COMSAT
Technical Review

COMSAT TECHNICAL REVIEW
Volume 6 Number 2, Fall 1976

A model for TDMA burst assignment and scheduling

A. K. SINHA

(Manuscript received March 25, 1976)

Abstract

This paper presents a simple mathematical formulation of the problem of time-division multiple-access (TDMA) burst assignment and scheduling for a general communications satellite system including an arbitrary earth station network, beam coverage pattern, and transponder configuration. Relevant concepts of beam overlap (over earth stations), burst overlap (in time), and earth station equipment requirements are introduced and precisely defined. In addition, useful parameters for evaluating the efficiency of system utilization are identified. Finally, a semianalytical algorithm is proposed for scheduling TDMA bursts so that earth segment equipment requirements are minimized and achievable scheduling efficiencies are optimized for a given traffic data base and system configuration. An example of a schedule obtained from a newly prepared computer program based on this approach is presented.

Introduction

In recent years the growing interest in the development and application of time-division multiple access (TDMA) for communications satellites in general [1]–[6] and the INTELSAT system in particular [7]–[9] stems from its

---
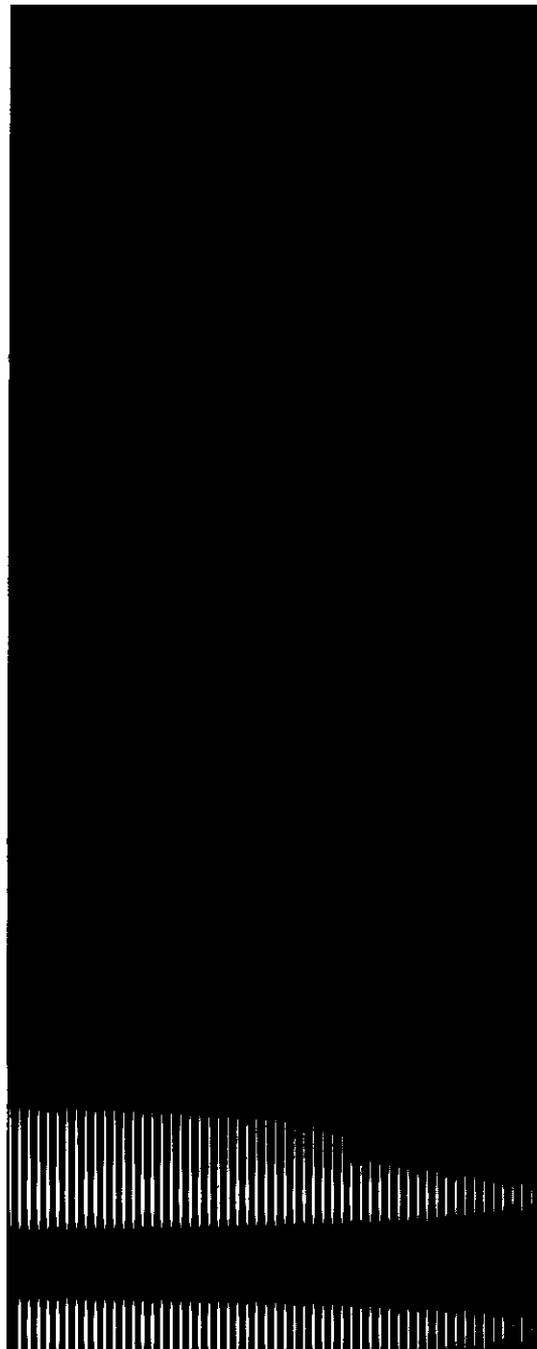
distinct advantages over frequency-division multiple access (FDMA). Because in TDMA the entire bandwidth of the transponder is allocated to a single link, the problems of intermodulation and the related requirement of power backoff caused by the nonlinear transponder response, which are major drawbacks of FDMA, are mitigated. Thus, the satellite power can be used more efficiently, thereby enhancing communications capacity. Furthermore, for a large and evolving communications network such as the INTELSAT system, the tedious task of determining frequency plans consistent with existing earth station equipment inventories at various stages of growth is replaced by the relatively simpler problem of reprogramming TDMA burst schedules. Inexpensive transponder hopping with or without frequency hopping, depending on the frequency reuse scheme, is possible for earth stations.

