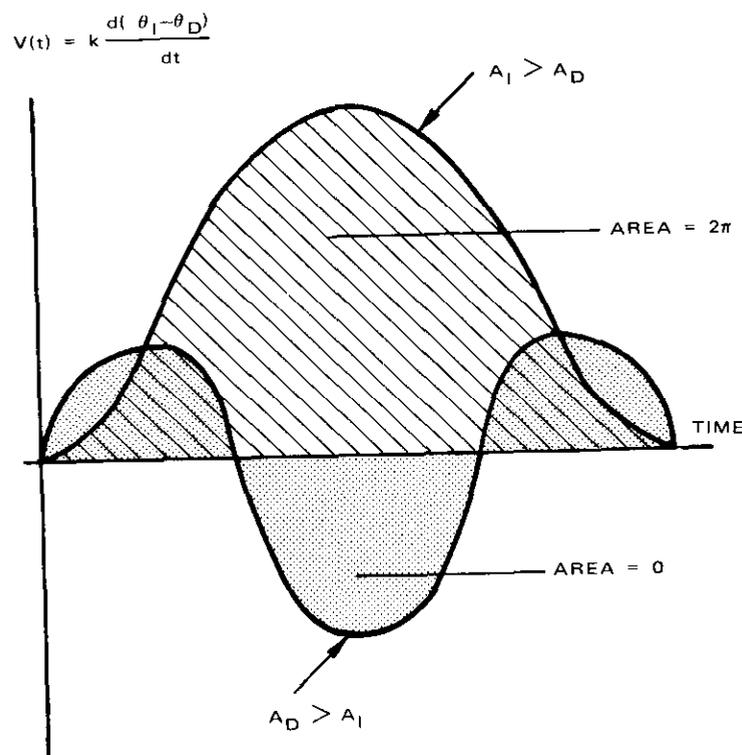


Figure 5. Demodulated Waveforms

for interference impulse studies. Unlike analyses in the frequency domain, the quasi-static representation of a frequency-modulated carrier is a single spectral line of constant amplitude but varying frequency. The instantaneous frequency is a linear mapping of the modulation waveform; the amplitude of the line is the carrier power.

The limits on the region of validity of the quasi-static approach must be carefully examined. One indicator of validity is a comparison of the power spectrum of the modulated carrier with the modulation process probability density. The spectrum, as practically measured, is a graph of the average power (intercepted by a narrow sweeping bandpass filter) as a function of frequency. Since the carrier amplitude is constant, the spectral shape is due to variation of the time spent by the carrier within a given bandwidth increment. As the bandwidth increment decreases, the spectral density approaches the probability density of the instantaneous

frequency of the carrier. It is therefore expected that the power spectrum should closely correspond to the probability density of the modulation.

The baseband modulating an FM carrier (of FDM signals) is commonly represented, for analytic purposes, as flat band-limited noise having Gaussian statistics to 5σ (14-dB pk/rms). The RF spectral shape therefore should be normal. Comparison of measured spectral density with the expected Gaussian shape indicates agreement beyond the -60-dB (power density) points for rms indices of 1.5 or greater [8]. As the index is further reduced, the spectral shape begins to deviate from normal; the unity index case is illustrated in Figure 6. The accuracy of the quasi-static approach, for engineering purposes, begins to degrade for rms modulation indices of one or less.

Simplified Analysis

With the fundamental click generation mechanism described, a simplified analysis may be performed. Assume that the desired carrier is unmodulated (CW), that transmit filters $G_{T_i}(f)$ and $G_{T_D}(f)$ are unity-gain all-pass networks, and that $G_{R_i}(f)$ has a trapezoidal response as diagrammed in Figure 7. The interference carrier is modulated by band-limited noise.

For the great majority of time, the interfering carrier is outside the passband of the receive filter. Occasionally, however, the carrier will deviate in toward the passband of the filter. Depending on the ratio A_i to A_D , the interference level may increase until the impulse condition is satisfied. A typical locus of the resultant phasor, $V_s(t)$, for this occurrence is illustrated in Figure 8.

As long as the instantaneous frequency of the interference carrier is outside the equal-ripple passband, the loci of the head of phasor $V_s(t)$ are negligibly small, varying-diameter spirals about the head of $V_D(t)$. A safe region, where the impulse condition cannot be satisfied, is defined as the interior of a circle of radius A_D about the head of V_D . When the interfering carrier deviates in toward the passband (as shown by b,c, Figure 8), the level increases (as the filter attenuation decreases) and the angular velocity decreases (as the instantaneous difference frequency decreases). The peak of the deviation (c,d, Figure 8) may result in an excursion into the unsafe region; the angular velocity at this point has decreased to a minimum. As the interfering carrier deviates back down the filter skirt (d,e, Figure 8), the amplitude decreases into the safe region, and the angular velocity again increases. The described frequency excursion has not resulted in an impulse, since the origin was not encircled. Path f-i in Figure 8, however, would result in an impulse

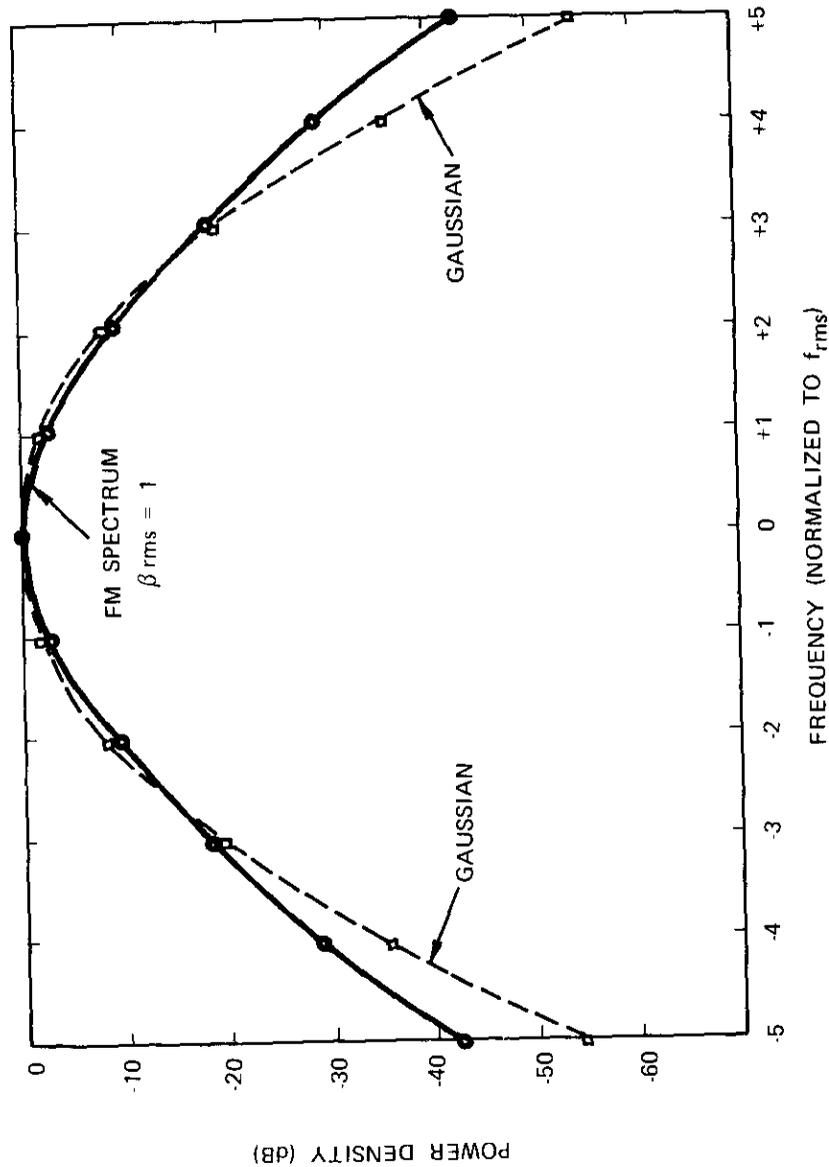


Figure 6. Comparison of RF Spectrum and Gaussian Curve

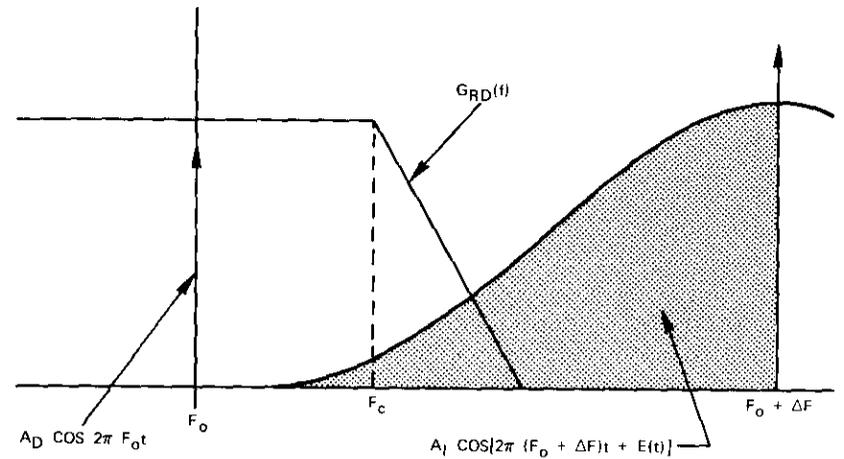


Figure 7. Simplified Frequency Plane Illustrating Adjacent Channel Interference

The rate and probability of these excursions into the unsafe region are calculated first. The probability density of the instantaneous frequency is given (for Gaussian modulation) by

$$p(f_i) = \frac{1}{\sqrt{2\pi}f_{rms}} \exp\left[-\frac{f_i^2}{2f_{rms}^2}\right]$$

where f_{rms} = the rms frequency deviation

f_i = the difference between the center and instantaneous frequencies of the interfering carrier.

The cumulative probability of an excursion into the unsafe region is therefore

$$P[A_I(t) > A_D(t)] = \frac{1}{\sqrt{2\pi}f_{rms}} \int_{|F_c - F_0 - \Delta F|}^{|2F_c - F_0 - \Delta F|} \exp\left[-\frac{f_i^2}{2f_{rms}^2}\right] df_i$$

where $A_I(t)$ and $A_D(t)$ are functions of time after filtering.

The rate of this excursion is obtained from a level-crossing analysis described in Appendix A. For completeness, the result is presented here for a flat baseband.

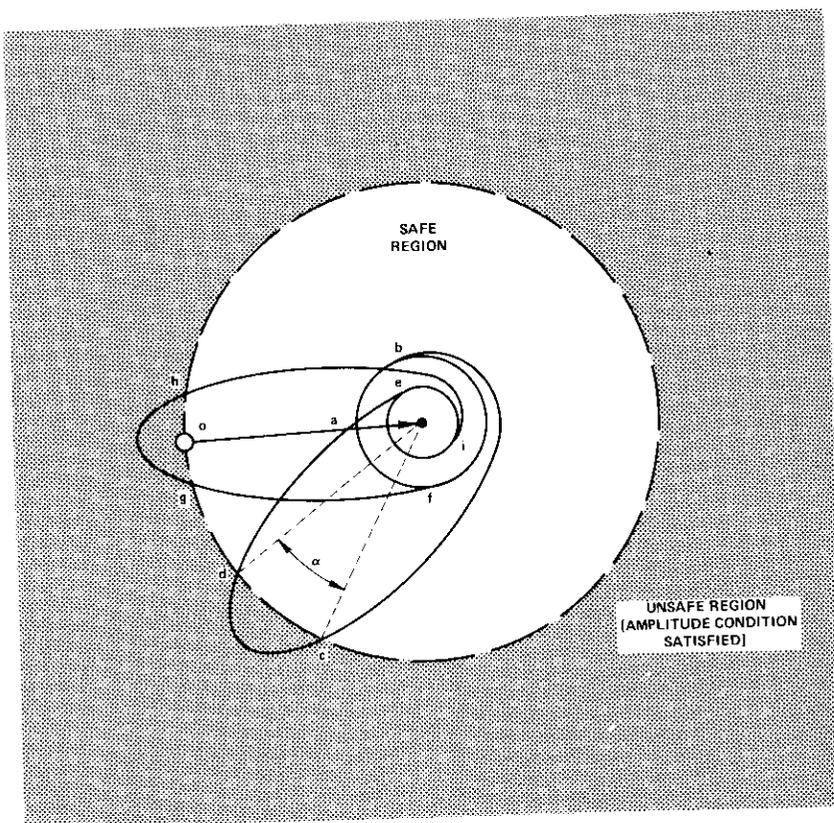


Figure 8. Phasor Loci

$$R = \sqrt{\frac{1}{3}} \cdot f_m \cdot \exp \left[-\frac{f_D^2}{2f_{rms}^2} \right]$$

where f_m = the highest baseband frequency

f_D = the specified frequency deviation from center frequency (level).

Thus, the rate of excursion into the unsafe region is

$$R(A_I > A_D) = \sqrt{\frac{1}{3}} f_m \exp \left[-\frac{(F_c - F_0 - \Delta F)^2}{2f_{rms}^2} \right]$$

where A_I and A_D are functions of time. The average time duration of an excursion is

$$D(A_I > A_D) = \frac{P(A_I > A_D)}{R(A_I > A_D)}$$

The phase condition for impulse is considered next. Due to the uncorrelated nature of the phasor phases (θ_I and θ_D), the probability density of the angular sector, α (Figure 8), is uniform between $-\pi$ and $+\pi$. Thus, the conditional probability of encirclement (given that the amplitude condition is satisfied between points d and c) is

$$P \left[\frac{\text{impulse}}{A_I > A_D} \right] = \frac{\alpha}{2\pi}$$

Angle α is dependent on the angular velocity of $V_s(t)$ and its duration in the unsafe region. For the "equal" carrier case (i.e., interference slightly greater than desired), the amplitude condition cannot be satisfied until the interference has reached the flat-filter passband. The upper limit on the phasor angular velocity in the unsafe region therefore is one-half of the flat-filter bandwidth. Angle α is determined as

$$\alpha = D(A_I > A_D) \cdot 2\pi |F_c - F_0|$$

and, therefore,

$$P \left[\frac{\text{impulse}}{A_I > A_D} \right] \leq D(A_I > A_D) \cdot |F_c - F_0|$$

The amplitude and phase conditions combined yield the desired impulse statistics. For example,

$$\begin{aligned} \text{rate of impulse} &= R(A_I > A_D) \cdot P \left[\frac{\text{impulse}}{A_I > A_D} \right] \\ &= \bar{F} \cdot P(A_I > A_D) < |F_c - F_0| \cdot P(A_I > A_D) \end{aligned}$$

where \bar{F} is the average rotational frequency, and the inequality results from the rotational rate variation within the filter passband.

Extension of Analysis

As an extension of this simplified analysis, assume that the desired carrier is also modulated by a flat baseband having Gaussian statistics. On peak deviations, the desired carrier will be attenuated by the filter skirts [i.e., $G_{TD}(f)$ and $G_{RD}(f)$]. This reduction in amplitude may be viewed as a decrease in the radius of the "safe region" boundary of Figure 8. The frequency-dependent amplitudes are thus

$$A_I(f) = |G_{RD}(f)| \cdot A_I$$

for interference, and

$$A_D(f) = |G_{RD}(f)| \cdot A_D$$

for desired carrier, where $G_{RD}(f)$ is the frequency response of the receive filter, and A_I and A_D are the CW peak voltages (time- and frequency-independent).

Both $A_I(f)$ and $A_D(f)$ are frequency-dependent functions of time, and, therefore, the computation of the amplitude probability is more complex than for the CW case. For each of the possible amplitudes of A_I , the probability that A_D is less must be calculated. The amplitude probability is given by

$$P(A_I > A_D) = \sum_{\text{all levels}} \left[\begin{array}{l} \text{probability that the interference is} \\ \text{at a given power level} \end{array} \right] \cdot \left[\begin{array}{l} \text{probability that the desired carrier is less} \\ \text{than that level} \end{array} \right]$$

$$P(A_I > A_D) = \sum_{n=1}^N P \left[a \left(n + \frac{1}{2} \right) > A_I(f) > a \left(n - \frac{1}{2} \right) \right] \cdot P(A_D < an)$$

where n is an integer variable, and $a \cdot N$ equals the peak value of A_I (N is sufficiently large, so that a is negligibly small).

This probability expression may be mapped to the frequency plane. The amplitude is not quantized directly; instead, the frequency scale is divided into many small equal intervals. For practical cases the intervals need only extend into the frequency region of relatively high carrier probability. If the intervals are sufficiently small, each may be characterized by a single (center) frequency and an interference carrier amplitude. There, at any given interval, n ,

$$\Delta P(A_I > A_D)|_n = P \left[F_I \cdot \left(n + \frac{1}{2} \right) > f_{Ii} > F_I \cdot \left(n - \frac{1}{2} \right) \right] \cdot P[A_D(f) < A_I(F_I \cdot n)]$$

where F_I = the width (frequency) of an interval

n = the number of the interval

f_{Ii} = the instantaneous interference frequency.

As the intervals decrease in width, the probability density becomes

$$p(A_I > A_D, f) = p(f_{Ii}) \cdot p[A_D(f) < A_I(f_{Ii})].$$

The phasor rotational rate for the CW case was shown to be the difference in instantaneous frequency between the two carriers. This is also

correct, on the average, for the modulated case, since the modulation of the desired carrier has a zero mean. The impulse rate density may be written as

$$\frac{d}{df_{Ii}} (\text{rate}) = |F_0 - f_{Ii}| \cdot p(A_I > A_D, f) \\ = |F_0 - f_{Ii}| \cdot p(f_{Ii}) \cdot p[A_D(f) < A_I(f_{Ii})].$$

The impulse rate therefore is

$$\text{rate} = \int_{-\infty}^{+\infty} |F_0 - f'| \cdot p(f') \cdot p[A_D(f) < A_I(f')] df'.$$

The analysis is generalized by including all of the filters originally described in the system. Thus, the amplitude responses are

$$A_I(f) = |G_{TI}(f)| \cdot |G_{RD}(f)| \cdot A_I = \text{cross-path gain} \cdot A_I$$

$$A_D(f) = |G_{TD}(f)| \cdot |G_{RD}(f)| \cdot A_D = \text{through-path gain} \cdot A_D$$

where $G_{xy}(f)$ are the filter frequency responses.

The phase response of the filters does not enter into the interference analysis. The instantaneous rotational rate will be modulated by the filter phase response as the carrier deviates across the filter response. It is assumed that deviations in one direction are complemented by those in the opposite direction and that the energy contributions from such modulations cancel when averaged over a short time relative to the memory of the filter. The filter phase response cannot be neglected when considering in-band performance, however.

While integration of the rate equation derived previously is not practical in closed form, an algorithm has been written to perform the computation. The logic is described in Appendix B.

Channel Degradation Due to Impulse Noise

Although the time waveform of each impulse cannot be specified, the area included is 2π radians (1 cycle). Examination of the cross-path gain* for practical cases reveals that the impulse duration is short compared with the period of the highest baseband frequency.

Use of filters of Carson's Rule bandwidth will result in a cross-path gain response peaked near the passband edge (Figure 9); consequently, most impulses will occur when the interfering carrier is in this region.

* Cross-path gain is defined as the amplitude of the transfer function of the path from the interfering carrier modulator to the desired carrier demodulator.

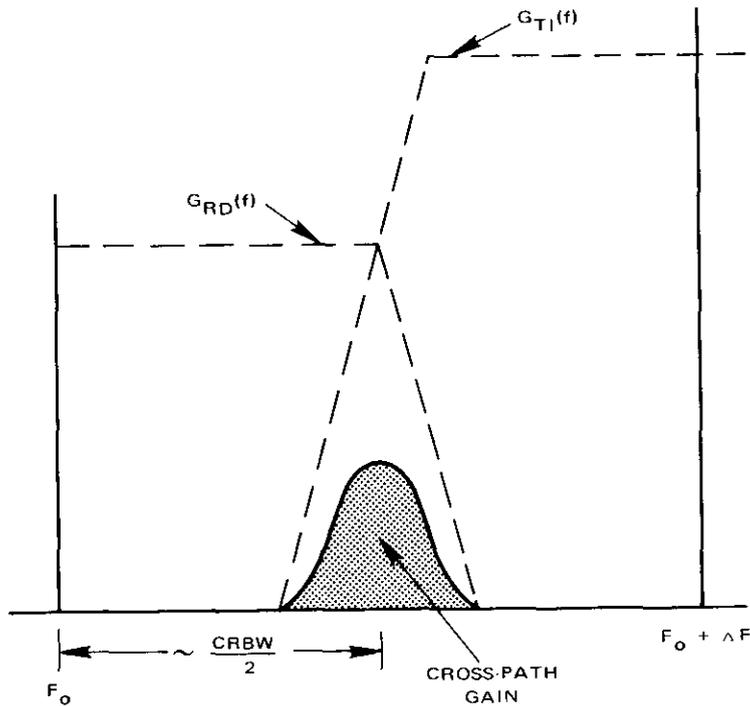


Figure 9. Illustration of Cross-Path Gain

If the rms modulation index is 1.5, the cross-path peak is located at 5.7 times the upper base-band frequency (from center frequency). Re-examining the vector diagram (Figure 8) indicates that most of the 2π phase-shift of an impulse occurs over a small fraction of the total vector rotational period. The time duration of the impulse therefore is much less than $1/6 f_m$. Each channel in the baseband will intercept approximately equal fractions of the total impulse power.

The individual channels in the composite baseband must be multiplexed to voice frequency before distribution to individual subscribers. A typical demultiplexing technique is schematically illustrated in Figure 10.

Two characteristics of impulse noise must be determined to evaluate the quality of a communications channel, the time average power, and the peak voltage response. As has been pointed out, each impulse is of negligible duration relative to the response time of a channel filter. The channel occupying the 0- to 3-kHz slot in the composite baseband will be

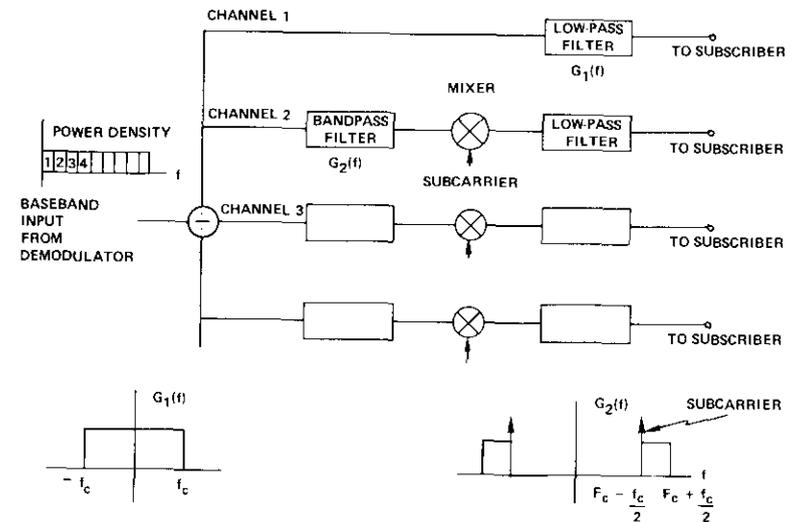


Figure 10. Diagram of Demultiplex Process

examined first; a low-pass filter, such as the one shown in Figure 11, performs the required demultiplexing.