In short, the TDMA mode of satellite communications offers the possibility of considerably enhancing the overall system utilization efficiency relative to that obtained with FDMA. However, the exact degree to which increased efficiency can be realized in practice depends on the efficacy of TDMA burst scheduling. In the case of a multitransponder satellite operating in the TDMA mode, the burst scheduling must, of course, be coupled with a suitable assignment of bursts to transponders.

Adequate tools for analyzing the TDMA burst assignment and scheduling problem must be developed for a realistic evaluation of new or proposed TDMA satellite systems, or for comparing alternative system configurations on the basis of performance, efficiency, and economic considerations. The problem is generally complex due to the number of system parameters and the complexity of their interrelationships. In particular, apart from the two basic "orthogonal" degrees of freedom, frequency and time, a new degree of freedom is created by the use of satellite onboard directional antennas which provide isolation between earth stations based on orthogonal polarization or spatial separation. This "spatial" degree of freedom allows for multiple reuse of the frequency spectrum. Extensive use of this degree of freedom is envisioned in the future in the so-called satellite-switched TDMA (SS/TDMA) or space-division multiple-access (SDMA) systems [10]–[13]. Thus, the basic degrees of freedom involved in a general TDMA satellite system—frequency, time, coding, polarization, and spatial isolation of beams—can generally be exploited to advantage in burst assignment and scheduling.

Although various aspects of the TDMA concepts and techniques have been analyzed at length, rather limited attention has been paid to the scheduling problem [14]. The necessity for a quantitative analysis of various algorithms for TDMA burst assignment and scheduling in a multitransponder multibeam satellite with arbitrary beam coverage, connectivity, frequency reuse, and transponder hopping as well as a set of pertinent system constraints and options can hardly be overemphasized. Since the size and complexity of the problem generally prohibit a completely analytical solution, the development of a computer simulation model based on the basic mathematical model is necessary to provide the best possible tradeoff among maximization of satellite capacity utilization, minimization of earth station equipment cost, and flexibility for network expansion and traffic growth.

The present paper deals with the formulation of a simple mathematical model and analytical algorithms for the TDMA scheduling problem in a general satellite communications system. From the previous discussion, it can be seen that TDMA scheduling clearly involves two important interrelated procedures: assignment of traffic to transponders and time ordering of the assigned traffic within the transponder frames. This paper proposes a simple algorithm based on sequential assignment of transponders and time slots that provides a systematic combined treatment of these two procedures. The algorithm is further generalized to accommodate specific system constraints as well as various schemes for determining hierarchy and priority.

For simplicity, the present discussion is confined to the case of a single non-SS/TDMA satellite of specified transponder design, including channelization, connectivity, capacity, and beam coverage patterns. Generalizations to the case of multisatellite systems, including SS/TDMA transponders, are fairly straightforward.

First, the basic elements of the model are considered to establish the notation and to organize related observations. Symbolic representations are introduced to concisely describe the earth and space segments of the system under study in terms of the traffic matrix and traffic growth, beam coverage, transponder channelization and connectivity, and earth station equipment inventory. The characteristic parameters of TDMA bursts are identified, and certain basic or operational system constraints for burst assignment and scheduling, including burst format, burst overlap, burst length, earth station equipment requirements, and other related factors, are formulated. Quantitative measures of system utilization efficiency are introduced, followed by a simple semianalytical algorithm for burst assignment and scheduling. Finally, the computer implementation of the

model is described and an example is presented. It is shown that satisfactory results can be obtained by using the model.

Mathematical model description

The basic set of system components consists of the traffic data base, satellite beam coverages, satellite configuration, earth station equipment inventory, and transponder utilization efficiency. A general mathematical description of these quantities and the implications thereof is provided in the following subsections.

Traffic data base

For λ earth stations labeled $S_1, S_2, \ldots, S_\lambda$, the traffic from station S_i to station $S_j (i \neq j)$ is given by the element $\Gamma_{ij} (1 \leq i \leq \lambda, 1 \leq j \leq \lambda)$ of the $\lambda \times \lambda$ square traffic matrix Γ, which is generally a function of time, reflecting the growth of the traffic over a number of years. If it is assumed that bursts are to be assigned subject to constraints imposed by the availability of a prespecified set of transponders and earth station equipment, there is obviously an upper limit on the traffic growth that can be accommodated in the system. The traffic growth can be symbolically represented by

$$\Gamma = \Gamma(g, t) \tag{1}$$

where g and t denote the growth rate and time, respectively. Beyond a certain stage of growth corresponding to time t_s,

$$\Gamma(g, t_s) = \overline{\Gamma} \tag{2}$$

no further growth may be accommodated by the system, and the matrix $\overline{\Gamma}$ is defined as the saturation matrix.