Assume a rectangular filter having cut-off frequency f_c . Clearly, when the filter is driven by an impulse of negligible duration and unit area, response is

$$y(t) = Kh(t) = K \frac{\omega_c}{\pi} \frac{\sin \omega_c t}{\omega_c t}$$

The peak voltage therefore is

$$y_{pk}(t) = K \frac{\omega_c}{\pi} = K2f_c$$

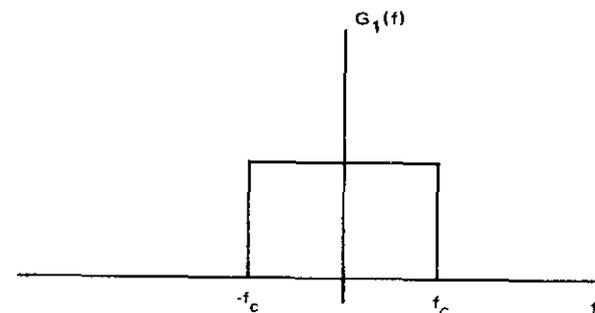


Figure 11. Low-Pass Channel Filter Response

where K is the demodulator deviation sensitivity. This result may be written as a ratio to remove the sensitivity constant. Thus,

$$\text{test tone/peak impulse} = 20 \log \Delta f_{it}/2f_c$$

where Δf_{it} is the rms deviation of a standard test tone.

The analysis of demultiplexing a higher-frequency channel is more complicated. The proper channel is selected from the composite baseband by a bandpass filter (Figure 12). From Appendix C the impulse response for this case is

$$y(t) = Kh(t) = \frac{K}{\pi t} [\cos(\omega_c t - \theta) - \cos \theta] \quad \text{for } t > 0$$

where θ is the phase of the subcarrier at the time of arrival of the impulse. The peak response amplitude therefore is a function of θ , but not of the subcarrier frequency; i.e., all channels are equally affected.

It is of interest to determine the probability statistics of the amplitude distribution of the peaks. An empirical approach using a digital computer was employed to determine maximums as a function of phase angle. The assumption of uniform phase distribution allows the resulting probability density to be plotted in Figure 13. An eighth-order polynomial is fitted to the data from 100 random-phase samples. From the illustrated distribution, the multiplex process attenuates the peak voltage of many of the impulses.

In addition to the peak voltage, a second impulse characteristic of importance is the average noise power contribution. While the absolute power can be computed, given the demodulator sensitivity, it is more meaningful to compute a ratio of noise power to signal power. The noise power ratio, which is simply the power ratio of undesired noise in a

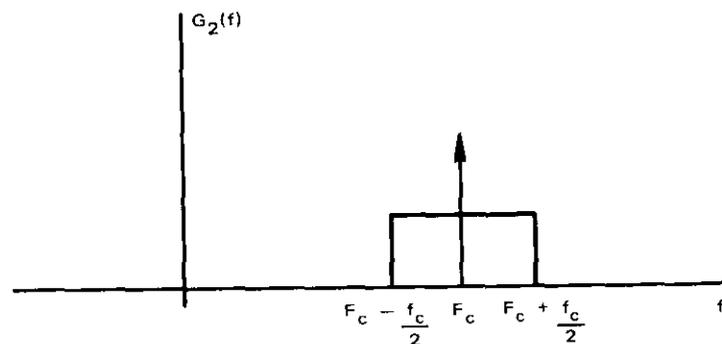


Figure 12. Demultiplex Bandpass Filter Response

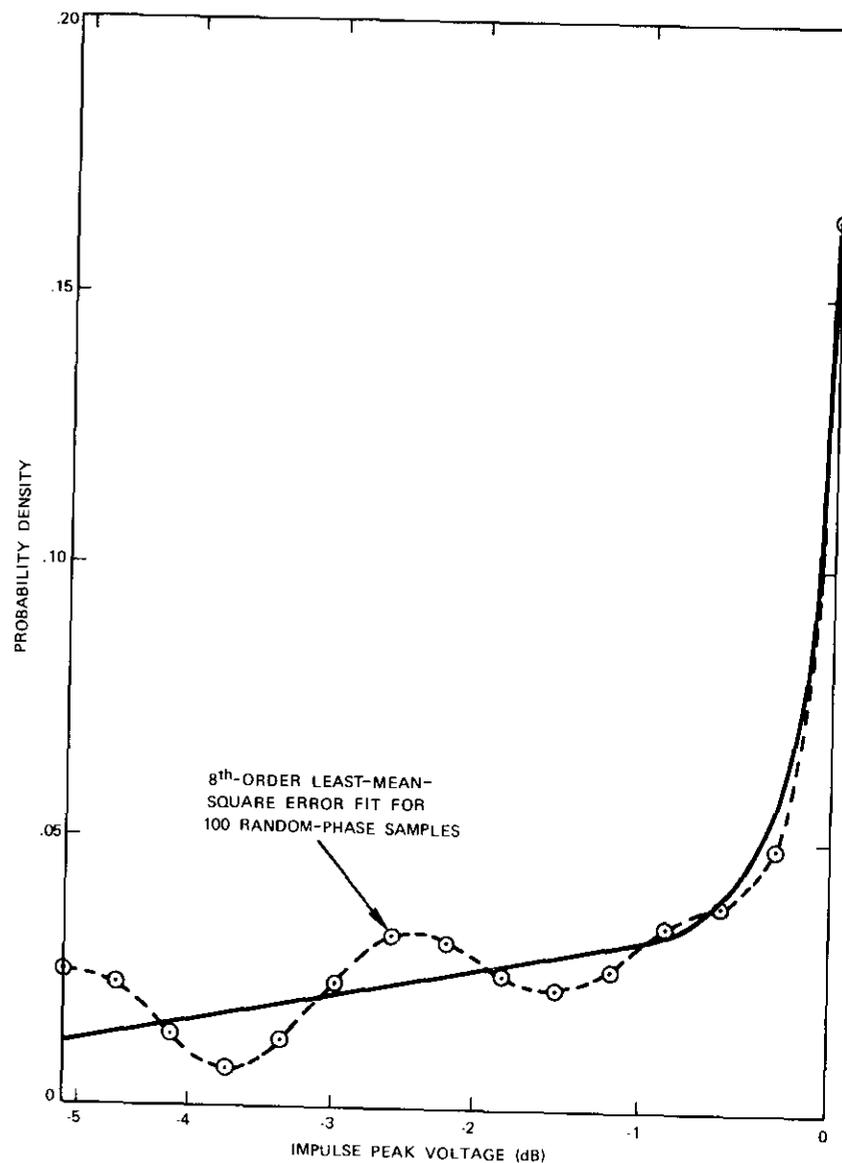


Figure 13. Peak Impulse Statistics

channel to the Gaussian noise used to simulate the desired channel, is a useful measurement of channel quality.

Returning to the impulse response of the channel low-pass filter,

$$y(t) = Kh(t) = \frac{K\omega_c}{\pi} \frac{\sin \omega_c t}{\omega_c t}$$

The total energy per impulse is given by

$$E = \int_{-\infty}^{\infty} v^2(t) dt = K^2 \int_{-\infty}^{\infty} h^2(t) dt = K^2 2f_c$$

where f_c is the low-pass filter cut-off frequency.

A repetitive train of impulses having an average rate of one per second therefore will have an average power equal to $K^2 2f_c$ at the filter output. The signal power in each channel is equal to $K^2 (\Delta f_{rms})^2/n$, where n is the total number of channels. The NPR due to impulse noise therefore is

$$NPR = 10 \log \frac{(\Delta f_{rms})^2}{2nf_c} \cdot \text{impulse rate.}$$

It is not necessary to separately compute the NPR of a higher-frequency channel, since the power spectrum is uniform across the passband.

CONVOLUTION NOISE

The second interference degradation, convolution noise, exists even when the desired carrier amplitude is greater than the interference. This analysis is simplified by initially neglecting the filters, then later including them. The input to the demodulator is $f_s(t) + f_i(t)$, where

$$f_s(t) = \text{desired carrier} = A_s(t) \cos [\omega_0 t + \phi(t)]$$

$$f_i(t) = \text{interfering carrier} = A_i(t) \cos [(\omega_0 + \Delta\omega) t + E(t)]$$

and $\overline{A_i(t)} < \overline{A_s(t)}$ (where the bar notation represents the time average), and $\phi(t)$, $E(t)$ are the modulation processes. For convenience, let $A(t) = A_i(t)/A_s(t)$, where, under all practical operating conditions, $A(t)$ will always be less than unity except for relatively infrequent impulse periods. Linearly summing the two carriers results in a composite signal:

$$M(t) = \cos [\omega_0 t + \phi(t)] + A(t) \cos [(\omega_0 + \Delta\omega) t + \xi(t)]$$

where $\xi(t)$ is the modulation distorted by filtering. If

$$K(t) \cos \alpha(t) = 1 + A(t) \cos [\Delta\omega t - \phi(t) + \xi(t)]$$

and

then

$$\begin{aligned} M(t) &= \cos [\omega_0 t + \phi(t)] \cdot K(t) \cos \alpha - \sin [\omega_0 t + \phi(t)] \\ &\quad \cdot K(t) \sin \alpha(t) \\ &= K(t) \cos [\omega_0 t + \phi(t) + \alpha(t)] \end{aligned}$$

$$\alpha = \tan^{-1} \frac{\{A(t) \sin [\Delta\omega t - \phi(t) + \xi(t)]\}}{\{1 + A(t) \cos [\Delta\omega t - \phi(t) + \xi(t)]\}}$$

An ideal limiter with infinite dynamic range and no AM/PM conversion will remove the amplitude variation, resulting in

$$M_1(t) = A \cos [\omega_0 t + \phi(t) + \alpha(t)].$$

The FM spectrum of $\alpha(t)$ in the baseband then may be obtained from the Fourier transform,

$$\text{baseband interference spectrum} = F[\dot{\alpha}(t)] = \omega^2 F[\alpha(t)]$$

where F is the Fourier transform operator. $\alpha(t)$ may be expanded as a series of harmonically related terms. For the typical case where $\overline{A(t)} \ll 1$, only the first term need be considered. When $A(t)$ approaches unity, however, the higher harmonic terms also approach unity amplitude, and the power spectral density of these terms must be computed for completeness.

For the case where $\overline{A(t)} \ll 1$, $\alpha(t)$ may be approximated by

$$\alpha(t) = A(t) \sin [\Delta\omega t - \rho(t) + \xi(t)]$$

which may be rewritten as

$$\begin{aligned} \alpha(t) &= \sin [\Delta\omega t - \rho(t)] A(t) \cos \xi(t) \\ &\quad + \cos [\Delta\omega t - \rho(t)] A(t) \sin \xi(t). \end{aligned}$$

Thus $\alpha(t)$ can be represented as the product of two time functions, $\sin [\Delta\omega t - \rho(t)]$ and $A(t) \cos \xi(t)$. Consequently, the resulting baseband noise spectrum can be derived by convolving the spectra of the two modulated RF carriers.

The application of filtering to the system serves to modify the RF spectra:

$$W'_D(f) = |G_{RD}(f)|^2 |G_{TD}(f)|^2 \cdot W_D(f)$$

$$W'_D(f) = |G_{RD}(f)|^2 |G_{TD}(f)|^2 \cdot W_D(f)$$

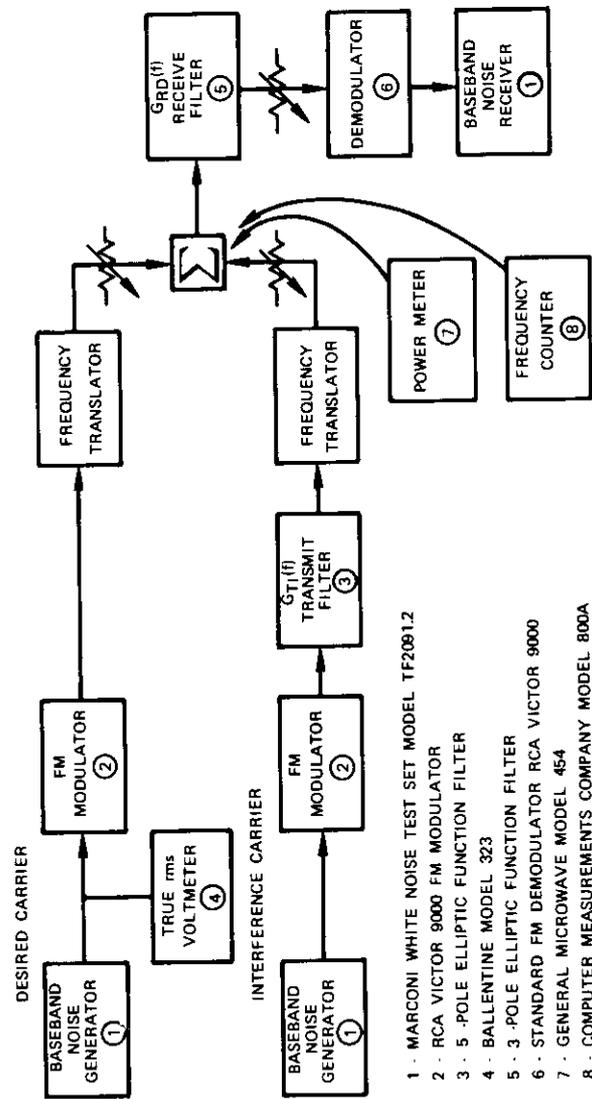


Figure 14. Measurement Set-Up

where $W_D(f), W_I(f)$ = the prefiltered desired and interfering carrier power spectra

$W'_D(f), W'_I(f)$ = the postfiltered desired and interfering carrier power spectra.

As in the unfiltered case, the noise contribution may be computed by convolving the filtered spectra, $W'_D(f)$ and $W'_I(f)$.

EXPERIMENTAL RESULTS

Measurements were performed for a specific case illustrated in Figure 14. Variable frequency translators are included in the transmission path to allow adjustment of frequency separation. The operating parameters as shown in Table 1 and the filter amplitude responses shown in Figures 15 and 16 represent typical INTELSAT operation.

The system characteristics were entered into the computer, and the results are compared in Figure 17. Cases with and without transmitter

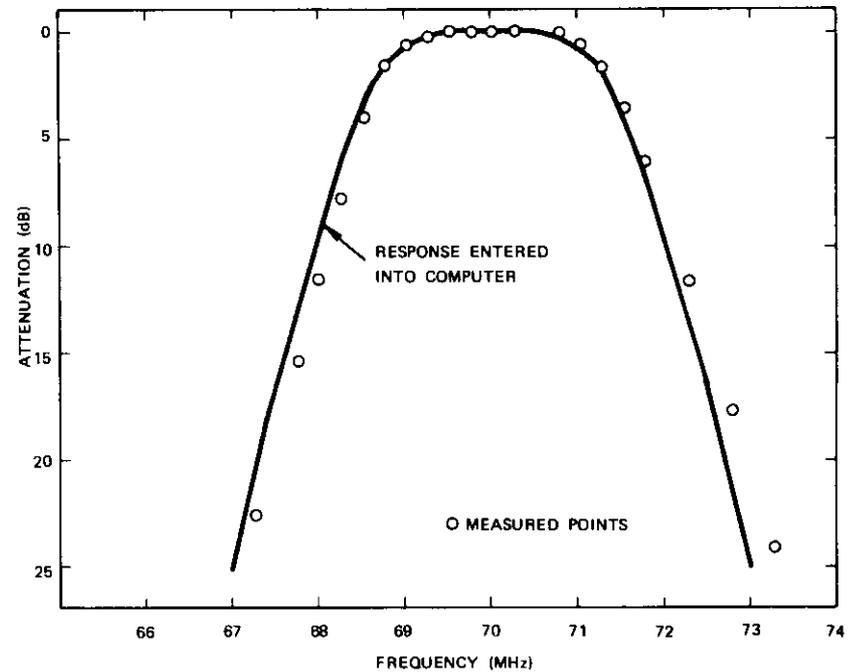


Figure 15. Receive Filter Response

TABLE 1. OPERATING PARAMETERS

Parameter	Desired Carrier	Interference Carrier
Number of Channels	24	132
Upper Baseband Frequency	108 kHz	552 kHz
rms Full-Load Deviation	363 kHz ($\beta = 3.36$)	1.253 MHz ($\beta = 2.27$)
Channel Test-Tone Deviation	216 kHz	530 kHz
Amplitude	0 dB	+6 dB
Center Frequency	70 MHz	Variable (>70 MHz)
Transmit Filter	No	Yes (see Figure 16)
Receive Filter	Yes (see Figure 15)	—
Carrier-Thermal Noise	>30 dB	>30 dB
Carson's Rule Bandwidth	2.5 MHz	9.0 MHz

filtering (on the interference) were studied as a function of carrier separation. The top and bottom channels exhibit the same *NPR* performance, indicating that the impulse noise is dominant for these cases. For these cases, the frequency separation corresponding to Carson's Rule band-

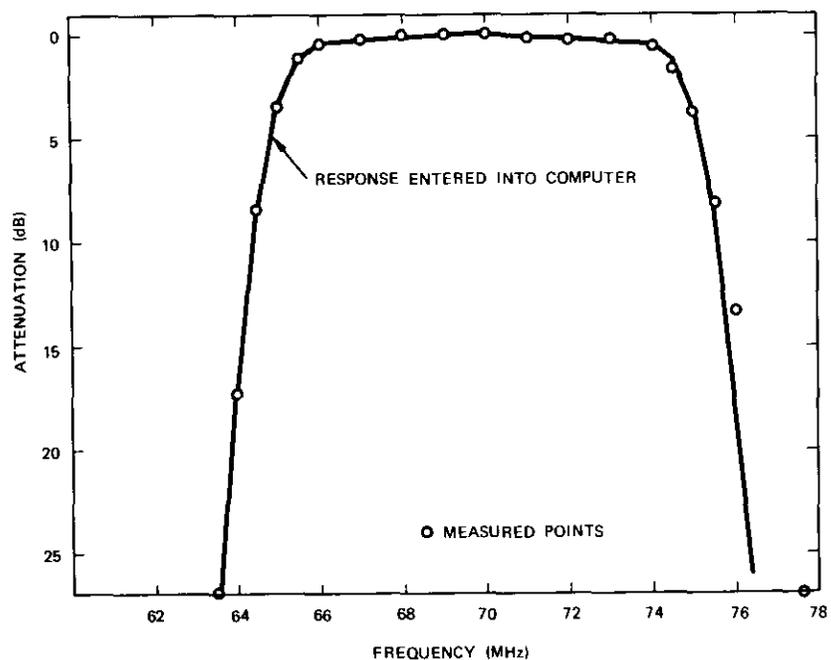


Figure 16. Transmit Filter Response

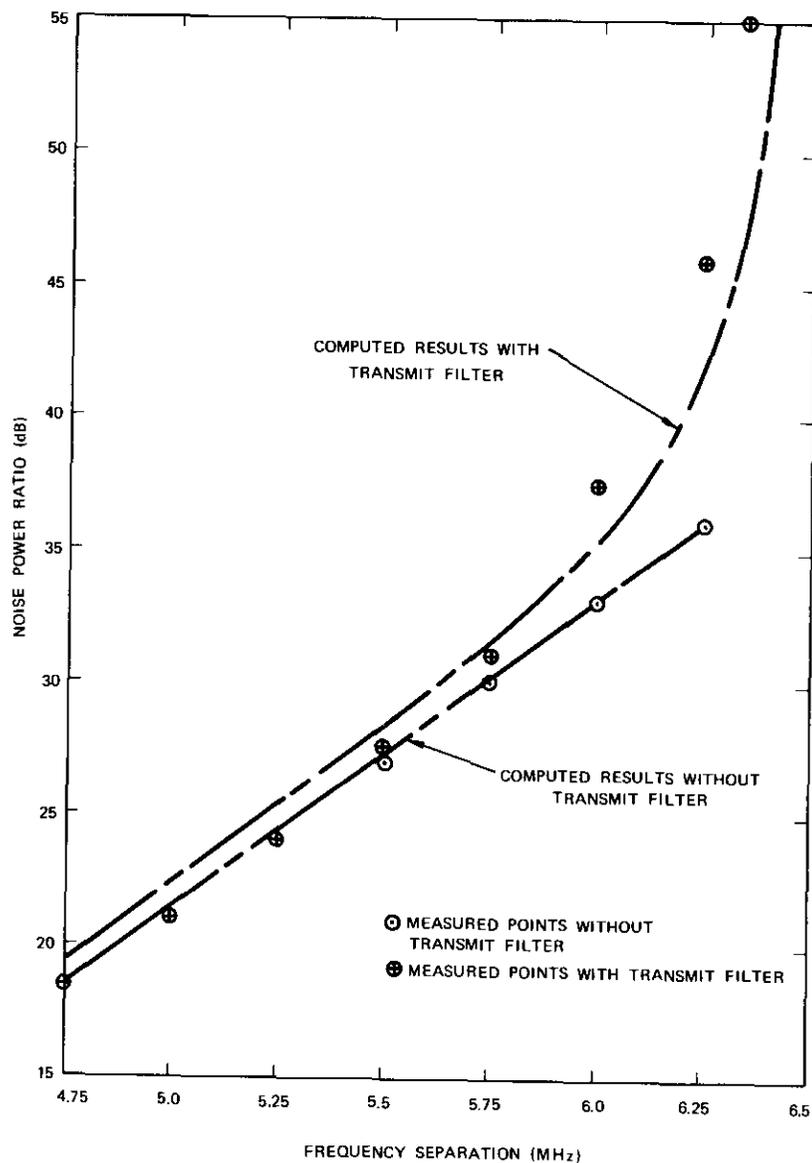


Figure 17. Comparison of Measured and Computed Results

width is 5.77 MHz. The transmit filter results in improved performance for the cases where the equal-ripple bandwidths of the filters do not overlap.

CONCLUSIONS

The analysis in this paper provides, with the aid of a digital computer, a means of closely predicting the performance of the system with respect to adjacent channel interference. The selection of proper filtering for limitation of impulse noise, however, can be facilitated by a more simplified analysis.

If it is assumed that the desired carrier is unmodulated, the probability of impulse will become negligible if the cross-path attenuation is made equal to or greater than the difference in carrier power (i.e., the interference carrier power is greater). The bandwidth characteristic of the cross-path gain is also of importance to the impulse performance. A narrow cross-path bandwidth reduces the time that impulse conditions are satisfied.

A useful procedure for establishing operating parameters is described. Given the channel capacity and modulation index requirements for each carrier, a bandpass filter is selected whose equal-ripple (or flat) bandwidth corresponds to the Carson's Rule bandwidth of the carrier (assuming a 10-dB pk-rms ratio for the deviation). When the filter is operated with group-delay equalization over the entire equal-ripple bandwidth, the distortion performance will have a 1- to 3-dB margin for over-deviation (i.e., *NPR* better than 50 dB). Selection of the order of the filter depends on the specific adjacent channel environment. In general, five to seven poles is a good first try for cases where the carrier power difference is up to 10 dB. The absolute filter skirt slope (in dB/MHz) should be the same for adjacent filters (i.e., more sections for wider bandwidth carriers). For this equal-slope condition, the cross-path gain characteristics will be considered between the filters, thus simplifying system design.