As mentioned previously, the schedules may be determined at any stage of growth. Alternatively, if it is desirable to avoid rescheduling at every stage for a growing traffic base, it is best to start with a traffic matrix which already includes the projected growth. The necessary operations can then be performed to provide a schedule and to determine the equipment requirements at saturation. Conversely, if the equipment requirements are subject to prespecified constraints, growth can be simulated in stages to determine through scheduling the maximum traffic or saturation date consistent with the specified constraints. In any case, after the scheduling for the saturation traffic matrix is completed, the traffic schedule for any stage

prior to saturation can be obtained simply by reducing all the traffic elements proportionately. This leaves additional gaps in the time frame which are gradually filled up as the traffic grows so that the original configuration is attained at saturation.

Satellite beam coverages

Assume that there are μ satellite antenna beams $B_1, B_2, \ldots, B_\mu$ covering the stations according to a specified distribution

$$\{S_l\} \in \{B_m\}, \quad 1 \leq l \leq \lambda, \quad 1 \leq m \leq \mu \quad . \tag{3}$$

The specified beam coverage pattern can be represented by the $\lambda \times \mu$ matrix B with the following designation for the elements:

$$B_{lm} = \begin{cases} 1, & S_l \in B_m \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Each beam B_m is connected to one or more transponders for up- and down-link transmission and can accordingly be associated with a characteristic frequency pair Ω_m/Ω_m', where Ω_m and Ω_m' represent the characteristic frequencies of transponders connected to B_m for up- and down-link transmission, respectively. In case of frequency reuse, two or more of the Ω_m (and Ω_m') values are identical. For frequency reuse with dual polarization, the respective frequency values will be denoted as $\Omega_m^{(\pm)}$, where the superscripts $(+)$ and $(-)$ represent the horizontal (or right circular) and vertical (left circular) polarizations, respectively. In particular, two beams with identical geographical coverage and/or identical operating frequencies but orthogonal polarizations are physically independent and will be treated as distinct beams.

The geographical overlap pattern of the set of μ beams described above can be completely derived from the matrix B, whose columns can be regarded as beam vectors. The magnitudes of individual vectors are related to the total number of stations covered by the respective beams, and the scalar product P_{mn} between two vectors represents the number of common stations in the corresponding beams B_m and B_n. Two beams B_m and B_n with no common station $(P_{mn} = 0)$ will be referred to as spatially disjoint; conversely, two beams with at least one common station will be referred to as spatially overlapping. The member of the set $\{B_m\}$ covering all the stations obviously has the highest magnitude and spatial overlap with each

of the remaining members of the set; this beam is termed a global beam. Spatially overlapping beams must differ in terms of characteristic frequency pair, polarization, or transponder channelization for optimal utilization of the satellite capacity. In particular, in a global beam, access to all stations is achieved at the expense of bandwidth reuse. In the other extreme case, if each beam in the set $\{B_m\}$ covers only one station ($\mu = \lambda$), then this access limitation permits maximum bandwidth utilization through frequency reuse up to λ times.

Satellite configuration

Assume that there are ν transponders $T_1, T_2, \ldots, T_\nu$ with a specified distribution of up- and down-link connectivity over the μ beams $B_1, B_2, \ldots, B_\mu$:

$$\{T_k\} \in \{B_m\}, \quad 1 \leq k \leq \nu, \quad 1 \leq m \leq \mu \quad . \tag{5}$$

The transponder beam connectivity may change systematically over the frame length, τ, in the case of an SS/TDMA system; otherwise the connectivity remains unaltered.