While only adjacent channel effects have been examined to this point, thermal noise must also be considered for system design. As was described, the cross-path gain characteristic should attenuate the interfering carrier to a level below that desired. A conservative but simplifying assumption regarding the thermal noise is that it may be characterized as a phasor rotating about the head of the desired carrier phasor. There is a finite probability that the amplitude will peak (i.e., at 13 dB above its nominal rms) so as to add to the adjacent carrier interference. The filter spacing and slope therefore should be designed to provide a minimum cross-path attenuation as follows:

$$CPA_{dB} = 20 \log \frac{A_I}{A_D} + 10 \log \frac{A_D^2}{N_T} + P_{N_T}$$

when

$$10 \log \frac{A_D}{N_T} + P_{N_T} > 0 \text{ dB}$$

where CPA_{dB} = cross-path attenuation, dB

A_I = carrier amplitude (interference)

A_D = carrier amplitude (desired)

N_T = total thermal noise power

P_{N_T} = peak-to-rms ratio of thermal noise.

Conclusions may be drawn from the theoretical and experimental investigations performed. With proper transmit and receive filtering, carrier spacings 10 percent greater than Carson's Rule bandwidth conservatively provide high-quality telephone channels as defined by the CCIR. With a standard demodulator, an *NPR* of 45 dB (top and bottom channels) was attained under this condition. Under the same condition, a phase-locked-loop (PLL) demodulator yields 63 dB and 50 dB in the bottom and top channels, respectively. The difference in adjacent channel interference performance between the standard and PLL is evident. In either case, both transmit and receive filters are required for optimum performance. It is practical to operate with steep-sided, equalized, high-order (5-pole + equalizer) filters having an equal-ripple bandwidth of Carson's Rule (bandwidth of the filtered carrier) and to obtain *NPR*'s of 50 to 55 dB (in the worst channel).

REFERENCES

- [1] F. L. H. M. Stumpers, "Interference Problems in Frequency Modulation," *Philips Research Reports*, Vol. 2, No. 2, April 1947, pp. 136-160.
- [2] M. S. Corrington, "Frequency Modulation Distortion Caused by Common- and Adjacent-channel Interference," *RCA Review*, Vol. 7, No. 4, Dec. 1946, pp. 522-560.

- [3] C. L. Ruthroff, "Computation of FM Distortion in Linear Networks for Bandlimited Periodic Signals," *Bell System Technical Journal*, Vol. 47, No. 6, July-Aug. 1968, pp. 1043-1063.
- [4] R. G. Medhurst, E. M. Hicks and W. Grossett, "Distortion in Frequency-division-multiplex F.M. Systems Due to an Interfering Carrier," *Proceedings of the IEE (U.K.)*, Vol. 105B, No. 21, May 1958, pp. 282-292.
- [5] R. G. Medhurst, "F.M. Interfering Carrier Distortion: General Formula," *Proceedings of the IEE (U.K.)*, Vol. 109B, No. 44, March 1962, pp. 149-150.
- [6] R. Hamer, "Radio-frequency Interference in Multi-channel Telephony F.M. Radio Systems," *Proceedings of the IEE (U.K.)*, Vol. 108B, No. 37, Jan. 1961, pp. 75-89.
- [7] M. Schwartz, W. R. Bennett and S. Stein, *Communication Systems and Techniques*, New York: McGraw-Hill, 1966, pp. 144-145.
- [8] R. Hamer and R. A. Acton, "Power Spectrum of a Carrier Modulated in Phase or Frequency by White Noise," *Electronic and Radio Engineer*, Vol. 34, No. 7, July 1957, pp. 246-253.

APPENDIX A. LEVEL-CROSSING ANALYSIS

An analysis of the expected number of crossings of the normal process, $y(t)$, at some arbitrary level, y_0 , for the interval t_1, t_2 appears in Reference [A1]. The expression for the number of crossings derived there is

$$N_0 = \frac{1}{\pi} P_y e^{-y_0^2/2b_0}$$

where

$$b_0 = \overline{y^2}$$

$$P_y^2 = \frac{b_2}{b_0} = \frac{\int_0^\infty W_y(f) \omega^2 df}{\int_0^\infty W_y(f) df}$$

and $W_y(f)$ is the spectral density of the process. For a rectangular baseband spectrum,

$$P_y^2 = \frac{\int_0^{f_m} \omega^2 df}{\int_0^{f_m} df} = \frac{4\pi^2 f_m^2}{3}$$

where f_m is the upper baseband frequency. The number of excursions beyond a given level, y_0 , therefore is

$$n = \sqrt{\frac{1}{3}} \cdot f_m \cdot e^{-1/2(y_0/f_{rms})^2}$$

where f_{rms} is the rms deviation.

REFERENCE

- [A1] D. Middleton, *An Introduction to Statistical Communication Theory*. New York: McGraw-Hill, 1969, pp. 426-428.

APPENDIX B. COMPUTER PROGRAM

Figure B-1 is the flow chart of the algorithm for performing the analysis. The system parameters are entered first. The filter amplitude responses are specified either in polynomial or tabular form. The remaining parameters (i.e., modulation indices, upper baseband frequencies, carrier amplitudes, and frequency separation) are entered in a conversational mode.

The convolution noise degradation is computed first; the respective RF spectra are calculated and then weighted by the appropriate filter responses. The modified spectra are then convolved for the top baseband channel. For most practical cases, the convolution noise will be greatest in this channel. The resulting *NPR* is printed out.

The impulse noise computation is performed next. The frequency domain between the carrier center frequencies is divided into small equal intervals. Each interval, in turn, is characterized by

- center frequency,
- the probability that the instantaneous frequency of the interference will be within the interval, and
- the amplitude when the interference is within the interval.

The probability that the desired carrier amplitude is less than or equal to the interference (for each interval) is computed. The product of the two probabilities yields the joint probability of satisfaction of the amplitude criterion. The center frequency of the interval is used next to determine the approximate vector rotational rate. Multiplication of this rate by the joint probability results in the rate of impulse attributable to the specific interval. The total impulse rate is obtained by summing over all intervals.

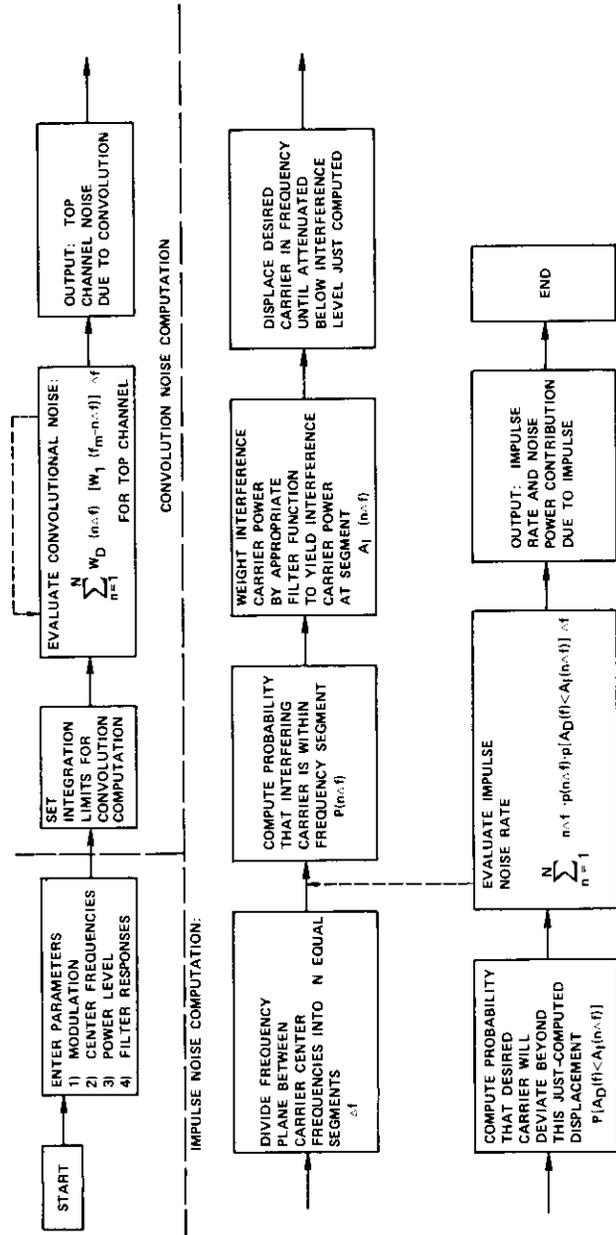


Figure B-1. Computer Program Flow Chart

APPENDIX C. DERIVATION OF THE RESPONSE TO IMPULSES OF A DEMULTIPLEX CHANNEL FILTER

The filter to be analyzed is illustrated in Figure C-1.

The baseband output of the demodulator is applied to a narrow-bandwidth rectangular bandpass filter, $G_1(f)$, centered on the desired channel of the composite band (Figure C-2). The filter output is synchronously down-converted by a local pilot carrier phase locked to a complementary carrier in the multiplex equipment. The interference-generated impulses occur randomly with respect to the pilot carrier phase. The output of the down-converter is then low-pass filtered to eliminate the pilot carrier component.

The impulse response of $G_1(f)$ first is derived as

$$h(t) = F(f)^{-1} = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\omega) e^{j\omega t} d\omega$$

where Ω_c is the channel center frequency, and ω_c is the channel bandwidth. Thus,

$$h(t) = \frac{1}{2\pi} \left[\int_{-\Omega_c - (\omega_c)/2}^{-\Omega_c + (\omega_c)/2} e^{j\omega t} d\omega + \int_{\Omega_c - (\omega_c)/2}^{\Omega_c + (\omega_c)/2} e^{j\omega t} d\omega \right] \\ = \left(\frac{\omega_c}{\pi} \frac{\sin \omega_c t / 2}{\omega_c t / 2} \right) e^{j\Omega_c t} = \Upsilon(t)$$

This response is mixed with the pilot carrier, $e^{j\{(\Omega_c + (\omega_c/2)t + \theta)\}}$, where θ represents the random carrier phase. The pilot carrier frequency is assumed to correspond to the upper edge of the channel. Therefore,

$$\text{channel output} = \Upsilon(t) \cdot e^{j\Omega_c t} \cdot e^{j\{(\Omega_c + (\omega_c/2)t + \theta)\}}$$

Reduction of this expression and filtering of high-frequency terms yields

$$\text{channel output} = \sin \left(\frac{\omega_c t}{2} + \theta \right) \cdot \frac{\omega_c}{\pi} \frac{\sin \omega_c t / 2}{\omega_c t / 2} \\ = \frac{2}{\pi t} \sin \frac{\omega_c t}{2} \sin \left(\frac{-\omega_c t}{2} + \theta \right)$$

but, $\sin A \sin C = 1/2[\cos(A - C) - \cos(A + B)]$. Therefore,

$$\text{channel output} = \frac{1}{\pi t} [\cos(\omega_c t - \theta) - \cos \theta]$$

which is the desired result.

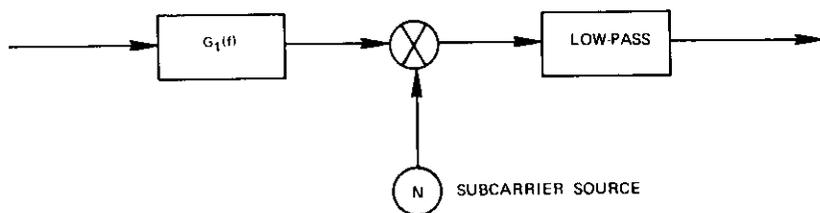


Figure C-1. Diagram of a Single Multiplex Channel

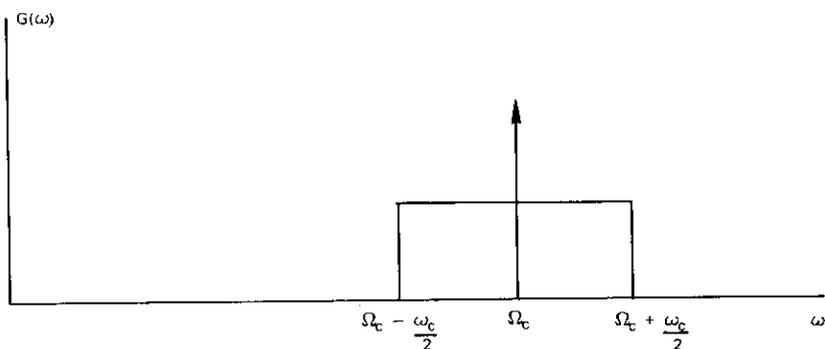


Figure C-2. Channel Bandpass Filter Response

ACKNOWLEDGMENT

The author wishes to acknowledge the invaluable contributions made by Mr. A. Berman and Mr. L. Pollack of COMSAT Labs. The filter synthesis expertise of Mr. F. Assal is also gratefully acknowledged.



Marvin R. Wachs received the B.E.(EE) degree from the City College of New York in 1965 and the M.S. degree in Electrical Engineering from George Washington University in 1969. For the past six years he has been employed as a member of the technical staff at COMSAT Laboratories specializing in the area of RF transmission as applied to satellite transponders. Mr. Wachs is a member of Tau Beta Pi and Eta Kappa Nu.

Index: communication satellites, data transmission, earth terminals, ground support equipment, multi-channel communications, multiplexing

THE SPADE SYSTEM AS APPLIED TO DATA COMMUNICATIONS AND SMALL EARTH STATION OPERATION

EUGENE R. CACCIAMANI, JR.

ABSTRACT

The SPADE system is capable of providing many types of services other than single-channel-per-voice-carrier transmissions. The SPADE channel unit can be operated in either continuous or voice-activated mode, depending on whether data or voice is transmitted. Narrow-band data normally transmitted over 4-kHz telephone circuits can be processed directly by the existing channel unit configuration. Wide-band data (40.8 kbps or 48 kbps) can also be transmitted by a single-channel unit. In addition, the bit-error-rate performance of the wide-band data signals can be significantly improved by utilizing error-correcting techniques which can be readily added to the basic channel unit configuration. The composite required bandwidth is equal to that used for single-channel voice.

Special applications of the SPADE system and its inherent flexibility have made consideration of small earth station terminal networks particularly attractive. Trade-offs which consider earth station antenna size, service requirements, channel coding requirements, and available satellite transponder e.i.r.p. are evaluated, and typical configurations that can be implemented as a direct result of current experimental studies are presented.

INTRODUCTION

The SPADE system [1], [2] utilizes single-channel-per-voice-carrier PCM/PSK transmissions. In the operational configuration, the system has been designed for 7-bit PCM encoding and operation at 56 kbps with a transmitted bit rate equal to 64 kbps. Since 4-phase coherent PSK modulation is used, the transmitted symbol rate is 32,000 symbols per second.

This paper is based upon work performed in COMSAT Laboratories under Corporate sponsorship and under the sponsorship of the International Telecommunications Satellite Consortium (INTELSAT). Any views expressed in this paper are not necessarily those of INTELSAT.

The purpose of this paper is to discuss not only the digital channel unit configuration designed for voice operation, but also to show how the SPADE channel unit can be used for other types of service. In addition to single-channel-per-voice-carrier transmissions, narrow-band baseband data can be handled directly by using the standard voice channel unit configuration. Wide-band data (40.8 kbps) can also be transmitted by using the basic single voice channel unit with certain peripheral modifications that can be readily added. In addition, narrow-band baseband signals can be multiplexed and then transmitted in a manner similar to that for wide-band data.

Another application is the transmission of voice by using delta modulation and encoding the voice at one-half the normal rate used for 7-bit PCM encoding, i.e., 28 kbps. Tests have shown that a good quality voice channel can be maintained at this rate and that the resultant bandwidth reduction can be used advantageously when considering small G/T^* value earth station operating networks [3]. By reducing the per-channel encoded bandwidth requirement by one-half, for example, we can increase the system capacity by a factor of approximately two. In addition, if a power-limited condition still exists, rate-1/2 convolutional encoding (or, if necessary because of bandwidth limitations, higher-rate codes) can be used to further improve the system channel capacity. In any event, system capacity generally can be improved for small G/T station operation by using suitable coding techniques for either 7-bit PCM at 56 kbps or delta modulation at 28 kbps [3].

SINGLE-CHANNEL PCM/PSK VOICE CHANNEL UNIT

Figure 1 is a block diagram of the voice channel unit implementation. The first SPADE operational network will use an INTELSAT IV global-beam transponder with 4-phase coherent PSK modulation [1], [2], [3]. Threshold performance for the PCM codec is defined at a bit-error-rate probability equal to 1×10^{-4} ; the carrier-to-noise ratio in the channel bandwidth required to provide threshold performance is 13.0 dB. A system capacity of 800 channels can be derived if voice-activated [1], [2], [3] operation and a system thermal noise margin of 4.2 dB are assumed.

A single-channel unit consists of a pair of transmit and receive subunits that provide digital encoding and decoding of the analog voice, digital synchronization to recover frame reference and remove

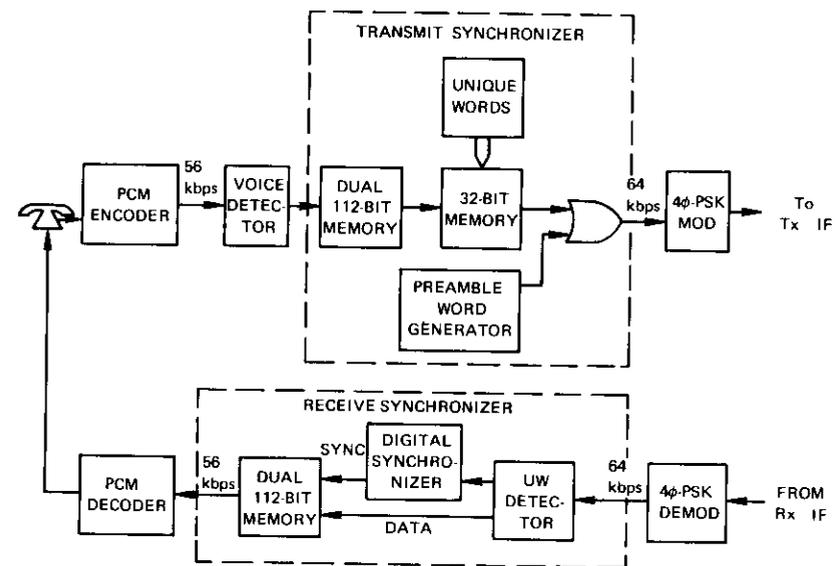


Figure 1. Single-Channel PCM/PSK Voice Channel Unit

phase ambiguity of the recovered carrier, and modulation and demodulation of the transmitted and received data. The PCM encoder uses standard A-law companding, and the modulation is 4-phase coherent PSK. The transmit synchronizer consists of a digital voice detector, dual 112-bit memory, a preamble word generator, and associated input/output timing. The PCM encoded output is at 56 kbps, and the digital voice detector is used to determine the presence of speech and provide on-off control of the carrier. If on-off switching of carrier power is used, the resultant saving in accessed satellite transponder power is approximately 4 dB [4]. The data are stored in memory, and two 16-bit unique words are inserted every 224 data bits. The resultant transmitted bit rate is 64 kbps, and the symbol rate, using 4-phase PSK, is 32 kbps.

During the beginning of each voice burst, a preamble word and two unique words are inserted to provide carrier and bit-timing recovery and phase ambiguity resolution at the corresponding receiver. The format is shown in Figure 2. The transmitted carrier and bit timing must be coherently recovered within 1.875 milliseconds at the beginning of each voice burst, and simultaneous detection of the two unique words is required to resolve the phase ambiguity of the 4-phase PSK recovered carrier [5].

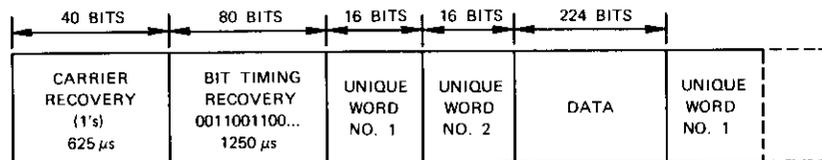


Figure 2. Preamble and Frame Format

In addition to providing resolution of recovered carrier phase ambiguity, unique word detection is used to provide digital frame synchronization. The receive synchronizer observes each unique word detected output and requires two consecutive detections, which are exactly 224 bits apart, at the beginning of each voice burst in order to lock into a closed-aperture mode. During closed-aperture operation, the synchronizer looks for a unique word detection every 224 bits, referenced to the original recovered frame timing. If a detection is absent, the synchronizer inserts a sync signal and continues this procedure until five consecutive miss detections are observed. The synchronizer then drops out of lock and begins again in the open-aperture search mode.

During closed-aperture operation, the occurrence of false unique word detections is ignored; hence, the consideration of false detections is limited to initial acquisition (open-aperture mode) of the digital synchronizer. After the demodulated data have been passed through the synchronizer, they are transmitted to a dual 112-bit memory, where the unique words are removed. The resultant 56-kbps bit stream with associated timing is received by the PCM decoder, where it is reconverted into its original analog waveform.

NARROW-BAND DATA TRANSMISSIONS

Data that are normally transmitted over voice bandwidths, such as 1200 bps, 2400 bps, and 4800 bps, can be transmitted directly over the voice channel unit shown in Figure 1. The data input is sampled by the PCM encoder, and the voice detector is by-passed to provide continuous transmission. Laboratory tests have demonstrated that this is a realizable configuration and that no degradation is introduced by using the PCM encoder to resample the data input.

WIDE-BAND DATA TRANSMISSIONS

It is apparent that data (narrow band or wide band) can be transmitted as a separate service by using a different channel unit configuration for each data service considered. In general (for rates less than

the nominal SPADE transmitted rate), this requires less bandwidth and power per transmission. However, since the preponderance of traffic using SPADE will be telephony, it is important to show that the same channel unit can be used to serve other traffic needs as well.

It is generally preferred to implement and maintain a system consisting of a limited number of fundamental components that can be readily configured to serve many types of traffic. If, for example, 40.8-kbps data transmissions are considered, one approach is to provide a different modem and transmit the data at 46.6 kbps,* which is considerably less than the 64-kbps transmitted rate of the standard voice channel configuration. Another approach is to make the 40.8-kbps data rate compatible with the desired input rate to the standard channel unit, either by "bit stuffing" or by using a rate-changing code such as a rate-3/4 convolutional code. This would enable the new service to conform to the existing system configuration. In addition, the use of convolutional encoding would provide an improved bit-error-rate performance.

Figure 3 shows the basic channel unit (excluding the PCM codec and voice detector) which is preceded at the transmitter by a rate-3/4 convolutional encoder and followed at the receiver by a rate-3/4 threshold decoder. This additional logic requires approximately 50 dual in-line integrated circuits, and its implementation in an experimental SPADE channel unit has been relatively simple. The timing is generated from the 40.8-kbps data signal by a clock recovery circuit which is used to phase lock the internal timing generators of the channel unit. The output rate of the rate-3/4 convolutional encoder is 54.4 kbps. The channel unit stuffs unique word bits at the 8/7 ratio, and thus requires a transmission rate equal to 62.17 kbps. The resultant bit rate is less than, but sufficiently close to, the modem-designed operating frequency at 64 kbps, so the design tolerance of the modem loops will recover the 62.17-kbps signal with little or no additional degradation, and the same channel filters can be used.

In addition, the threshold $P_{BE} = 1 \times 10^{-4}$ is enhanced to approximately $P_{BE} = 1 \times 10^{-8}$ by using 3/4 threshold decoding. Thus, if the configurations shown in Figures 1 and 3 are used, the basic channel unit can be readily switched from the single-channel voice/narrow-band data mode to the wide-band data mode by switching out the PCM codec, voice detector, and input timing source at the transmitter and by switching in the convolutional encoder and external timing circuits

* This assumes that the same SPADE synchronizer will be used and the resultant bit rate, with unique word stuffing, will be $40.8 \times 8/7$ kbps.

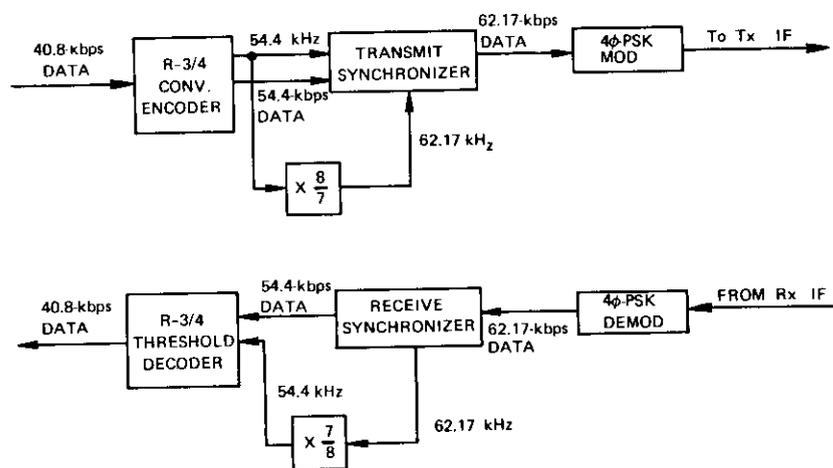


Figure 3. Wideband Data Channel Unit

at the transmitter and the threshold decoder at the receiver. The basic unit modem, memory, and synchronization requirements remain the same.

SMALL EARTH STATION APPLICATIONS

The SPADE system, using an INTELSAT IV global-beam transponder with standard INTELSAT earth stations ($G/T = 40.7 \text{ dB}/^\circ\text{K}$) and 4-phase coherent PSK modulation, will achieve a maximum capacity of 800 voice channels. For voice detection and a system margin equal to 4.2 dB, the results suggest that the capacity is simultaneously bandwidth and power limited. Hence, if earth stations with smaller G/T values are considered, the capacity will be power limited and there will be a reduction in system capacity.

One approach to minimize the power-limited condition is to use error coding (such as convolutional codes discussed earlier) to provide a trade-off of the excess available bandwidth to reduce the net per-channel required power.

Figure 4 shows typical results obtained by using the rate-3/4 convolutional encoder/threshold decoder discussed earlier and a rate-1/2 codec. Both codecs are relatively simple and efficient implementations, and represent low additional cost to the existing channel unit. Typically the rate-1/2 codec consists of about 17 integrated circuit packages at a total component cost of \$50.00, and the rate-3/4 codec consists of 50 integrated circuit packages at a total component cost of \$250.00.

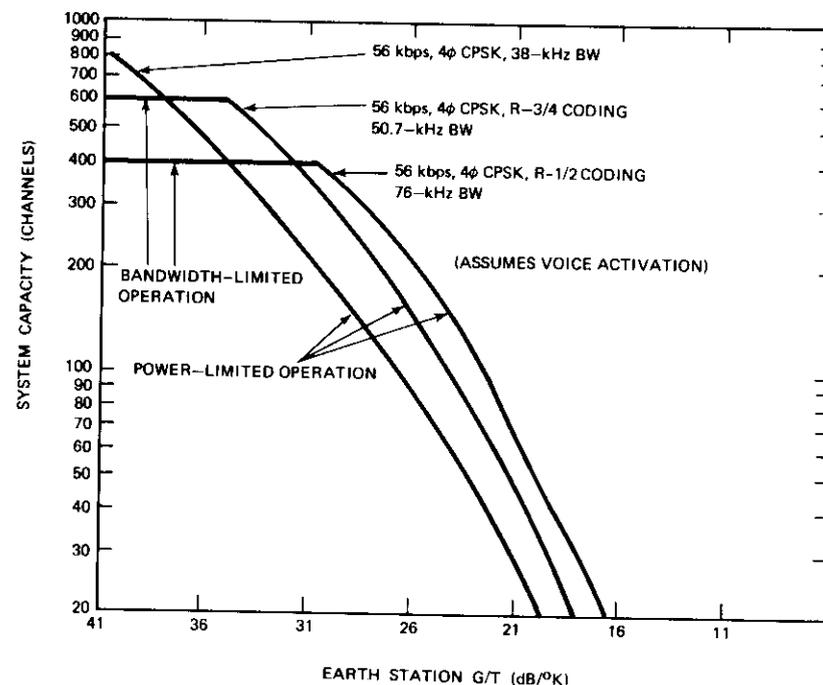


Figure 4. System Capacity vs. Earth Station G/T —Fixed-Input Bit Rate, Variable Bandwidth

Codes with other rates as well as different types of codecs, such as sequential and Viterbi decoders, also can be used. The ultimate choice will depend on the system efficiency and cost trade-offs considered. In general, the latter codecs (including block code implementations) involve more complex and costly implementations, with preliminary cost estimates ranging from 10 to 100 times greater than those for threshold decoder implementations.

The assumed system conditions used in Figure 4 include a global-beam INTELSAT IV transponder, minimum up-link margin of 2.5 dB, minimum down-link margin of 3.0 dB, threshold P_{RF} of 1×10^{-4} , saturation flux density, W_s , of $-67 \text{ dBW}/\text{m}^2$, and voice activation. The capacity calculations include the effect of intermodulation distortion and operation through a nonlinear transponder. The results are optimized for each G/T value considered. The results show that, to improve system capacity for G/T values less than $39 \text{ dB}/^\circ\text{K}$, rate-3/4 coding should be used, and for G/T values less than $32 \text{ dB}/^\circ\text{K}$, rate-1/2 coding

Another approach that can be used to increase system capacity for small earth terminal operation is to implement a digital voice channel that operates at a lower sampling rate than the standard SPADE 56-kbps PCM channel. Experiments using delta modulation have shown that excellent to reasonably good voice circuits can be maintained at sampling rates from 42 kbps down to 18 kbps. In addition, since delta modulation experiences a "graceful degradation" as the channel error rate worsens, the threshold P_{BE} can be defined as 1×10^{-3} .

Delta modulation was selected as the voice encoder because of recent improvements achieved by Philips Laboratories [6] in implementations which operate extremely well at low sampling rates. Several units have been tested at COMSAT Laboratories and have been found to work satisfactorily at sampling rates as low as 18 kbps. A recent report to the CCIR [7] also has shown that delta modulation, operating at 18 kbps, provides better performance (comparing output articulation index to channel carrier-to-noise density ratio) than a comparable 18-kbps PCM system.

The smaller bandwidth utilization of the lower sample rate voice channels can be used advantageously by adding suitable error coding until the resultant bandwidth is equal to the standard SPADE 56-kbps channel. This addition of coding provides further improvement in the per-channel power requirements. Figure 5 shows the combined improvement in system capacity that can be obtained by using rate-1/2 and rate-3/4 coding in conjunction with 28-kbps and 42-kbps sampled voice, respectively. The results shown in Figure 5 are based on a threshold P_{BE} of 1×10^{-4} , although, as already discussed for delta modulation, threshold P_{BE} can be lowered to at least 1×10^{-3} , resulting in a further improvement.

EXPERIMENTAL RESULTS

In order to demonstrate the feasibility of implementing a channel unit with the operational flexibility which has been discussed in this paper, an existing experimental channel unit has been configured to operate in the following three modes:

- 56-kbps digitally encoded voice (PCM or delta modulation) or data, 2-phase coherent PSK;
- 28-kbps digitally encoded voice (delta modulation) or data, 2-phase coherent PSK, with rate-1/2 convolutional encoding; and
- 40.8/42-kbps digitally encoded voice or data, 2-phase coherent PSK, with rate-3/4 convolutional encoding and threshold decod-

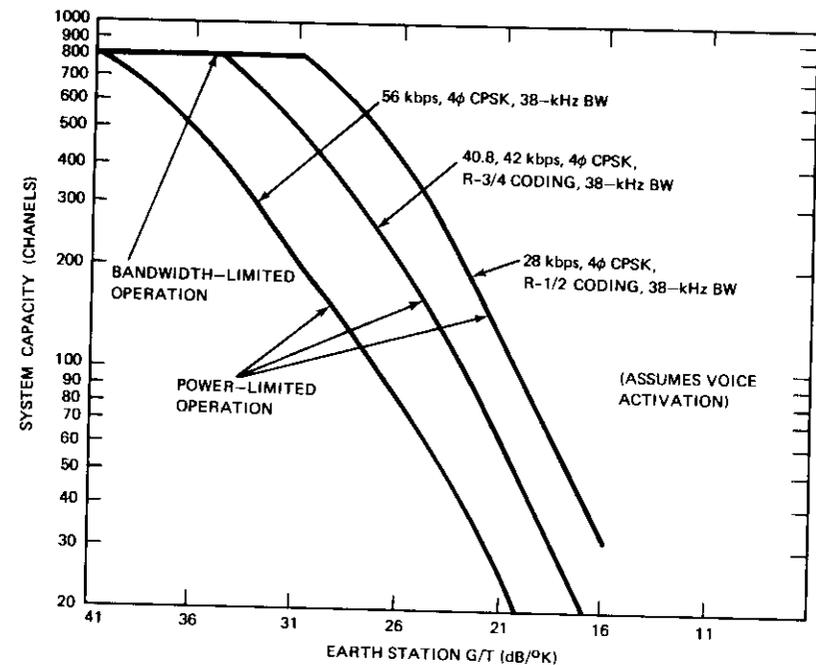


Figure 5. System Capacity vs. Earth Station G/T —Fixed Bandwidth, Variable-Input Bit Rate

The channel noise bandwidth for each of the three modes of operation is 80 kHz, since 2-phase CPSK was used in the experimental implementation. The transmitted bit rate is approximately 64 kbps. Although the results presented are for 2-phase CPSK, they can be readily extrapolated to determine the expected performance of 4-phase CPSK.

Figure 6 shows the measured performance curves of the three operational modes, each of which occupies the same channel noise bandwidth. The results show the significant improvements gained by using coding techniques in conjunction with single-channel PSK transmissions.

The criteria used to select the codecs for this experiment are as follows:

- the unit must be relatively simple to implement,
- the cost of implementation must be low (since the installation is required on a per-channel-unit basis), and
- the codec must provide reasonable improvement in E_b/N_0 .

In Figure 6, the results using coding can be improved by approximately 1 to 3 dB by using other decoding algorithms, such as sequential decod-

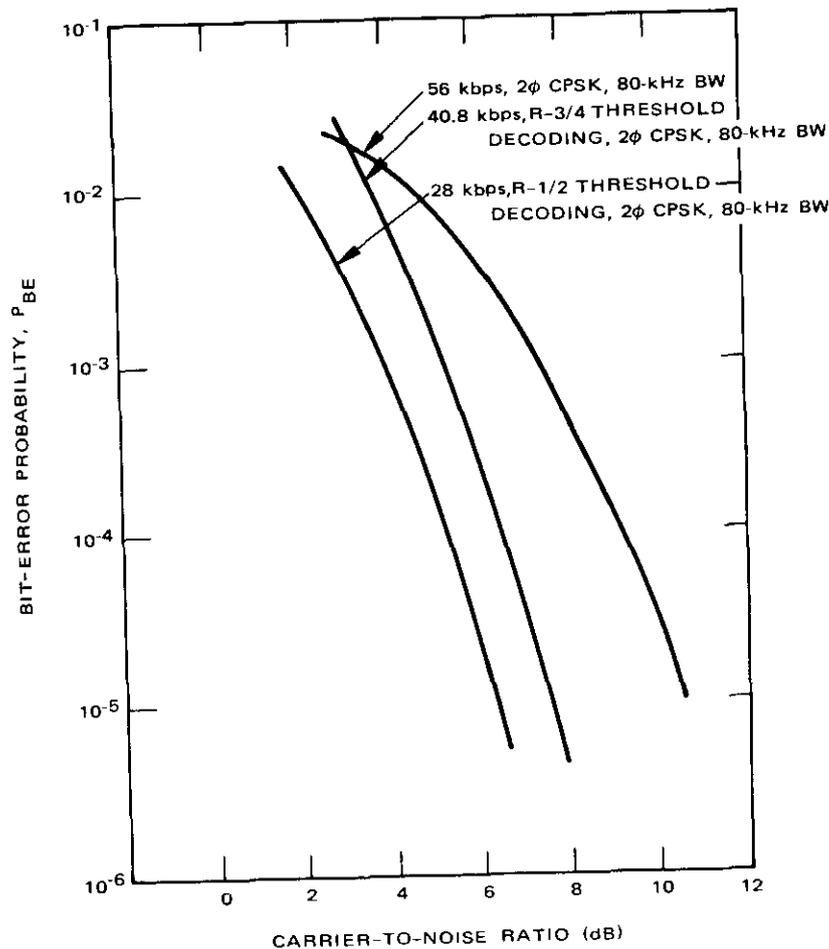


Figure 6. Measured System Performance Over INTELSAT III, F-6

ing or a Viterbi decoder. Other codes that also can be considered include block codes with or without "soft decision." At present, all of these alternative implementations result in considerably more complex and expensive implementations than threshold decoders, and one must evaluate the overall need for additional improvement in system performance versus the additional cost impact on the system.

The field tests were conducted using the COMSAT 16-foot terminal at L'Enfant Plaza, Washington, D.C., where a test signal was transmitted to the Fram. West Virginia, earth station via the INTELSAT III, F6

satellite and then looped back to the 16-foot terminal. In addition to the measured performance results shown in Figure 6, subjective voice quality evaluations were made, particularly in regard to the 28-kbps voice channel. The tests demonstrated that a good quality voice channel could be achieved by using 28-kbps delta modulation, and that the resultant bandwidth reduction allowed the use of rate-1/2 coding to achieve a further reduction in required threshold C/N. Thus, a good quality voice service can be provided, for example, to small earth terminals at significantly less satellite e.i.r.p. than is required for a standard 56-kbps uncoded voice channel.

The other significant test result verified the ability to transmit and receive 40.8-kbps data over a single SPADE channel unit operating at improved bit-error-rate performance (due to rate-3/4 coding) and using the same power and bandwidth as standard PCM voice.

CONCLUSIONS

A digital satellite communications system such as SPADE provides a broad range of potential services by using an efficient and modular design. The basic voice channel unit can transmit telephony, narrow-band data, and telegraphy without any modifications and, depending on power-bandwidth trade-off factors, can also use coding techniques to improve performance. With certain modifications, wide-band data can be transmitted over the same channel unit and occupy the same or less bandwidth than a single digitally encoded voice channel. In addition, wide-band data transmissions can be significantly improved by utilizing convolutional encoding techniques. For special applications, low-bit-rate voice encoding can be used to provide economical service to small earth stations.

REFERENCES

- [1] A. M. Werth, "SPADE: A PCM FDMA Demand Assignment System for Satellite Communications," *IEEE 1970 International Conference on Communications*, San Francisco, Calif., June 1970, pp. 46-22-46-32.
- [2] J. G. Puente and A. M. Werth, "Demand Assignment Service for the INTELSAT Global Network," *IEEE Spectrum*, Vol. 8, No. 1, January 1971, pp. 56-69.
- [3] E. R. Cacciamani, "The SPADE Concept Applied to a Network of Large and Small Earth Stations," *AIAA 3rd Communications Satellite Systems Conference*, Los Angeles, Calif., April 1970, AIAA Paper No. 70-420.
- [4] R. B. McClure, "The Effect of Earth Station and Satellite Parameters on the SPADE System," *Conference on Earth Station Technology* (IEE Conference Publication No. 72), London, Oct. 1970, pp. 17-22.

- [5] E. R. Cacciamani and C. J. Wolejsza, "Phase Ambiguity Resolution in a Four-phase PSK Communications System," accepted for publication, *IEEE Transactions on Communication Technology*, COM-19, No. 6, Dec. 1971.
- [6] J. A. Greefkes and F. de Jager, "Continuous Delta Modulation," *Philips Research Reports*, Vol. 23, No. 2, April 1968, pp. 233-246.
- [7] S. J. Campanella and J. A. Sciulli, "A Comparison of Voice Communication Techniques for Aeronautical and Marine Applications," submitted for publication, *COMSAT Technical Review*.

ACKNOWLEDGMENTS

The author would like to thank A. Dohne, W. Schmidt, and A. Walker for their contributions in deriving the results established in this paper, and J. Beyer and R. Briskman for their cooperation and use of the 16-foot terminal.



Eugene R. Cacciamani received a B.E.E. degree in 1958 from Union College and an M.E.E. degree in 1962 from The Catholic University of America. From 1958 to 1962 he served as an officer in the U. S. Air Force, working on research and development of special purpose digital computer technology. From 1962 to 1965 he was a senior member of the Technical Staff at RCA where he worked on the development of information storage and retrieval systems and command and control systems. He joined COMSAT in 1965 and is head of the Modulation Systems Section at COMSAT Laboratories. As program manager of the DICOM project he is responsible for the development of new digital voice and data transmission techniques as applied to small earth station technology.

Index: communication satellites, digital techniques, error correction codes, synchronism, time division multiplexing

APPLICATIONS OF ERROR-CODING TECHNIQUES TO SATELLITE COMMUNICATIONS

W. W. WU

ABSTRACT

Applications of coding are presented which demonstrate that error coding provides significant improvements in the performance of communication satellite links. Specifically, four types of high-rate codes are illustrated and computer simulation results are discussed. For synchronization, a coding algorithm is suggested for systematically obtaining a set of synchronizable words. For message control, a block codec design is described and the experimental results are examined. For practical shift register sequence generation, a new method is shown to generate long sequences from modular short sub-generators.

INTRODUCTION

As digital transmission systems come into operational use in communications satellites, the potential of coding technology for reliable information transmission is evident for the following reasons:

a. A telephone call reaching the wrong destination through a satellite link is clearly different from a telephone call reaching a wrong number across the street. It is important to establish circuits correctly so that channel allocation and identifications will be maintained. Transmission of data and TV requires lower error rates than telephone transmission. As computer communications through satellites increase, an almost error-free transmission link is essential. Since messages without synchronization have no meaning, special synchronization techniques have been developed from coding theory.

b. The channel noise of satellite links in normal operating conditions can be represented by a stationary Gaussian random process. For single access with PSK modulation, the link can be approximated by a binary

This paper is based upon a part of the coding work performed at COMSAT Laboratories under the sponsorship of the International Telecommunications Satellite Consortium (INTELSAT). Any views expressed in this paper are

symmetric channel. These channel models are most suitable for random error correction. Significant improvements can even be gained by coding procedures which partially match channel characteristics.

c. It is known from information theory that the rate-distortion bound, a lower bound on E_b/N_o , can be closely approached without recourse to feedback only if error coding is used. Thus, it is impossible to rely on the improvement of conventional modems alone to achieve the rate-distortion bound for the satellite link channel model.

d. Due to both power and bandwidth constraints in communications satellite systems, it is desirable to transmit information at a rate as close as possible to the channel capacity. It will be shown that high-efficiency (high-rate) codes to achieve error-free transmission with reasonable implementation complexity are economically feasible for such links.

In this paper, some specific applications of coding are presented which demonstrate that error coding can provide significant improvements in information transmission through satellite links. First, known codes are reviewed and their differences in terms of encoding processes, error exponents, and speeds of operation are examined. Secondly, high-rate codes are presented. These codes require a minimal amount of additional bandwidth for parity checking. Performance curves and computer simulation results are used to document their effectiveness. Next, a block code which enhances the reliability of the control message in an experimental 50.0-Mbps TDMA system is described. The design of the codec and the related experimental evaluation are illustrated. Then, the procedures leading to the selection of a set of synchronizable code words with prescribed autocorrelation and cross-correlation properties are given. Finally, a simple and new generalization of shift register sequence generators is brought forth in terms of a theorem. Comparisons with the classical solutions and some practical consequences and advantages are also examined.

FUNDAMENTAL CODES

Basically, there are two types of parity check codes, block and convolutional. Figures 1a and 1b are generally encoding diagrams with corresponding notations of these two types of codes. Block codes [1], [2] are described as ideals generated by the residue class containing $g(x)$ in the algebra of polynomials modulo $f(x)$ [3]. If $f(x) = x^{nb} - 1$, the code is said to be cyclic. It can be decoded either with majority logic [4], [5] or algebraically [3], [6].

Convolutional codes [7] can be represented in any of the following

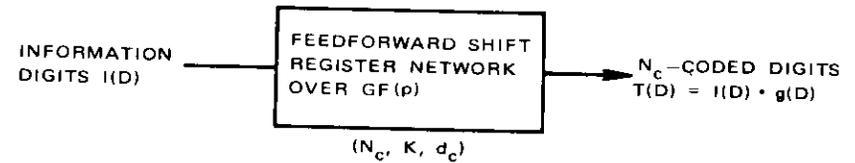


Figure 1a. General Convolutional Encoder

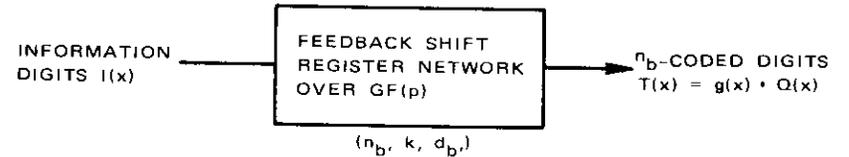


Figure 1b. General Block Encoder

sequential method [8], [9], threshold techniques [4], or a maximum likelihood algorithm [10]. Both block codes and convolutional codes are generated from their respective generator polynomials, $g(x)$ and $g(D)$.

For binary block codes, encoding begins with a k -bit information sequence, $I(x)$, which is divided by $g(x)$ as $I(x)$ shifts through the encoder. The division process results in a quotient polynomial, $Q(x)$, and a remainder, $R(x)$. The block coded sequence is

$$T(x) = x^{nb-k} I(x) + R(x) = g(x) \cdot Q(x)$$

which is n digits long.

For convolutional codes, the information sequence is denoted as $I(D)$, while the coded transmitted sequence is $T(D) = I(D)g(D)$. In general, the power series, $T(D)$, extends as long as $I(D)$ is continuous. Depending on the decoding method, the constraint lengths of convolutional codes are defined as the truncated number of coded digits which are affected by a given information digit. In other words, it is the code word length that guarantees minimum distance.

For block codes, there is no memory between code blocks. For convolutional codes, each bit is encoded with memory and delivered continuously. Block codes are generated from feedback shift registers which result in limitations of the speed of the codec operation due to the accumulative propagation delays from the feedback signals. Convolutional codes are implemented by feed-forward shift registers, which can operate at the speed of individual logic elements.

One of the main results of the coding theory is the reliability function, $E_s(R)$ [11]. The error probability at the decoder output for a code of

length N_r is given by $P_e = \exp[-N_r E_x(R)]$. Figure 2 compares the $E_x(R)$'s for the sphere packing $E_{sp}(R)$, random coding $E_r(R)$ [11] of block codes, systematic $E_{sc}(R)$ [12], and nonsystematic $E_{nc}(R)$ [10] convolutional codes. The $E_x(R)$'s are compared for a channel transition probability equal to 5×10^{-4} . These reliability function bounds yield a measure of the effectiveness of specific coding techniques.