Each transponder is characterized by a specified frequency bandwidth and reference capacity over the frame length. For convenience, this association can be described as follows. Assume that there are ν discrete frequency divisions $\bar{c}_1, \bar{c}_2, \ldots, \bar{c}_\nu$ having bandwidths of $\bar{b}_1, \bar{b}_2, \ldots, \bar{b}_\nu$ and capacities of $c_1, c_2, \ldots, c_\nu$, respectively. Capacity C_k of transponder $T_k(k = 1, 2, \ldots, \nu)$ can be expressed by specifying the particular frequency divisions with which the transponder is associated, or, equivalently, the corresponding starting and ending division numbers, k_1 and k_2, respectively.

If it is assumed for simplicity that the bandwidths of various divisions are integral multiples of a fundamental unit bandwidth b_o,

$$C_k = c_o \Delta_k \tag{6}$$

where c_o is the capacity associated with b_o, and

$$\Delta_k = k_2 - k_1 + 1 \tag{7}$$

is the number of fundamental unit bandwidth divisions handled by transponder T_k.

Earth station equipment inventory

Burst assignment and scheduling may be subject to constraints on the earth station equipment inventory or, conversely, equipment requirements may be determined as a result of the scheduling procedure. The earth segment or equipment inventory can generally be represented by the set $\{Q_{lp}, l = 1, 2, \ldots, \lambda; p = 1, 2, \ldots\}$, where Q_{lp} is the amount of equipment of type p at station S_l. Before earth station equipment requirements are considered, it is useful to introduce the concept of burst overlap and to discuss the basic system constraints.

BURST OVERLAP

The criteria for burst overlap in the time and frequency domains will be examined first; the impact of such overlaps on the earth station equipment inventory will be considered later.

The application of any TDMA scheduling algorithm will result in a series of bursts, each having a certain up-link carrier frequency (and a corresponding down-link frequency), a certain time duration in a given transponder, and a certain starting time. Each burst originates at a certain station and terminates at certain other stations. Thus, if individual bursts are sequentially labeled $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{N_\beta}$, where N_β is the total number of bursts within the satellite over the repeat length, τ, then each $\hat{\beta}_\rho (1 \leq \rho \leq N_\beta)$ is characterized by a carrier frequency, start and end times, transmitting and receiving stations, up- and down-link beams, and a transponder.

For all practical purposes, it is sufficient to consider only a discrete set of carrier frequencies $\omega_1, \omega_2, \ldots, \omega_L$, distributed over the set of bandwidth divisions $\{c_n, 1 \leq n \leq \nu\}$, and hence over the set of transponders $\{T_k, 1 \leq k \leq \nu\}$.

The time variable may be treated as a continuous variable or a discrete one.* Each option offers certain advantages and disadvantages. A continuous treatment is more convenient for analytical purposes as well as for minimizing inefficient utilization of the time space. A discrete treatment, on the other hand, is more suitable for bookkeeping purposes, such as TDMA frame synchronization, and for modular utilization of the time space. The discrete treatment, in which the full frame length (repeat

---

*A suitable time slot size may be chosen on the basis of the conventional option of assigning traffic in multiples of a quantum size, such as 6, 12, or 60 voice channels.

length), τ, is divided into a large number, N, of time slots $t_1, t_2, \ldots, t_N$ of equal duration

$$\delta = \frac{\tau}{N} \tag{8}$$

will be adopted here. If necessary, continuous treatment can be approximated to any degree by simply taking the limits $\delta \to 0$, $N \to \infty$ such that the product $\delta N \to \tau$, the constant repeat length.

As mentioned previously, the individual burst $\hat{\beta}_\rho$ can be labeled according to the base carrier frequency, $\hat{\omega}_\rho$; the beginning time slot, $\hat{t}_\rho$; the end time slot, $\hat{t}_{\rho'}$; the transmitting station, $\hat{S}_\rho$; the receiving station(s), $S_{\rho'}(S_{\rho''}, S_{\rho'''}, \ldots)$; the up-link beam, $\hat{B}_\rho$; the down-link beam, $\hat{B}_{\rho'}$; and the transponder, $\hat{T}_\rho$. Then, symbolically,

$$\hat{\beta}_\rho = \hat{\beta}_\rho(\hat{S}_\rho \to \hat{S}_{\rho'}, \hat{S}_{\rho''}, \ldots; \hat{\omega}_\rho; \hat{t}_\rho \to \hat{t}_{\rho'};$$
$$\hat{B}_\rho \to \hat{B}_{\rho'}; \hat{T}_\rho), \quad 1 \leq \rho \leq N_\beta, \quad \hat{t}_{\rho'} \geq \hat{t}_\rho \tag{9}$$

where the entries on the right are members of the prescribed sets* and station, beam, and transponder accessibility and connectivity are assumed. (That is, $\hat{T}_\rho$ is assumed to have up-link connectivity with $\hat{B}_\rho$ covering $\hat{S}_\rho$ and down-link connectivity with $\hat{B}_{\rho'}$ covering $\hat{S}_{\rho'}, \hat{S}_{\rho''}, \ldots$.)