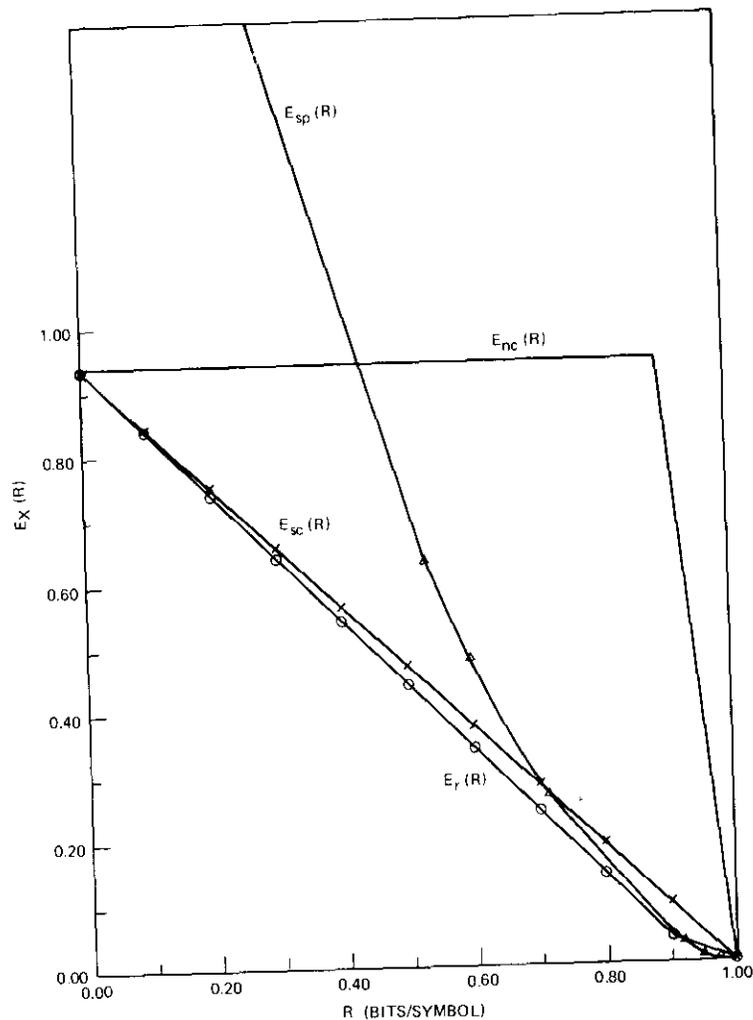


Figure 2. A Comparison of $E_x(R)$ at $p = 5 \times 10^{-4}$ (BSC)

HIGH-EFFICIENCY CODING

Most coding techniques involve the idea of parity checking, i.e., a mapping from k (or K) digits of information to n_b (or N_c) digits of coded blocks. Since $n_b - k$ (or $N - K$) is greater than 0, additional digits are required. Due to the bandwidth limitation of the satellite system, large amounts of redundancy (parity check digits), which would reduce the error rate, cannot be tolerated. Such a constraint differs from coding application constraints for deep-space information transmission. As a result, high-rate codes to achieve low error probability are almost a necessity. Unfortunately, most coding applications have been concerned with low-rate codes. In the following paragraphs, four types of high-rate codes are described for high data-rate (30- to 70-Mbps) digital transmission systems.

Long Block Codes

Although long BCH (Bose-Chaudhuri-Hocquenghem) codes are asymptotically inefficient [13] (i.e., the ratio of the minimum distance, d , to the block length, n_b , approaches zero as n_b increases), BCH codes are still quite powerful for practical values of n_b , and R . For example, the BCH code with $n_b = 1023$, $k = 943$, $d = 17$, and $R = 0.92$ can convert an error rate of 5×10^{-4} to 4×10^{-9} . The bit-error-probability performance in an additive Gaussian channel in terms of E_b/N_0 is shown in Figure 3. Modern BCH codec implementation is eased by combining Berlekamp-Massey's iterative algorithm [6], [14] for obtaining the error-locating polynomial, Chien's method [15] for searching for its roots, and Burton's procedure [16] for eliminating the inversion step.

Iteration of Single-Error-Correcting Block Codes (ISECB)

For binary symmetric channels, Elias' sequences of iterated single-error-correcting (SEC) codes yield a probability of error which can be made to approach 0, while the rate can be bounded away from 0 as the length of the codes approaches infinity [7]. For finite and reasonable code word lengths, high-rate codes can be obtained. The trick is to use high enough rates for the component codes so that, after multiplication, the resultant code rate still can be considered to be efficient.

SEC codes can be generated from a primitive generator polynomial, $g(x)$, of arbitrary degree m_b . The necessary and sufficient condition to ensure the primitivity of $g(x)$ is that it cannot divide $x^{n_b} - 1$ for n less than $2^{m_b} - 1$. SEC codes can be characterized by the following parameters:

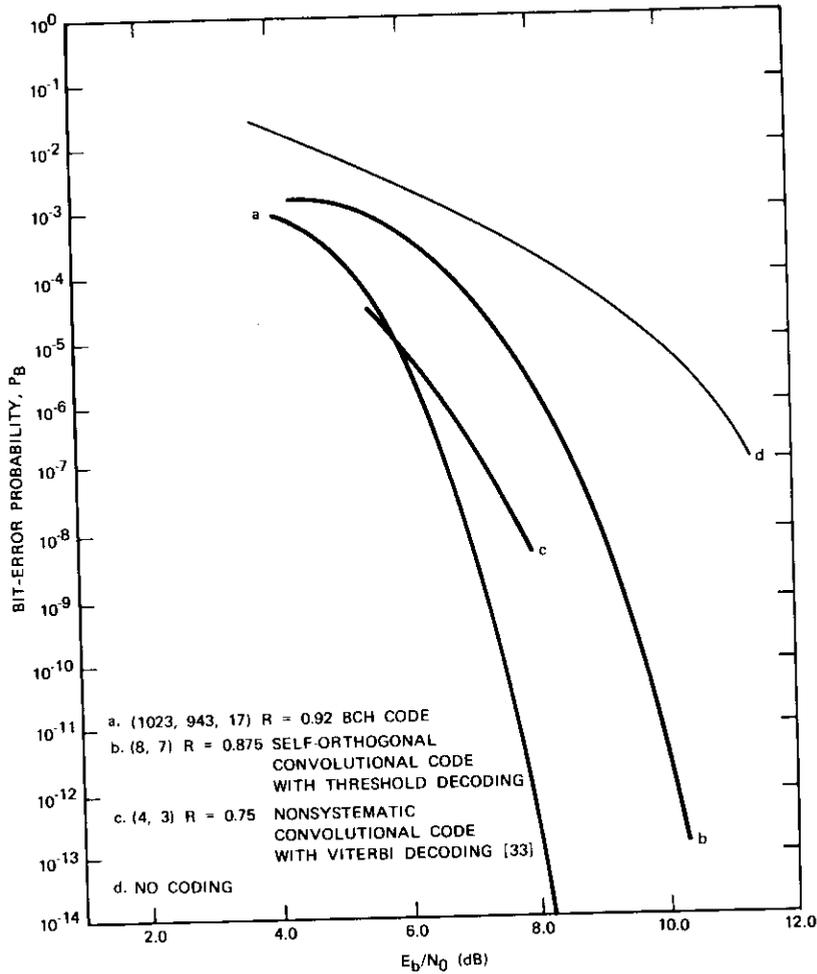


Figure 3. Singular High-Rate Codes

block length: $n_b = 2^{mb} - 1$
 number of parity check digits: $m_b = n_b - k$
 number of information digits: $k = 2^{mb} - m_b - 1$
 rate: $R = \frac{k}{n_b} = 1 - \frac{m_b}{2^{mb} - 1}$ (1)

must approach 0. Since 2^{mb} increases faster than m_b , it is not necessary to have a very large value of m_b for R to approach unity. It is easy to verify that, for $m_b = 11$, $R = 0.99$. For the iterations of s SEC codes, the overall code rate is

$$R_b = \prod_{i=0}^{s-1} \left(1 - \frac{m_{b_i}}{2^{m_{b_i}} - 1} \right) \quad (2)$$

where $m_{b_i} = n_{b_i} - k_i$.

In general, for s iterations of certain SEC codes, the overall code bit-error probability is

$$P_B = f_s \{ f_{s-1} [\dots f_1(p)] \} \quad (3)$$

In the case of BSC with 3-dimensional codes,

$$P_B = \frac{1}{n_1} \sum_{i=2}^{n_1} \binom{n_1}{i} \left\{ \frac{1}{n_2} \sum_{j=2}^{n_2} \binom{n_2}{j} \left[\frac{1}{n_3} \sum_{k=2}^{n_3} \binom{n_3}{k} p^k q^{n_3-k} \right]^j \cdot \left[1 - \frac{1}{n_3} \sum_{k=2}^{n_3} \binom{n_3}{k} p^k q^{n_3-k} \right]^{n_2-j} \right\} Q^{n_1-i} \quad (4)$$

where $f_i(\cdot)$ = the decoding bit-error probability of the i th code

$$Q = 1 - \{ \cdot \}$$

$\{ \cdot \}$ = the quantity in braces in equation (4)

$$q = 1 - p$$

p = BSC transition probability.

The implementation of ISECB codes is straightforward [17], [18] and it can often be simplified by interleavers [19]. Except for storage requirements, few calculations are needed for the encoding and decoding. The performance of the ISECB codes listed in Table 1 is shown in Figures 4 and 5.

TABLE 1. ISECB CODES

Code	n_1	n_2	n_3	R
ISECB ₁	2047	4095		0.99
ISECB ₂	255	511		0.95
ISECB ₃	31	63		0.76
ISECB ₄	2047	4095	8191	0.99
ISECB ₅	255	511	1023	0.94
ISECB ₆	31	63	127	0.72

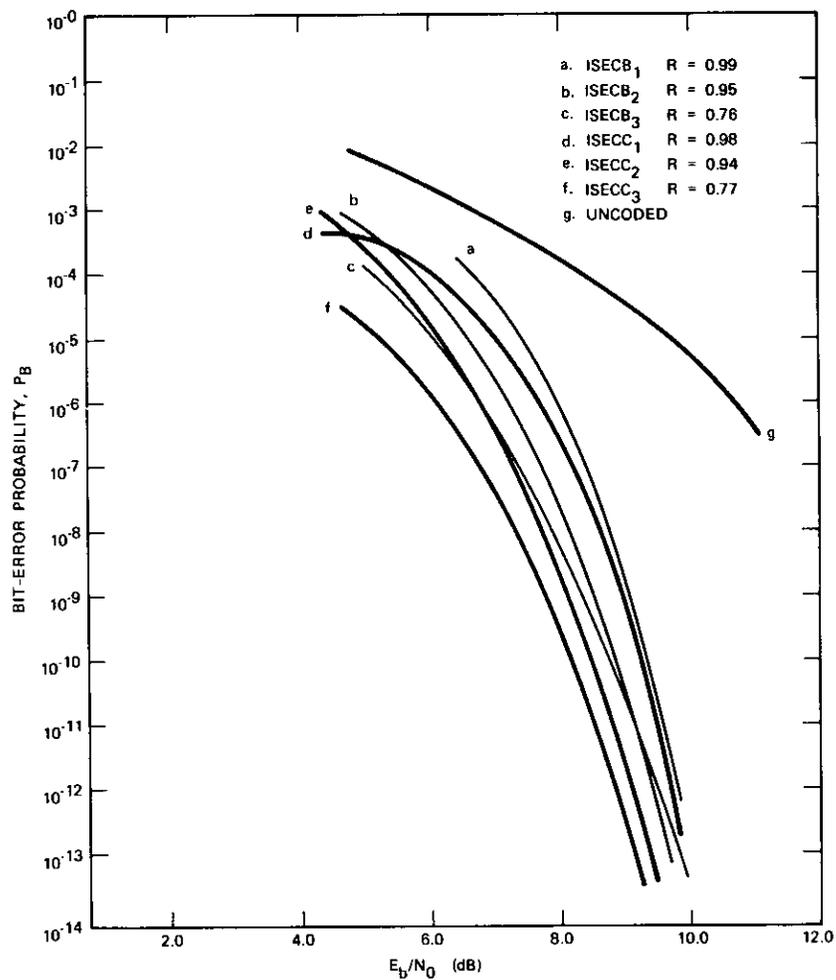


Figure 4. Performances of Two-Dimensional Iterative Codes in Additive Gaussian Channel

Self-Orthogonal Convolutional (SOC) Codes

Orthogonalizable codes are known to be threshold decodable [4], [5]. Orthogonality in threshold decoding refers to the use of a set of syndrome sequences (parity check sums) to check on an error digit. The error digit is checked by each of the syndrome sequences in the orthogonalized set and no other error digit is checked by more than one such sequence in the set.

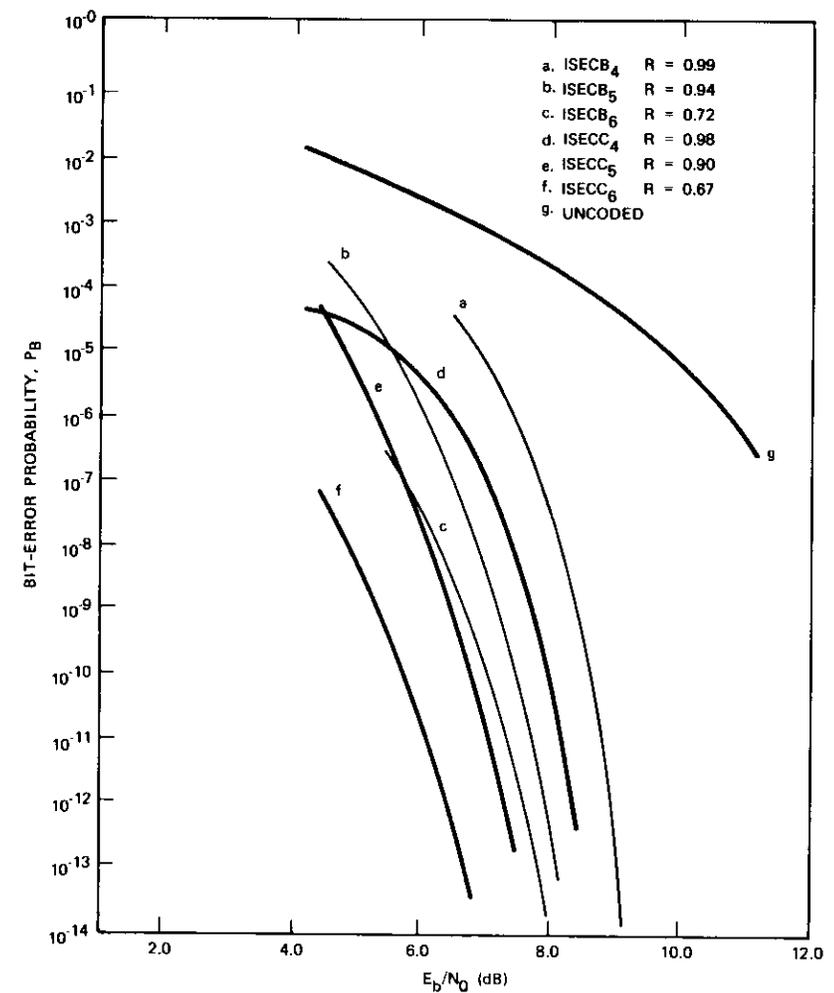


Figure 5. Performances of Three-Dimensional Iterative Codes in Additive Gaussian Channel

A binary (N_c, K) convolutional code is said to be J orthogonalizable if it is possible to form J orthogonal parity check sums on an error digit. It has been shown by Massey [4] that, if d_c , the minimum distance of the convolutional code, is equal to $J + 1$, the code is said to be completely orthogonalizable and capable of correcting $J/2$ or fewer errors within the constraint length. A convolutional code is said to be self-orthogonal if and only if the set of syndrome digits is orthogonal on the block zero

error digit, e_0 . An $(N_c, N_c - 1)$ self-orthogonal convolutional code is specified by a set of $N_c - 1$ distinctive subgenerators:

$$G(D) = \{G_i(D) : i = 1, \dots, N_c - 1\} \quad (5)$$

The set of $G(D)$ may be systematically obtained by means of difference sets [20], [21]. A difference set, D , is a collection of $q + 1$ distinct residues, d_1, \dots, d_{q+1} modulo M , for which the congruence relation,

$$d_i - d_j \equiv a \pmod{M} \quad (6)$$

has exactly λ distinct solution pairs, d_i, d_j , in D for every $a \not\equiv 0 \pmod{M}$, and i, j inside $D, i \neq j$, which implies that the set of residues of the difference modulo M are all non-zero. A necessary condition for the existence of D is that the parameters M, q , and λ must satisfy the condition:

$$q(q + 1) = \lambda(M - 1). \quad (7)$$

When $M - 1 \div q + q^2, D$ is said to be perfect. Robinson [20] tabulated a set of self-orthogonal convolutional codes from perfect difference sets.

As a first example, from the set of first triangular rows, (3,12,4) (8,9,1) (6,5,2) is the (4,3), $J = 4$, self-orthogonal code. Through addition by pairs, the half-T elements are obtained; i.e.,

3	12	4	8	9	1	6	5	2	
15			17			11			
19			18			13			(8)

The columns of equation (8) determine the set of subgenerator polynomials:

$$G(D) = \begin{cases} G_1(D) = 1 + D^3 + D^{15} + D^{19} \\ G_2(D) = 1 + D^8 + D^{17} + D^{18} \\ G_3(D) = 1 + D^6 + D^{11} + D^{13} \end{cases} \quad (9)$$

As a second example, the (8,7) threshold decodable code is from the

(2,	6,	24,	56,	54)	
(3,	16,	33,	26,	68)	
(11,	1,	50,	23,	46)	
(21,	4,	14,	43,	44)	
(5,	15,	27,	37,	60)	
(58,	38,	10,	7,	28)	
(41,	36,	31,	9,	13)	(10)

which was obtained by Klieber [22]. The generator polynomial, $G(D)$,

can be similarly shown to consist of the following set of subgenerator polynomials:

$$G(D) = \begin{cases} G_1(D) = 1 + D^2 + D^8 + D^{32} + D^{88} + D^{142} \\ G_2(D) = 1 + D^3 + D^{19} + D^{52} + D^{78} + D^{146} \\ G_3(D) = 1 + D^{11} + D^{12} + D^{62} + D^{85} + D^{131} \\ G_4(D) = 1 + D^{21} + D^{25} + D^{39} + D^{82} + D^{126} \\ G_5(D) = 1 + D^5 + D^{20} + D^{47} + D^{84} + D^{144} \\ G_6(D) = 1 + D^{58} + D^{96} + D^{106} + D^{113} + D^{141} \\ G_7(D) = 1 + D^{41} + D^{77} + D^{108} + D^{117} + D^{130} \end{cases} \quad (11)$$

The parity matrix is

$$H_\Delta = [\Delta_1 \Delta_2 \dots \Delta_7] \quad (12)$$

where each Δ_i ($i = 1, \dots, 7$) is a 146×146 matrix whose elements are in accordance with $G_i(D)$ of equation (11). The syndrome equation is given in terms of the error vector, \bar{e} ; i.e.,

$$[\bar{S}] = [\Delta_1 : \Delta_2 : \dots : \Delta_7 : I] [\bar{e}] \quad (13)$$

where I is the identity matrix of the same dimension as Δ .

Both examples were simulated on the IBM 360/65 computer. The simulation of the (8,7) code is shown in Figure 6. The source messages are 3-bit-per-sample digitized television signals. For a high rate, Massey's $(1 - R) n_A$ encoding configuration [4] is less complex than Wozencraft and Reiffen's Rn_A configuration [8]. However, in the encoding replica part of the decoder, the Rn_A configuration is preferred because the use of the $(1 - R) n_A$ configuration requires additional memory buffers in decoding. The circuits for $G(D)$ and $G'(D)$ are shown in Figures 7 and 8. The syndrome shift register connection is shown in Figure 9. The seven threshold circuits (TH-CKT) in the codec block diagram are identical. One of the TH-CKT's, which consists of four 3-input binary adders, is shown in Figure 10. The six inputs ($J = 6$) from the syndrome shift register correspond to the six non-zero coefficients (or positions) of the set of subgenerator polynomials.

The guaranteed performance of the (8,7) codec is shown in Figure 3. Computer runs indicate that the decoder often corrects more errors than it guarantees. Another encouraging result is that, with the code constraint length $(1 + m_c) N_c = 147 \times 8 = 1176$, no error propagation has been observed. This result shows that, with properly designed, hazard-free logic circuits, error propagation can be eliminated.

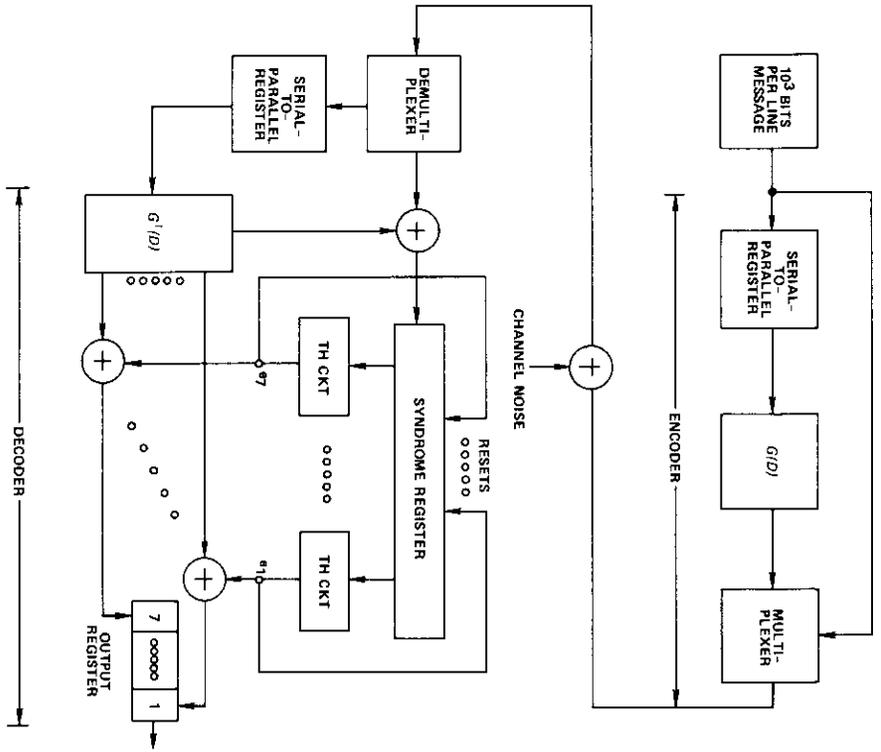


Figure 6. Simulated Rate-7/8 Convolutional Code

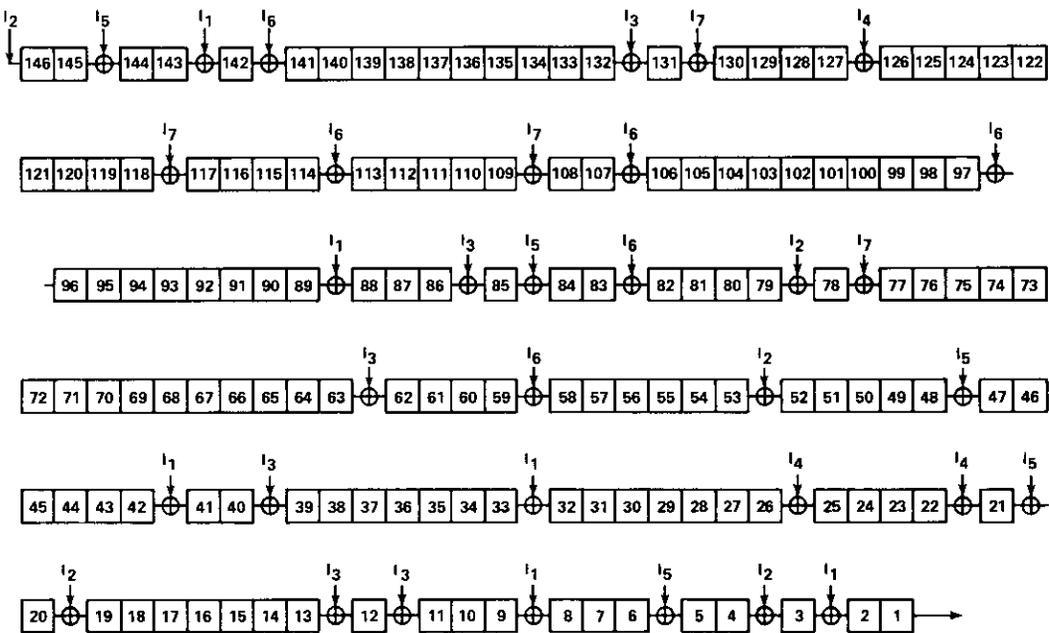


Figure 7. Detailed Connections of Massey's Encoder Configuration. $I_1 \dots I_7$ are Input Digits.

Convolutional Product Codes

In this section, the iteration of single-error-correcting convolutional (ISECC) codes is discussed. A type of convolutional single-error-correcting code shown by Wyner and Ash [23] has the following properties:

$$\begin{aligned}
 N_c &= 2^{m_c} \\
 K &= 2^{m_c} - 1 \\
 n_c &= (1 + m_c) 2^{m_c} \\
 m_c &= N_c - K \\
 R &= \frac{K}{N_c} = 1 - 2^{-m_c}. \tag{14}
 \end{aligned}$$

Extending Elias' technique for block codes to convolutional codes results in the iteration of a number of Wyner and Ash codes to obtain a product code of rate R_c , where

$$R_c = \prod_{i=0}^{s-1} [1 - 2 \exp(-m_{c_i})]. \tag{15}$$

A comparison of the rates of equations (2) and (15) shows that the resultant convolutional code provides a higher rate for the same amount of parity check digits per block. Freiman and Robinson [24] have evaluated the equivalent performance of the noniterated single-error-correcting codes (block and convolutional) according to the criterion of mean time to uncorrectable error. For iterative codes, the difference becomes significant.

The bit-error-probability performances of the ISECC codes listed in Table 2 in additive Gaussian noise are also shown in Figures 4 and 5. It is easy to see that, for $m_c = 3$, the $R = 0.77$, 2-dimensional ISECC code outperforms the rate 0.76 ISECB codes of the same dimension.

The unavoidable long decoding time delay of ISECB and ISECC codes

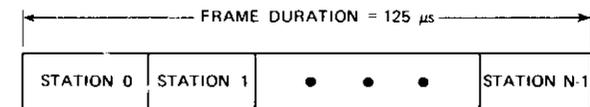
TABLE 2. ISECC CODES

Code	n_c	m_c	R
ISECC ₁	(1024) ²	7	0.98
ISECC ₂	(192) ²	5	0.94
ISECC ₃	(32) ²	3	0.77
ISECC ₄	(1024) ³	7	0.98
ISECC ₅	(192) ³	5	0.90
ISECC ₆	(32) ³	3	0.67

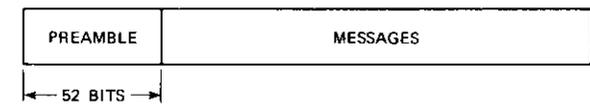
is not a handicap for 1-way transmission. The memory requirements of both types of codes can be significantly reduced by interleavers.

CODING FOR CONTROL MESSAGES

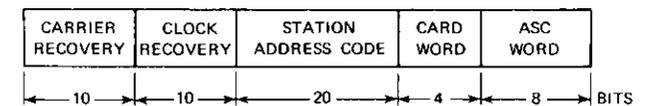
In June 1970, a 700-voice-channel PCM/PSK/TDMA digital system was successfully tested through an INTELSAT III satellite over the Pacific Ocean. The baseband information was processed at 50.0 Mbps with links among Hawaii, Australia, and Japan. A description of the system can be found in Reference [26]. The transmission format of the control subsystem is shown in Figure 11. Figure 11a shows the frame format formed from the participating ground station burst, Figure 11b shows the portion of each burst, Figure 11c shows the content of each



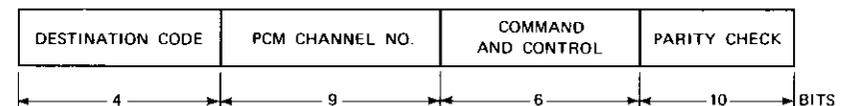
(a) FRAME FORMAT



(b) BURST FORMAT



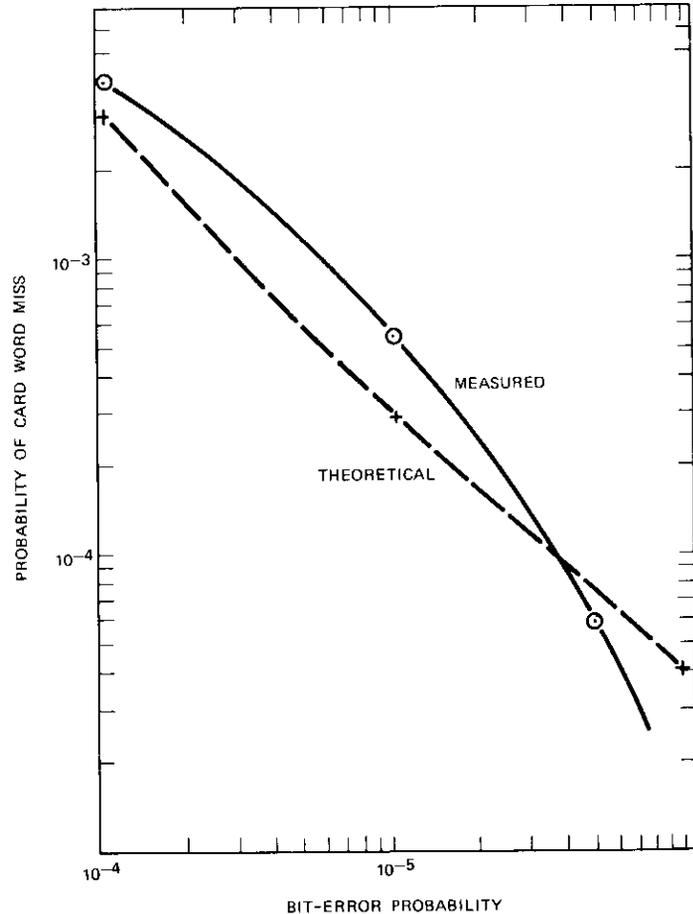
(c) PREAMBLE FORMAT



(d) CARD WORD FORMAT

preamble, and Figure 11d shows the channel allocation routing-destination (CARD) information, which is the control message and is coded with the (31,21,5) BCH code. The CARD word is multiplexed for transmission at four bits per frame. The decoder accumulates eight frames to complete a word.

To establish a telephone circuit, an error-free CARD word is needed for earth station identification and PCM channel reallocation. The field test result of the CARD word miss-detection probability versus bit-error probability is shown in Figure 12. During the test interval over the



satellite links, at no time was there evidence of CARD word information being incorrectly processed.

The following is a description of the codec design for error correction as well as for error detection. The encoder, with $g(x) = 1 + x^3 + x^5 + x^6 + x^8 + x^9 + x^{10}$, is shown in Figure 13, where $g(x)$ is the product of the following two minimal polynomials: $m_1(x) = 1 + x^2 + x^5$ and $m_3(x) = 1 + x^2 + x^3 + x^5 + x^6$. The powers of the primitive element α generate a $GF(2^5)$, as shown in the first column of Table 3. The k configuration is used in encoding to avoid additional CARD word memory requirements.

In decoding, the syndrome calculation is accomplished by multiplying the received code sequence, $r(x)$, by the parity check matrix of the code

TABLE 3. THE PARITY CHECK MATRIX OF TDMA-1 CODE

α^0	1	0	0	0	0	0	0	0	1	0	α^{21}
α^1	0	1	0	0	0	0	1	0	1	0	α^6
α^2	0	0	1	0	0	0	1	0	1	1	α^9
α^3	0	0	0	1	0	0	0	1	1	0	α^{12}
α^4	0	0	0	0	1	1	1	1	1	1	α^{15}
α^5	1	0	1	0	0	1	1	0	0	0	α^{18}
α^6	0	1	0	1	0	0	0	0	1	1	α^{21}
α^7	0	0	1	0	1	0	1	1	1	1	α^{24}
α^8	1	0	1	1	0	1	1	0	1	0	α^{27}
α^9	0	1	0	1	1	0	1	0	0	1	α^{30}
α^{10}	1	0	0	0	1	0	0	1	0	0	α^2
α^{11}	1	1	1	0	0	1	0	1	0	0	α^5
α^{12}	0	1	1	1	0	1	0	1	1	0	α^8
α^{13}	0	0	1	1	1	1	1	1	0	0	α^{11}
α^{14}	1	0	1	1	1	1	0	1	1	1	α^{14}
α^{15}	1	1	1	1	1	1	1	0	0	1	α^{17}
α^{16}	1	1	0	1	1	0	0	1	1	0	α^{20}
α^{17}	1	1	0	0	1	1	1	1	1	0	α^{23}
α^{18}	1	1	0	0	0	1	1	1	0	1	α^{26}
α^{19}	0	1	1	0	0	1	0	0	1	0	α^{29}
α^{20}	0	0	1	1	0	0	1	0	0	0	α
α^{21}	0	0	0	1	1	0	0	0	0	1	α^4
α^{22}	1	0	1	0	1	0	0	1	0	1	α^7
α^{23}	1	1	1	1	0	1	0	0	0	1	α^{10}
α^{24}	0	1	1	1	1	0	0	1	1	1	α^{13}
α^{25}	1	0	0	1	1	1	1	0	1	1	α^{16}
α^{26}	1	1	1	0	1	0	1	1	0	0	α^{19}
α^{27}	1	1	0	1	0	1	0	1	0	1	α^{22}
α^{28}	0	1	1	0	1	1	0	0	1	1	α^{25}
α^{29}	1	0	0	1	0	0	1	1	0	1	α^{28}
α^{30}	0	1	0	0	1	1	0	0	0	0	α^{31}

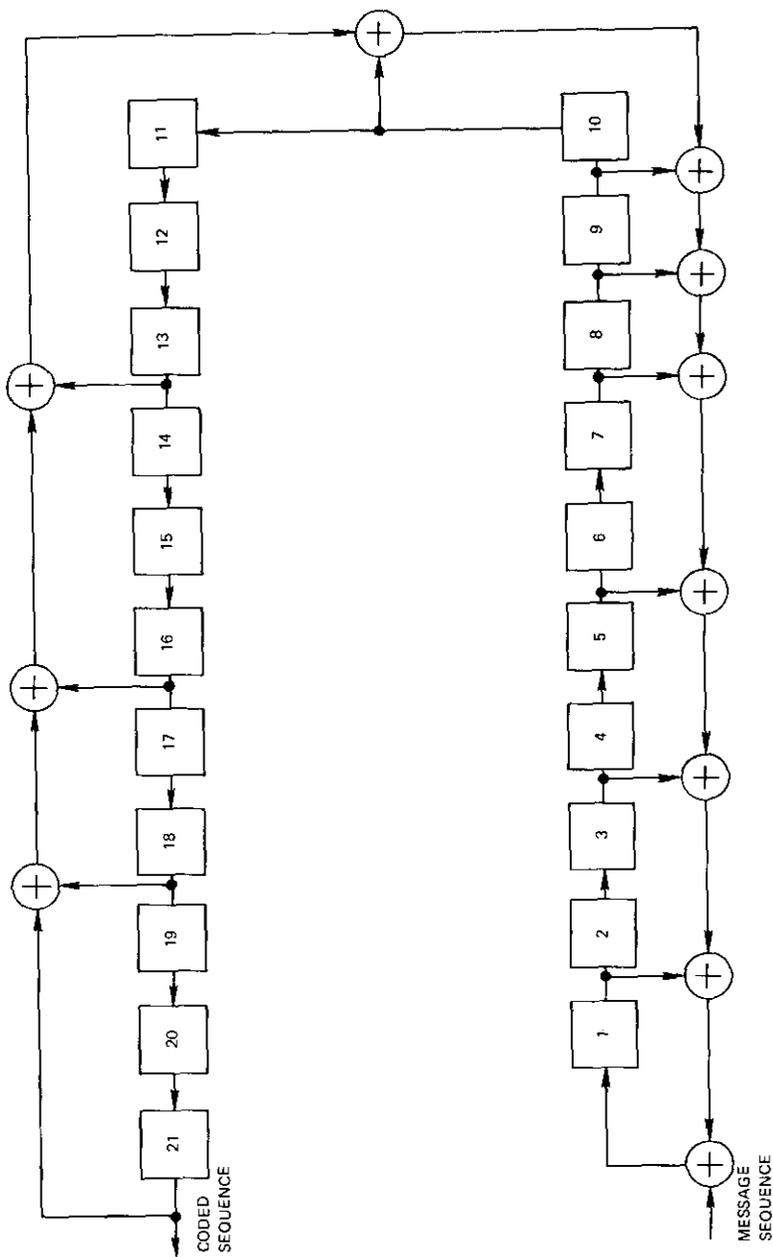


Figure 13. (31, 21, 5) BCH Encoder

(Table 3). Then the power sum symmetrical functions are obtained from the error locations. The roots of the error-locating polynomial are found by using Chien's search. A Delta matrix is formed from the power sums, and decoding circuits are synthesized from the elements of this matrix. If the control message is 21 digits of the 101010 . . . pattern, then the generated parity check is 1110001101. If two random errors have occurred due to channel noise, e.g., in the 7th and 26th digits, $r(x)$ differs from $T(x)$ in the addition of term x^5 and the deletion of x^{24} . Thus, the received block is 101010010101010101011110101101. The power sum of this code with errors that occurred in these positions is

$$S_1 = e(\alpha) = \alpha^6 + \alpha^{25} = (\alpha + \alpha^3) + (1 + \alpha^3 + \alpha^4) = 1 + \alpha + \alpha^4$$

$$S_3 = e(\alpha^3) = [e(\alpha)]^3 = [\alpha^6 + \alpha^{25}]^3 = 1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4. \quad (16)$$

In this case, the error-locating polynomial is simply

$$\sigma_2 + \sigma_1 x + x^2 = (x + \beta_1)(x + \beta_2).$$

From Newton's identity, $\sigma_2 S_1 = 3\sigma_3 + \sigma_1 S_2 + S_3$, where $\sigma_3 = 0$, $\sigma_2 = S_1^2 + S_3/S_1$, $S_1 = \sigma_1$, and $S_2 = S_1^2$, the values of σ_1 and σ_2 are: $\sigma_1 = 1 + \alpha + \alpha^4$ and $\sigma_2 = 1$.

Next, $\bar{\sigma}_1 = \alpha\sigma$ and $\bar{\sigma}_2 = \alpha^2\sigma^2$ are calculated as shown in Table 4. One is looking for rows such that $\bar{\sigma}_1 + \bar{\sigma}_2 = 1$. Since the 6th and 25th rows satisfy the previous condition, a "1" is provided to correct the 7th (α^{31-23}), and the 26th (α^{31-6}) digit inverses at the shift register output.

For the (31, 21, 5) BCH code, the Delta matrix reduces to

$$\Delta = S_1 + S_1^2 + S_1^3 + S_3. \quad (17)$$

The coefficients of S_1 , S_1^2 , S_1^3 , and S_3 are given by the following equations:

$$S_1 = a_0\alpha^0 + a_1\alpha + a_2\alpha^2 + a_3\alpha^3 + a_4\alpha^4 \quad (18a)$$

$$S_1^2 = (a_0\bar{a}_4 + a_3a_4)\alpha^0 + [a_4(a_1 + a_2) + a_3] \alpha$$

$$+ [a_3\bar{a}_4 + a_4(a_0 + a_2)]\alpha^2 + [a_3\bar{a}_4 + a_4(a_1 + a_2)]\alpha^3$$

$$+ (a_2 + a_3a_4) \alpha^4 \quad (18b)$$

$$S_1^3 = [a_0\bar{a}_4 + a_1(a_2 + a_3 + a_4) + \bar{a}_1a_2a_4]\alpha^0$$

$$+ [\bar{a}_0\bar{a}_4a_2 + \bar{a}_0a_3 + a_0a_1 + a_4(a_0 + a_2 + a_3)]\alpha$$

$$+ [a_0(a_1 + a_2 + a_4) + \bar{a}_0a_1a_4 + a_1a_2 + a_3a_4(a_0 + a_1 + a_2)]\alpha^2$$

$$+ [a_4(a_0\bar{a}_2 + a_1\bar{a}_3) + a_1 + \bar{a}_2a_3 + a_2\bar{a}_4 + a_3a_4]\alpha^3$$

$$+ [\bar{a}_0a_2 + (a_2 + a_3)(a_0a_4 + a_1) + a_3\bar{a}_2a_4]\alpha^4 \quad (18c)$$

$$S_3 = b_0\alpha^0 + b_1\alpha + b_2\alpha^2 + b_3\alpha^3 + b_4\alpha^4. \tag{18d}$$

The function of the decoding circuit is to respond whenever Δ equals 0. The circuit is synthesized with two levels of OR logic, and NAND gates. The coded experimental waveform is shown in Figure 14.

TABLE 4. THE ELEMENTARY SYMMETRIC FUNCTIONS OF THE ERROR-LOCATION NUMBERS

$\bar{\sigma}_1$	$\bar{\sigma}_2$
$1 + \alpha$	α^2
$\alpha + \alpha^2$	α^4
$\alpha^2 + \alpha^3$	$\alpha + \alpha^5$
$\alpha^3 + \alpha^4$	$1 + \alpha^2 + \alpha^3$
$1 + \alpha^2 + \alpha^4$	$1 + \alpha^4$
$1 + \alpha + \alpha^2 + \alpha^3$	$\alpha + \alpha^2 + \alpha^3$
$\alpha + \alpha^2 + \alpha^3 + \alpha^4$	$1 + \alpha^2 + \alpha^3 + \alpha^4$
$1 + \alpha^3 + \alpha^4$	$1 + \alpha + \alpha^3 + \alpha^4$
$1 + \alpha + \alpha^2 + \alpha^4$	$1 + \alpha$
$1 + \alpha + \alpha^3$	$\alpha^2 + \alpha^3$
$\alpha + \alpha^2 + \alpha^4$	$1 + \alpha^2 + \alpha^4$
$1 + \alpha^3$	$\alpha + \alpha^2 + \alpha^3 + \alpha^4$
$\alpha + \alpha^4$	$1 + \alpha + \alpha^2 + \alpha^4$
1	$\alpha + \alpha^2 + \alpha^4$
α	$\alpha + \alpha^4$
α^2	α
α^3	α^3
α^4	$1 + \alpha^2$
$1 + \alpha^2$	$\alpha^2 + \alpha^4$
$\alpha + \alpha^3$	$\alpha + \alpha^3 + \alpha^4$
$\alpha^2 + \alpha^4$	$1 + \alpha + \alpha^2$
$1 + \alpha^2 + \alpha^3$	$\alpha^2 + \alpha^3 + \alpha^4$
$\alpha + \alpha^3 + \alpha^4$	$1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4$
$1 + \alpha^4$	$1 + \alpha + \alpha^4$
$1 + \alpha + \alpha^2$	$\alpha + \alpha^2$
$\alpha + \alpha^2 + \alpha^3$	$\alpha^3 + \alpha^4$
$\alpha^2 + \alpha^3 + \alpha^4$	$1 + \alpha + \alpha^2 + \alpha^3$
$1 + \alpha^2 + \alpha^3 + \alpha^4$	$1 + \alpha^3 + \alpha^4$
$1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4$	$1 + \alpha + \alpha^3$
$1 + \alpha + \alpha^3 + \alpha^4$	$1 + \alpha^3$
$1 + \alpha$	1

CODING FOR SYNCHRONIZATION

The significance of synchronizable word selection and detection in a digital burst synchronization for communications satellites has been previously studied [27]. One of the problems in word selection has been the lack of a simple general procedure for determining a group of unique

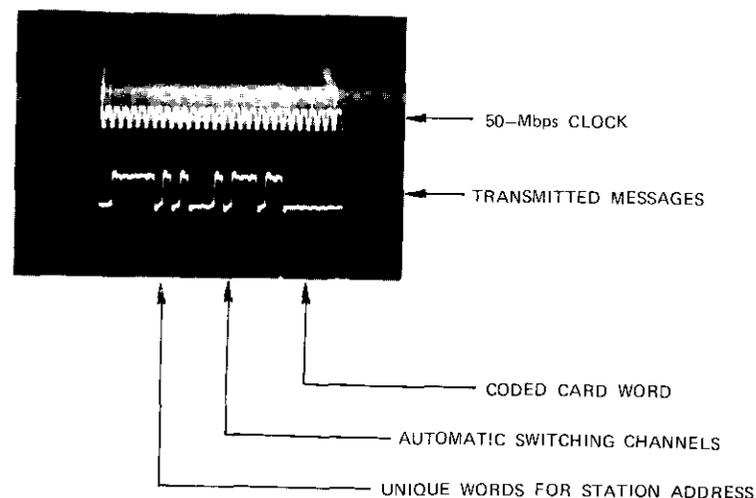


Figure 14. Experimental Waveform

words (UW's). A method has been suggested by Wu [28] to provide a solution by describing an algorithm which systematically leads to a set of code words provided that an extended block code exists.

Unique words are defined as code words patterned to satisfy both autocorrelation and cross-correlation criteria under specified error probabilities, since these properties affect the reliability of earth station identification and its corresponding message timing. Hence, multipoint communication requires not only maximal autocorrelation values, but also minimum cross-correlation values. In addition, whenever each word characterizes a ground station, a large number of such words is required. Such a requirement differs from that for deep-space single-spacecraft communication.

The error probabilities mentioned previously are known as false-detection (F) and miss-detection (Q) probabilities. For a particular n-bit unique word, UW_o , the decision rule of F and Q is expressed in Table 5, where \overline{UW}_o denotes any sequence which is not UW_o .

TABLE 5. DECISION ON UW_o

Decision \ Transmission	UW_o	\overline{UW}_o
	Accept	Correct
Reject	Q	Correct

From the table,

$$F = P(\text{accept } UW_o/\overline{UW}_o \text{ was sent}) \left| \begin{array}{l} \\ \text{random bits in} \\ \text{the detector} \end{array} \right. = \sum_{i=0}^t \binom{n}{i} \frac{1}{2^n} \quad (19)$$

$$\begin{aligned} Q &= P(\text{reject } UW_o/\overline{UW}_o \text{ was sent}) \\ &= P(\text{detection beyond tolerable errors}) \\ &= \sum_{i=t+1}^n \binom{n}{i} p^i (1-p)^{n-i} \end{aligned} \quad (20)$$

where p is the binary symmetric channel transition probability and t is the number of tolerable errors.

For satellite channels, $Q \ll F$; hence, F dominates. If possible, it is preferable to keep n small and t large, but this makes it difficult to obtain a small value of F . If n , p , Q , and F are given, a set of UW may be obtained as follows (see Figure 15):

a. The cross-correlation values are first checked to determine whether they meet the minimum distance (d_b) requirement and satisfy F , Q , and $F \approx Q$.

b. The code selection need not be limited to single block codes. Modified cyclic codes [29], [30] and product codes [31] can all be used without consideration of decoding complexity because only the word structure is of interest if no attempt is made to decrease F .

c. The average autocorrelation value is checked by Moon and Moser's [32] theorem, which states that, given $\epsilon > 0$, then

$$[\text{the average autocorrelation value}] \leq (2 + \epsilon) \sqrt{n \log n} \quad (21)$$

for almost all 2^n sequences, except those sequences for which $[\cdot]$ approaches 0 as n approaches ∞ .

Two possibilities then exist. First, all code words obtained in step b may be used as UW 's. This situation arises when a very large number of words is required (e.g., for certain applications encountered in navigation satellites), and, hence, some code words can exhibit poor autocorrelation values. If these values are not acceptable, then another approach may be taken. If the number of UW 's needed is less than 2^l , where $2^l < 2^k$, then the autocorrelation search is performed on the set of code words that satisfied step b. Since the autocorrelation search is simpler and the search takes place within a reduced code word subspace of 2^k , the amount of computation is reduced.

Known good autocorrelation words can be used to advantage as the k

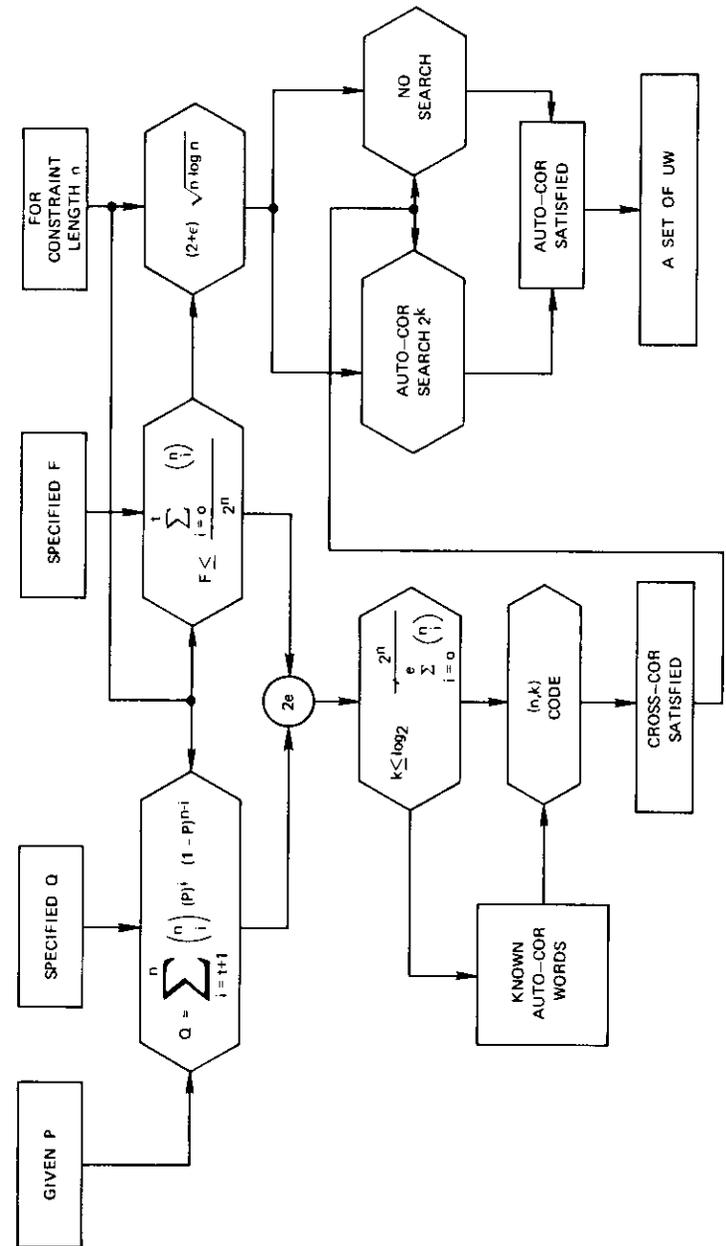


Figure 15. Unique Word Determination

information digits before coding. The expected autocorrelation value is increased from $(2 + \epsilon) \sqrt{k \log k}$ to $(2 + \epsilon) \sqrt{n \log n}$ due to coding. As an example, the following three equal length words are coded with the (23,12,7) Golay code with $g(x) = 1 + x^2 + x^3 + x^5 + x^6 + x^{10} + x^{11}$:

$$1 \ 1 \ 0 \ 1 \ 1 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \quad (22)$$

$$1 \ 0 \ 1 \ 0 \ 1 \ 1 \ 1 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \quad (23)$$

$$1 \ 0 \ 1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \quad (24)$$

The results are shown in Figures 16 and 17. For the averaged autocorrelation value of equation (22), the difference is between $(2) \sqrt{12 \log 12} = 7.2$ before coding and $(2) \sqrt{23 \log 23} = 11.2$ after coding. The difference in cross-correlation values for equations (23) and (24) at $\tau = 0$ is 8.

A GENERALIZATION OF SHIFT REGISTER SEQUENCE GENERATORS

Maximum length sequence generators have been known to be useful for correlation detection, noise generation, and coding in communication. Although feedback shift register sequence generators (FSRG's) have been extensively studied, a new generalization which is obtained by interleaving generalized feedback and/or feedforward SRG's is presented here. The new generalized approach is based on the following theorems.

Theorem 1

The maximal period of a sequence generated by Reed and Turn for nonlinear (κ, m) FSR's is $\kappa 2^m$. The maximal period of a sequence generated by Reed and Turn linear (κ, m) FSR's is $\kappa(2^m - 1)$ [33].

As shown in Figure 18, the Reed-Turn FSRG differs from the conventional FSRG's because it introduces an instruction sequence, $\pi_i(D)$, which controls the feedback function, f , and produces a set of transformations, $\alpha_i(D)$. Then each shift part of the m -stage shift register is instructed to change states in order to sustain a maximal period.

Proof

Assume that the initial state of the register is $X^{(0)}$ and the initial symbol of the control sequence is $\pi_i(1)$. Then $\alpha_i(1)$ is selected and the shift operation, c , is applied, i.e., $X^{(1)} = X^{(0)}[\alpha_i(1) c]$. Next, the symbol $\pi_i(2)$ is applied and the process is repeated. If the input has a length κ after s shifts ($s > \kappa$), the state of register is

$$X^{(s)} = X^{(0)} [\alpha_i(1) c] [\alpha_i(2) c] \dots [\alpha_i(s_\kappa) c] \quad (25)$$

where $s_\kappa = s \bmod (\kappa)$.

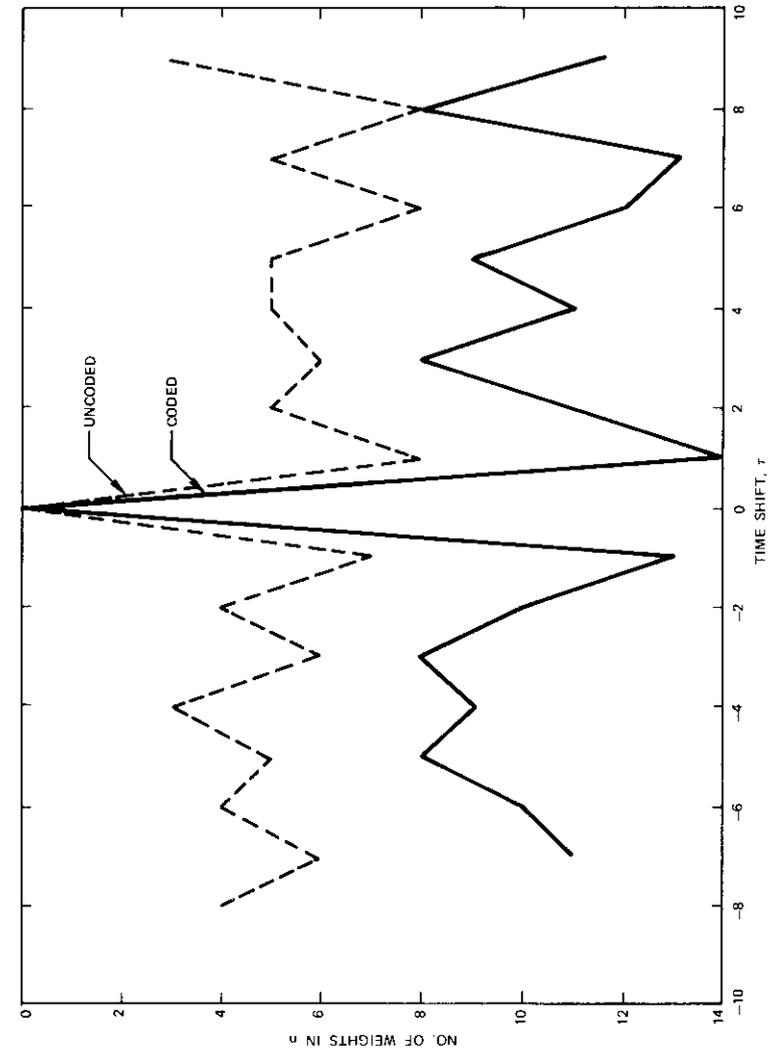


Figure 16. The Autocorrelation of the Unique Word (22)

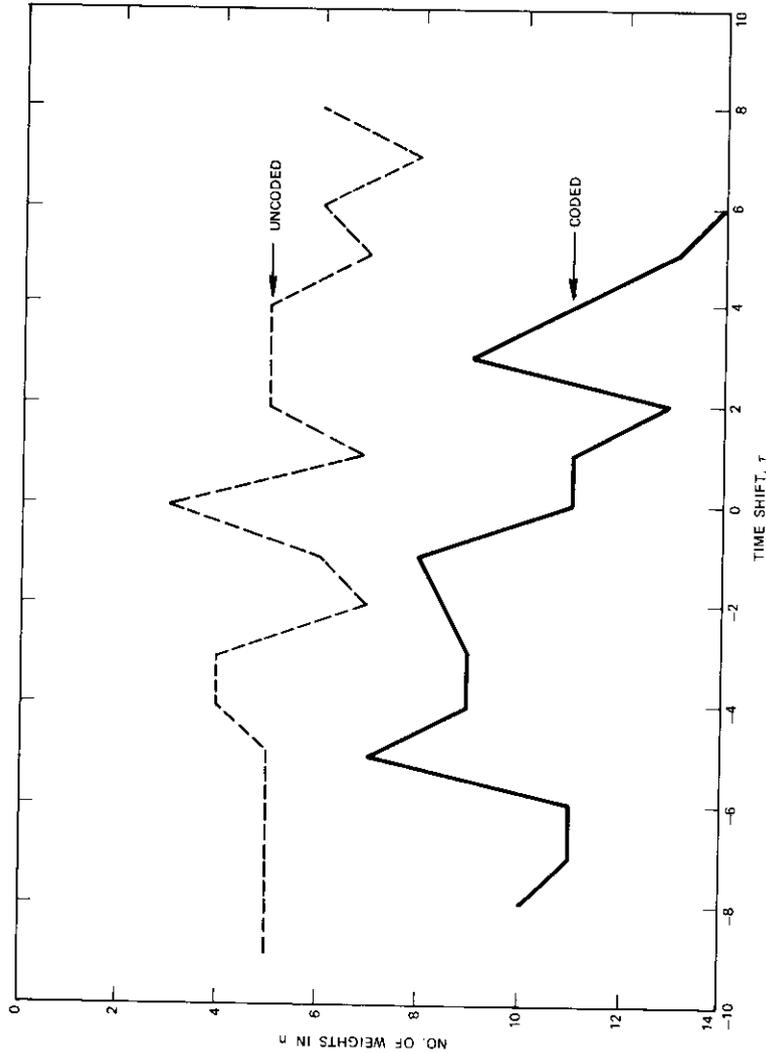


Figure 17. The Cross-Correlation of Unique Words (23),(24)

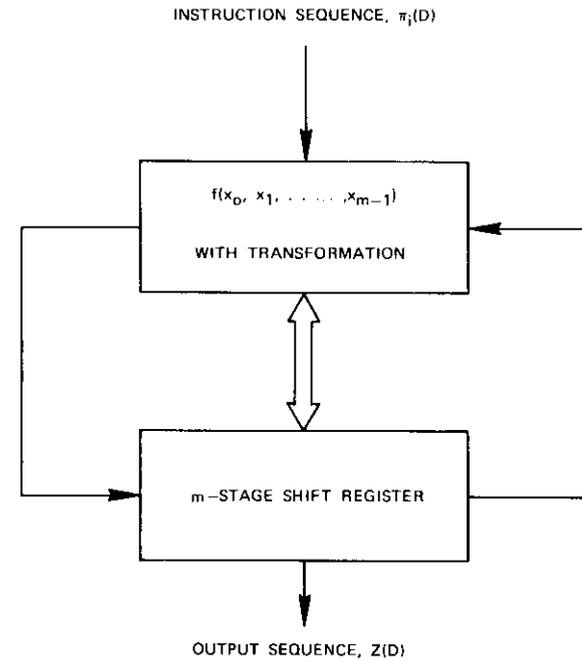


Figure 18. A Reed-Turn Shift Register Sequence Generator

For classical FSRG's, the maximum sequence length of an m -stage register is $2^m - 1$ if α is linear, and 2^m if α is nonlinear. Consider a Reed-Turn generator which has a sequence of transformations $(\alpha_1 c) \dots (\alpha_\kappa c)$, with a period κ , and a classical FSRG which has a sequence of transformations αc , such that

$$(\alpha c) = (\gamma_1 c) (\gamma_2 c) \dots (\gamma_\kappa c). \quad (26)$$

After $2^m - 1$ applications of (αc) , the classical FSRG is again in its initial state, $X^{(0)}$. On the other hand, for the Reed-Turn generator, the transformation $(\alpha_1 c)$ must be applied next. Each application of the right-hand product of equation (26) produces κ output bits to form the sequence. Thus, the length of the sequence is $\kappa(2^m - 1)$. A similar proof holds for nonlinear (κ, m) FSRG's.

The determination of the required feedback logic and the number of distinct transformations is generally very difficult for the Reed-Turn generators. The following approach provides a much simpler solution to the practical design of shift register sequence generators.

Theorem 2

A generalized shift register generator, G, can be constructed from N general feedback and/or feedforward subgenerators, which each have an arbitrary length, m_0, m_1, \dots, m_{N-1} . The output sequence generated by G in the sequence domain is

$$Z(D) = \sum_{d=0}^{n-1} A_d D^d \tag{27}$$

The period, n, of the generator, G, is bounded by

$$n \leq \prod_{\alpha=0}^{N-1} t_\alpha 2^{m_\alpha} \tag{28}$$

where t_α = the number of feedback transformations corresponding to the subgenerator α

A_d = the mod 2 sum of the product of the distinctive subgenerator state at time d from the set S

and

$$S = \sum_{W=0}^{n_{N-1}} \dots \sum_{j=0}^{n_1} \sum_{i=0}^{n_0} S_i(X_0, X_1, \dots, X_{m_{N-1}}) \\ S_j(X_0, X_1, \dots, X_{m_1}) \\ \vdots \\ S_W(X_0, X_1, \dots, X_{m_{N-1}}) \tag{29}$$

Proof

Each of the N subgenerators performs most effectively in accordance with theorem 1. For an α subgenerator, G_α , which has m_α shift register stages and t_α feedback transformations, the sequence length is $n_\alpha = t_\alpha 2^{m_\alpha}$. For each shifting bit, there is a corresponding register state, \bar{s}_α . Assume that there is another subgenerator, G_β , which has m_β shift register stages and which generates a sequence of length $n_\beta = t_\beta 2^{m_\beta}$. The shift register state vectors are represented by \bar{s}_β . Each state assignment, \bar{s}_α , of G_α is multiplied by every state, \bar{s}_β , of G_β . The sequence length of the generator, G, which is formed from a combination of G_α and G_β , is $n_\alpha n_\beta = t_\alpha t_\beta 2^{m_\alpha + m_\beta}$. This is the number of coefficients of A_d , which can be either 0 or 1 in the binary field. Each A_d is obtained from the mod 2 sums after each multiplication of \bar{s}_α by \bar{s}_β . Next let the combined generator, G, combine with another subgenerator, G_γ , with feedback transformation t_γ , to form a generator G' . The sequence length of G' is bounded by $t_\alpha t_\beta t_\gamma 2^{m_\alpha + m_\beta + m_\gamma}$. By extending the number of such generalized subgenerators to N as shown in Figure 19, we obtain theorem 2.

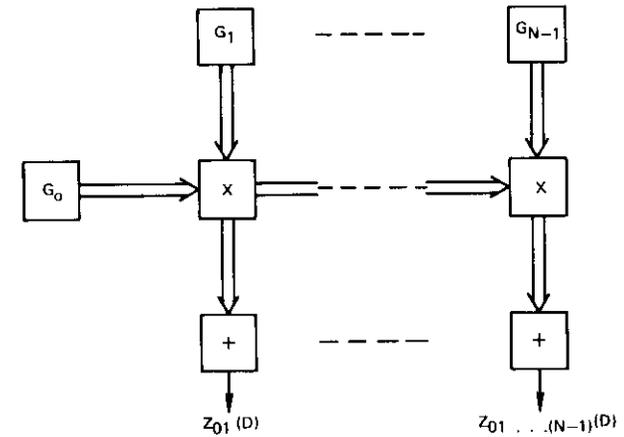


Figure 19. A General Shift Register Sequence Generator

A direct consequence of theorem 2 is that, if G is realized from classical nonlinear feedback shift register subgenerators which generate sequences with periods of $2^{m_0}, 2^{m_1}, \dots, 2^{m_{N-1}}$, then the combined generator generates a sequence with a length not more than

$$2 \exp \sum_{\alpha=0}^{N-1} m_\alpha$$

For N linear subgenerators, the number of different subgenerator combinations is

$$\prod_{\alpha=0}^{N-1} \frac{\phi(2^{m_\alpha} - 1)}{m_\alpha}$$

where $\phi(2^{m_\alpha} - 1)$ is the Euler's function.

This product number is larger than the number of combinations which may be obtained from any individual subgenerator, and is generally less than the number of combinations which may be obtained from a single shift register with the equivalent combined stages. It is advantageous for all m_α 's in theorem 2 to be equal because longer sequence lengths can be obtained from identical modular subgenerators. Well-known results from linear, nonlinear, feedback, or feedforward shift register subgenerators can be effectively combined in a systematic way. Arbitrary n can be obtained by properly selecting t, m, and N.

CONCLUDING REMARKS

For low-rate applications, as in the case of deep-space missions, error

coding has proved its usefulness. In this paper, several highly efficient coding techniques are presented for applications related to communications satellite links. The expected performances are described, and computer simulation results are presented for two cases. For message control, a block codec is illustrated; the design of the codec and its experimental evaluations are demonstrated. For synchronization, an algorithm is suggested which allows one to obtain systematically a set of unique words. For noise generation, coding, or signal spectrum spreading communications, a new generalization of shift register sequence generators is derived from the theory of convolutional coding. This method provides a way of generating long sequences from simpler modular subgenerators.

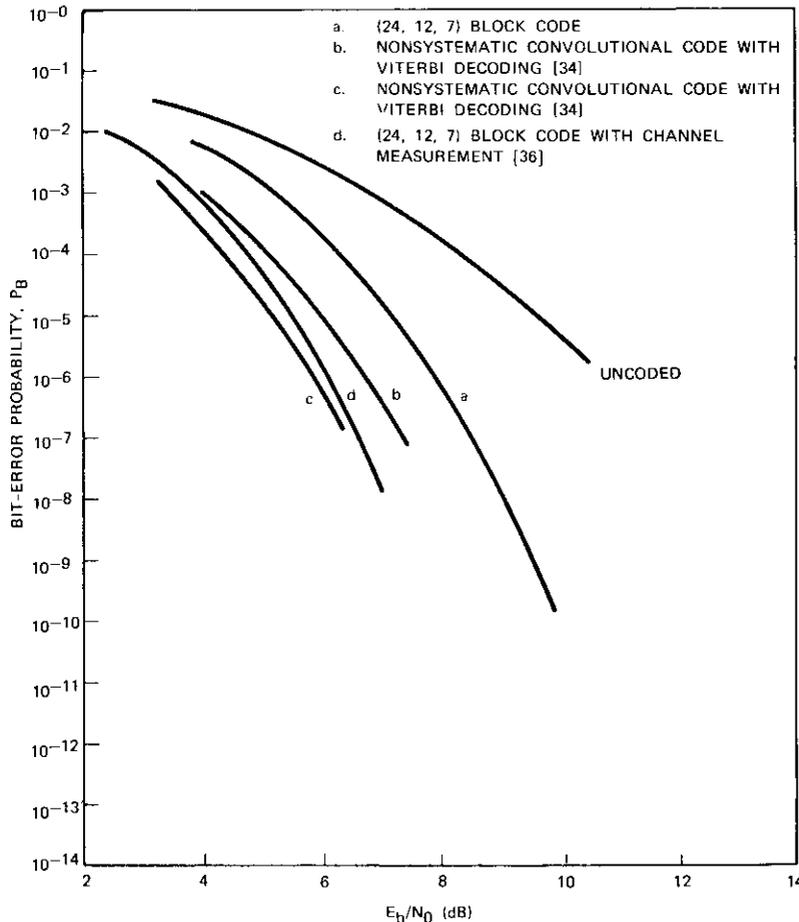


Figure 20. Code Comparisons at Rate-1/2

As shown in Figure 2, nonsystematic convolutional codes with maximum likelihood decoding yield higher $E_x(R)$ at high rates. The decoding algorithm is an iterative computational procedure in which the number of calculations grows exponentially with the code constraint length, K . The decoder needs storage for each of the 2^{K-1} encoder states. The total memory requirement is about $5K2^{K-1}$ [34]. For larger values of K and high speed, it is difficult to avoid increased complexity.

Unfortunately, sequential decoding algorithms with convolutional codes have practical limitations, although their performance is attractive [9],[35]. The computational complexity and buffer overflow problems are shortcomings, especially when high-speed, high-efficiency, and cost-conscious commercial communications are involved. The probability of buffer overflow is exponentially proportional to the coding rate.

The soft decision block decoders [36] use a set of channel measurements which provide improved performance for block codes. At $R = 1/2$, a comparison is made in Figure 20. A comparison of equipment complexity between a longer block code with conventional decoding and a shorter block code which provides the same performance with channel measurement decoding needs to be implemented.

ABBREVIATIONS AND SYMBOLS

BSC	Binary symmetric channel
CARD	Telephone channel allocation, routing, and destination
c	Shifting operation
codec	Encoder-decoder pair
D	Delay operator
D	Difference set
d	Code minimum distances (d_b for block codes, d_c for convolutional codes)
d_i	Distinct residues in a difference set
E_b	Signal energy
$E_x(R)$	Reliability function
F	False-detection error probability
$f_i(\cdot)$	Decoding bit-error probability after the i th code
GF($\#$)	Finite field with $\#$ elements
$g(D)$	Generator polynomial for convolutional code
$g(x)$	Generator polynomial for block code
ISECB	Interleaved single-error-correcting block
ISECC	Interleaved single-error-correcting convolutional
K	Number of inputs of the convolutional encoder

K	Nonsystematic convolutional code constraint length
k	Information symbols of block code
l	Positive integer less than k
M	Difference set parameter
m	Number of shift register generator stages
m_b	Number of parity check digits of block code
m_c	Number of parity check digits of convolutional code per block
modem	Modulator-demodulator pair
N	Number of generalized subgenerators
N_o	Number of outputs per block of convolutional code
N_o	Noise density
N_r	Code word length
n	Synchronizable word length
n_b	Word length of block code
n_c	Constraint length of convolutional code with feedback decoding
n	Sequence length of shift register generator
P_b	Bit-error probability
PCM	Pulse-code modulation
PSK	Phase-shift keyed
p	BSC transition probability
Q	Miss-detection error probability
Q	One minus the probability $\{ \cdot \}$, where $\{ \cdot \}$ is the quantity in braces in equation (4)
q	Difference set parameter
R	Code rate
S_i	Power sums
$T(\cdot)$	Transmitted sequence
TDMA	Time-division multiple access
t	Number of feedback transformations
t	Number of tolerable errors
α	Primitive element
β	Error-location numbers
Δ	Triangular matrix
$\underline{\Delta}$	Delta power sum matrix
κ	Number of feedback logics
λ	Difference set parameter
$\phi(\cdot)$	Euler's function

REFERENCES

- [1] R. W. Hamming, "Error Detecting and Error Correcting Codes," *Bell System Technical Journal*, Vol. 29, No. 2, April 1950, pp. 147-160.
- [2] D. Selpian, "A Class of Binary Signaling Alphabets," *Bell System Technical Journal*, Vol. 35, No. 1, Jan. 1956, pp. 203-234.
- [3] W. W. Peterson, *Error-correcting Codes*, Cambridge, Mass.: MIT Press, 1961.
- [4] J. L. Massey, *Threshold Decoding*, Cambridge, Mass.: MIT Press, 1963.
- [5] W. C. Gore, "Generalized Threshold Decoding of Linear Codes," *IEEE Transactions on Information Theory*, IT-15, No. 5, Sept. 1969, pp. 590-592.
- [6] E. R. Berlekamp, *Algebraic Coding Theory*, New York: McGraw-Hill, 1968.
- [7] P. Elias, "Error-free Coding," *IRE Transactions on Information Theory*, IT-1, No. 4, Sept. 1954, pp. 29-37.
- [8] J. M. Wozencraft and B. Reiffen, *Sequential Decoding*, Cambridge, Mass.: Technology Press of MIT and New York: Wiley, 1961.
- [9] F. Jelinek, "A Fast Sequential Decoding Algorithm Using a Stack," *IBM Journal of Research and Development*, Vol. 13, No. 6, Nov. 1969, pp. 675-685.
- [10] A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Transactions on Information Theory*, IT-13, No. 2, April 1967, pp. 260-269.
- [11] R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.
- [12] E. A. Bucher and J. A. Heller, "Error Probability Bounds for Systematic Convolutional Codes," *IEEE Transactions on Information Theory*, IT-16, No. 2, March 1970, pp. 219-224.
- [13] S. Lin and E. J. Weldon, Jr., "Long BCH Codes are Bad," *Information and Control*, Vol. 11, No. 4, Oct. 1967, pp. 445-451.
- [14] J. L. Massey, "Shift-register Synthesis and BCH Decoding," *IEEE Transactions on Information Theory*, IT-15, No. 1, Jan. 1969, pp. 122-127.
- [15] R. T. Chien, "Cyclic Decoding Procedures for Bose-Chaudhuri-Hocquenghem Codes," *IEEE Transactions on Information Theory*, IT-10, No. 4, Oct. 1964, pp. 357-363.
- [16] H. O. Burton, "Inversionless Decoding of Binary BCH Codes," *IEEE Transactions on Information Theory*, IT-17, No. 4, July 1971, pp. 464-466.
- [17] N. Abramson, "Cascade Decoding of Cyclic Product Codes," *IEEE Transactions on Communication Technology*, COM-16, No. 3, June 1968, pp. 398-402.

- [18] D. T. Tang and R. T. Chien, "Cyclic Product Codes and Their Implementation," *Information and Control*, Vol. 9, No. 2, April 1966, pp. 196-209.
- [19] J. L. Ramsey, "Cascaded Tree Codes," RLE Technical Report No. 478, MIT Research Laboratory of Electronics, Sept. 1970, pp. 88-99 (N71-11270).
- [20] J. P. Robinson and A. J. Bernstein, "A Class of Binary Recurrent Codes with Limited Error Propagation," *IEEE Transactions on Information Theory*, IT-13, No. 1, Jan. 1967, pp. 106-113.
- [21] M. Hall, Jr., "A Survey of Difference Sets," *Proceedings of the American Mathematical Society*, Vol. 7, No. 6, Dec. 1956, pp. 975-986.
- [22] E. J. Klieber, "Some Difference Triangles for Constructing Self-orthogonal Codes," *IEEE Transactions on Information Theory*, IT-16, No. 2, March 1970, pp. 237-238.
- [23] A. D. Wyner and R. B. Ash, "Analysis of Recurrent Codes," *IEEE Transactions on Information Theory*, IT-9, No. 3, July 1963, pp. 143-156.
- [24] C. V. Freiman and J. P. Robinson, "A Comparison of Block and Recurrent Codes for the Correction of Independent Errors," *IEEE Transactions on Information Theory*, IT-11, No. 3, July 1965, pp. 445-449.
- [25] G. D. Forney, *Concatenated Codes*, Cambridge, Mass: MIT Press, 1966.
- [26] J. G. Puente, W. G. Schmidt and A. M. Werth, "Multiple-access Techniques for Commercial Satellites," *Proceedings of the IEEE*, Vol. 59, No. 2, Feb. 1971, pp. 218-229.
- [27] W. Schrepp and T. Sekimoto, "Unique Word Detection in Digital Burst Communications," *IEEE Transactions on Communication Technology*, COM-16, No. 4, Aug. 1968, pp. 597-605.
- [28] W. W. Wu, "Coding Unique Words in Communication Satellites," *National Telemetry Conference*, Washington, D.C., April 1971, pp. 270-276.
- [29] S. Y. Tong, "Synchronization Recovery Techniques for Binary Cyclic Codes," *Bell System Technical Journal*, Vol. 45, No. 4, April 1966, pp. 561-596.
- [30] R. C. Bose and J. G. Caldwell, "Synchronization Error-correcting Codes," *Information and Control*, Vol. 10, No. 6, June 1967, pp. 616-630.
- [31] W. C. Gore, "Further Results on Product Codes," *IEEE Transactions on Information Theory*, IT-16, No. 4, July 1970, pp. 446-451.
- [32] J. W. Moon and L. Moser, "On the Correlation Function of Random Binary Sequences," *SIAM Journal on Applied Mathematics*, Vol. 16, No. 2, March 1968, pp. 340-343.

- [33] I. S. Reed and R. Turn, "A Generalization of Shift-register Sequence Generators," *Journal of the Association for Computing Machinery*, Vol. 16, No. 3, July 1969, pp. 461-473.
- [34] *Convolutional Codes: The Key to Effective Error Control and Efficient Communication*, seminar presented by the Linkabit Corporation, Washington, D.C., June 1970.
- [35] R. M. Fano, "A Heuristic Discussion of Probabilistic Decoding," *IEEE Transactions on Information Theory*, IT-9, No. 2, April 1963, pp. 64-74.
- [36] D. Chase, "Decoding with Channel Measurement Information," unpublished presentation, 1970 *International Symposium on Information Theory*, Noordwijk, The Netherlands, June 1970.

ACKNOWLEDGMENTS

I wish to thank L. S. Golding and P. L. Bargellini for their careful review and helpful comments. I also want to thank W. G. Schmidt for his suggestions on coding in the TDMA system and M. C. Hu for his help with the computer simulations.



William W. Wu holds a B.S.E.E. from Purdue University and a M.S.E.E. from Massachusetts Institute of Technology. He joined COMSAT Laboratories in 1967 as a member of the technical staff. He has been responsible for analysis and synthesis of error-correcting codes in communication satellite links. Before joining COMSAT he was a DSR staff member at the M.I.T. Center for Space Research, Cambridge, Mass.

CTR NOTES: Attenuation Statistics at 15.3 GHz for Clarksburg, Maryland

H. D. CRAFT, JR.*

Since early July 1970, nearly continuous measurements have been made at COMSAT Laboratories, Clarksburg, Maryland, of the attenuation on an earth-space path at 15.3 GHz. The measurements were made utilizing either the 15.3 GHz transmissions from the ATS-5 satellite or radiometric techniques on the sun or sky. The receiving antenna was a 16-foot paraboloid directed at the ATS-5 satellite (elevation angle approximately 35 degrees), except for a few short periods during which the sun was tracked.

For the data recorded from July 3, 1970, to July 3, 1971, Figure 1 shows the percentage of the time during which attenuation levels in excess of 10 dB were noted (10 dB is the maximum loss that can be measured with reasonable accuracy by radiometric techniques). The dashed curve of Figure 1 shows similar statistics, compiled by the Bell Telephone Laboratories sun tracker and radiometer systems, in New Jersey between December 8, 1967, and February 28, 1969.† The third curve in Figure 1 shows the statistics for the worst month at Clarksburg—July. (The data of July 1-2, 1971, have been combined with the data for July 3-31, 1970.)

Some caution must be exercised in using the annual curve shown in Figure 1 for the COMSAT data since the position of the high attenuation tail of that curve depends heavily upon a small number of events (thunderstorms). One or two extra storms per summer will have a substantial impact on that part of the curve. For instance, if the COMSAT data from July 1, 1970, to July 1, 1971, are used, then the percentages of time 10 dB loss is exceeded drop to 0.031% rather than the 0.047% shown. This is the result of a storm on July 2, 1971.

Figure 2 shows the statistics for the four seasons of the year. The

* Formerly with COMSAT Laboratories, Dr. Craft is now Technical Coordinator at the Arecibo Observatory (P.R.) of Cornell University. This note is based upon work performed under the sponsorship of the International Telecommunications Satellite Consortium (INTELSAT). Any views expressed in this note are not necessarily those of INTELSAT.

† H. W. Evans, "Attenuation on Earth-Space Paths at Frequencies up to 30 GHz," *IEEE International Conference on Communications*, Montreal, Canada, June 1971, p. 27-1.

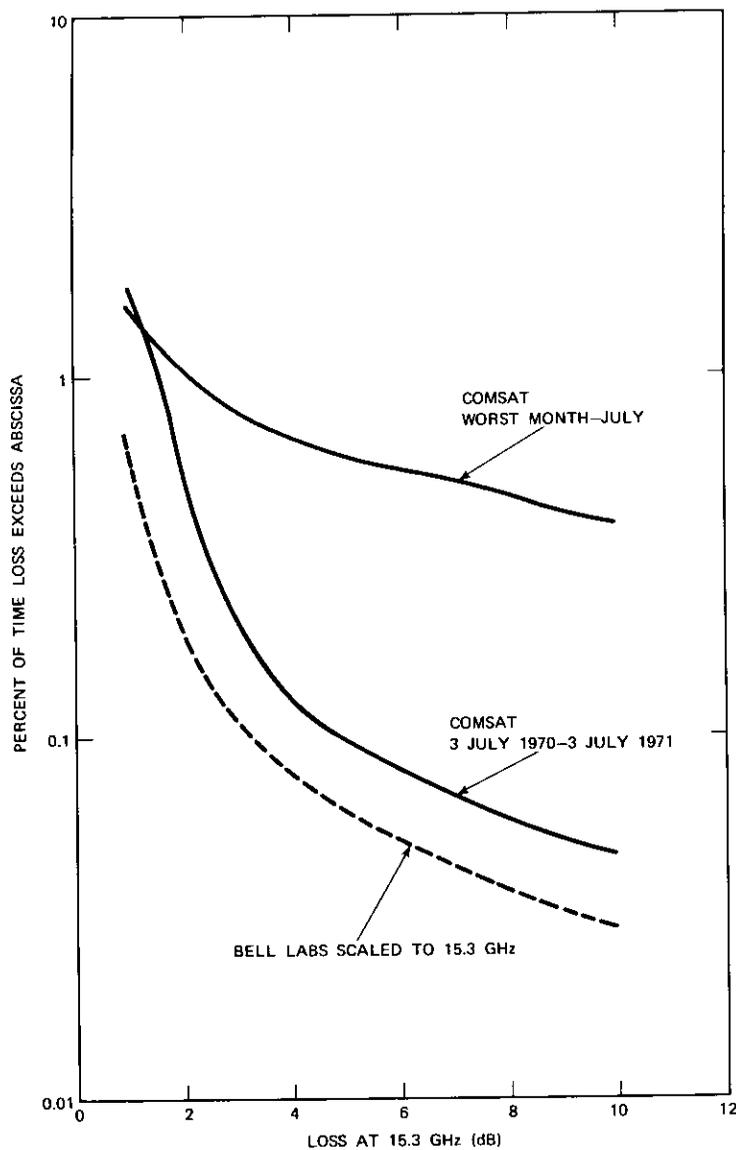


Figure 1. Percentage of the Time That the Earth-Space Path Loss at 15.3 GHz Exceeds the Values on the Abscissa for COMSAT Data Between July 3, 1970, and July 3, 1971

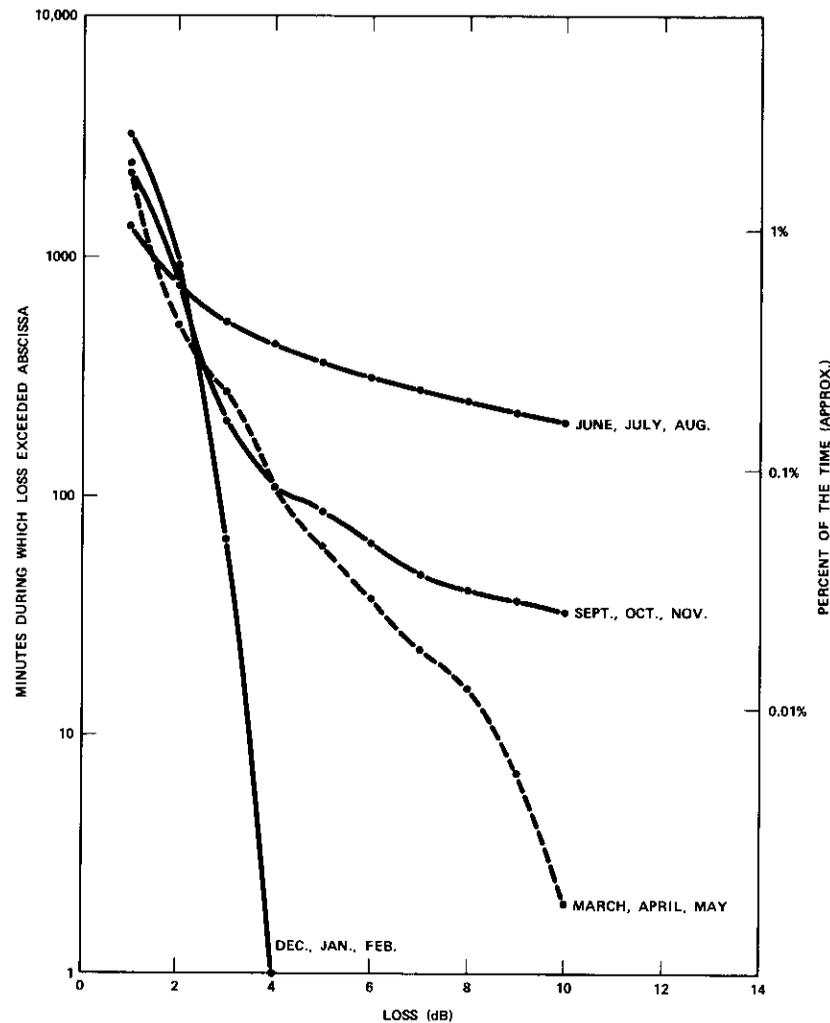


Figure 2. The Time That the Earth-Space Path Loss at 15.3 GHz Exceeds the Values on the Abscissa for the Four Seasons of the Year

position of the curves in this figure agrees quantitatively with the thunderstorm activity at Clarksburg.

Figure 3 indicates the distribution throughout the day of fades equal to or greater than 4 and 9 dB. There is a predominance of fades in the afternoon and early evening hours, the hours when thunderstorm activity is most probable.

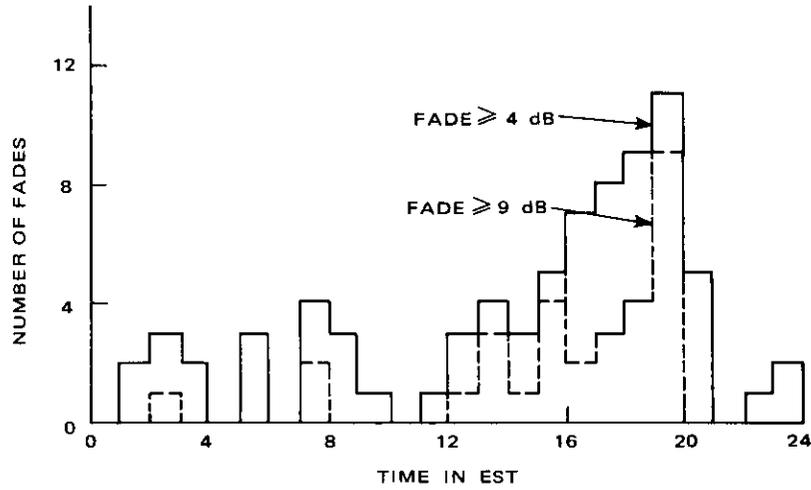


Figure 3. Histogram of the Time of Occurrence (Eastern Standard Time) of Fades Greater Than or Equal to 4 dB and 9 dB (COMSAT data, July 3, 1970—July 3, 1971)

Figure 4 indicates the durations of fades equal to or greater than 4 and 9 dB. Most fades are ten minutes or less in duration, but long duration fades are also possible. A fade greater than 9 dB and lasting for 74 minutes has a heavy impact on the COMSAT data shown in Figure 1. It is clear that, for any reliable yearly statistic to be derived, measurements must be made for more than just one year.

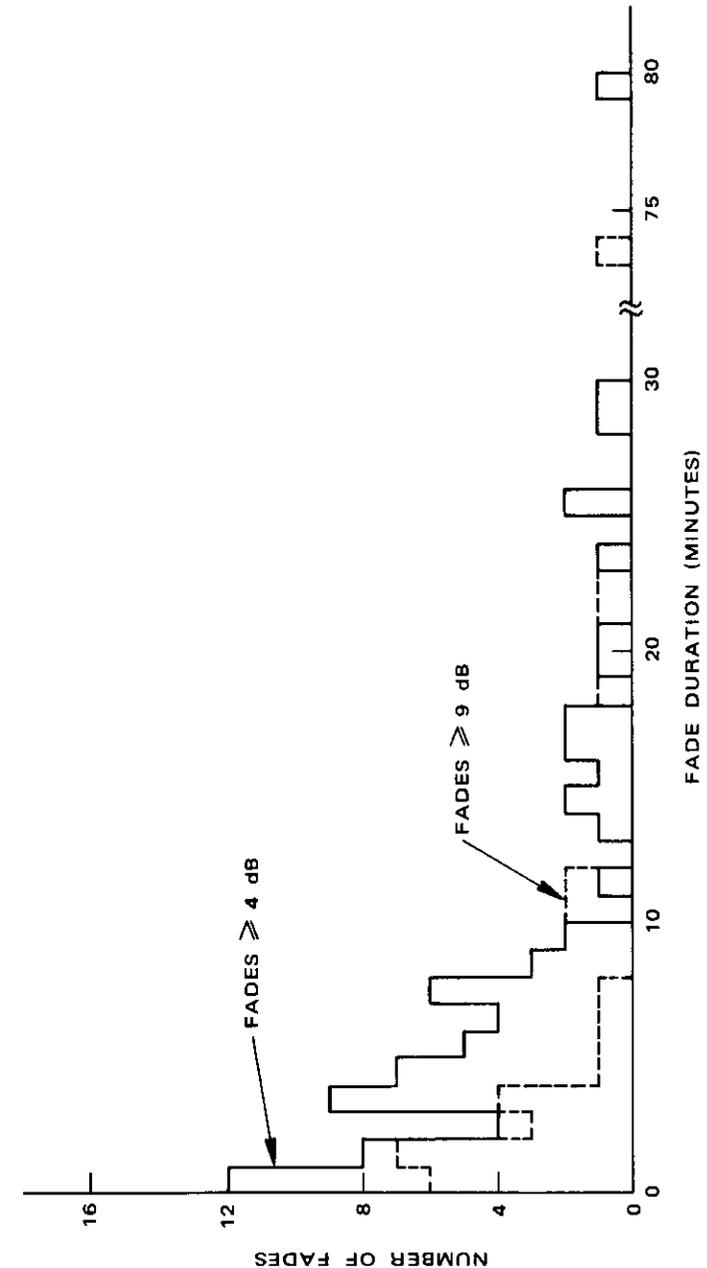


Figure 4. Fade Duration Histogram for Fades Greater Than or Equal to 4 dB and 9 dB (COMSAT data, July 3, 1970—July 3, 1971)

**RECHERCHES EXPERIMENTALES SUR LE BALLOTTEMENT
DU COMBUSTIBLE DANS UN SATELLITE
A DOUBLE ROTATION**

ERNESTO R. MARTIN

Sommaire

L'énergie dissipée par des éléments vibratoires du rotor en rotation d'un satellite à double rotation, comme le satellite de télécommunications INTELSAT IV, donne une certaine instabilité à la dynamique de l'engin. Des recherches expérimentales en vue de calculer la dissipation de l'énergie due au ballottement du combustible dans le réservoir ont révélé une turbulence inattendue des mouvements du liquide dans les réservoirs à combustible conicosphériques des INTELSAT IV. On a observé le mouvement du fluide à l'intérieur d'un réservoir en rotation soumis à des oscillations angulaires semblables à celles du satellite. Des essais précédents, au cours desquels le réservoir en rotation avait été soumis à une vibration rectiligne, mais à des accélérations linéaires analogues, avaient eu pour résultat un mouvement pendulaire, calme, pareil à celui d'un corps rigide. On a découvert que le mouvement de turbulence du fluide est provoqué lorsqu'un réservoir en rotation est soumis à des oscillations angulaires autour d'un axe qui n'est pas parallèle à l'axe de symétrie du réservoir.

**NOUVEAUX FILTRES PASSE-BANDE A GUIDE
D'ONDES, POUR LES TRANSPONDEURS DE SATELLITE**

A. E. ATIA ET A. E. WILLIAMS

Sommaire

Les transpondeurs de satellites de télécommunications de conception classique utilisent des systèmes de multiplexage qui reposent sur des filtres passe-bande Tschebyscheff ou Butterworth à guide d'ondes. Non seulement ces filtres représentent-ils une partie importante du poids total du transpondeur, mais leurs fonctions de filtrage n'offrent pas de caractéristiques optimales en ce qui concerne la sélectivité des fréquences, le temps de propagation de groupe, ou l'affaiblissement d'insertion dans la bande. Cette étude décrit comment l'on peut améliorer sensiblement les fonctions de filtrage en utilisant des modes orthogonaux dans des cavités de guides d'ondes soit carrées soit circulaires. Le poids des filtres en est réduit d'au moins 50%, et on peut réaliser des fonctions optimales de