Two bursts $\hat{\beta}_u$ and $\hat{\beta}_\nu$ will be assumed to have a time overlap if one or more of the time slot indices within the inclusive intervals $(\hat{t}_u, \hat{t}_{u'})$ and $(\hat{t}_\nu, \hat{t}_{\nu'})$ are identical. This will occur whenever the following two inequalities hold†:

$$\hat{t}_\nu < \hat{t}_{u'} \tag{10a}$$

and

---

*That is, $\{\hat{S}_\rho, \hat{S}_{\rho'}, \hat{S}_{\rho''}, \ldots\} \in \{S_l, 1 \leq l \leq \lambda\}$; $\{\hat{\omega}_\rho\} \in \{\omega_r, 1 \leq r \leq L\}$; $\{\hat{t}_\rho, \hat{t}_{\rho'}\} \in \{t_\eta, 1 \leq \eta \leq N\}$; $\{\hat{B}_\rho, \hat{B}_{\rho'}\} \in \{B_m, 1 \leq m \leq \mu\}$; and $\{\hat{T}_\rho\} \in \{T_k, 1 \leq k \leq \nu\}$.

†Due to the finite length of each time slot, an overlap may also be said to occur when $\hat{\beta}_u$ starts at the same time slot at which $\hat{\beta}_\nu$ ends, or vice versa. Such a time overlap is described by changing either one of the two (but not both) inequalities in (10a) and (10b) into an equality.

$$\hat{t}_u < \hat{t}_{\nu'} \quad . \tag{10b}$$

Similarly, two bursts $\hat{\beta}_u$ and $\hat{\beta}_\nu$ will be assumed to have a frequency overlap if they share the same carrier frequency ($\hat{\omega}_u = \hat{\omega}_\nu$), a beam overlap if they have beam overlaps in up- and/or down-link transmissions ($\hat{B}_u = \hat{B}_\nu$ and/or $\hat{B}_{u'} = \hat{B}_{\nu'}$), and a transponder overlap if they occupy the same transponder ($\hat{T}_u = \hat{T}_\nu$).

SYSTEM CONSTRAINTS

Some typical system constraints involved in a TDMA scheduling algorithm are as follows:

   a. Burst isolation: No two bursts $\hat{\beta}_u$ and $\hat{\beta}_\nu$ can have both a time overlap and a transponder overlap. Since this is a universal constraint, compliance is best achieved by requiring that no two non-disjoint beams B_m and $B_{m'}$ be connected to transponders having the same input/output carrier frequency or, more generally, sharing the same bandwidth and polarization.

   b. Burst format: Each burst must be accompanied by a preamble containing coded bits for demodulator synchronization, unique word, station identification, and engineering requirements [7]. Consequently, the total content, $\hat{\alpha}_u$ (in voice channels or bits), of a burst $\hat{\beta}_u$ is actually composed of two distinct parts:

$$\hat{\alpha}_u = \frac{C_u}{N}(t_{u'} - t_u + 1) = \hat{\Pi}_u + \hat{\Gamma}_u \tag{11}$$

where C_u is the reference capacity of transponder $\hat{T}_u$, $\hat{\Pi}_u$ is the content of the preamble, and $\hat{\Gamma}_u$ is the traffic component. For simplicity, the preamble component will be assumed to be a constant, v_o (i.e., $\hat{\Pi}_u = v_o$, $1 \leq u \leq N_\beta$). Furthermore, it will be assumed that any guard space required to prevent time overlap of consecutive bursts will be absorbed in the preamble component. At the start of each TDMA frame, a synchronization burst is transmitted by the reference station for network timing and synchronization [15] [19]. This will also be treated as a special form of preamble.

   c. Maximum burst length: In practice it is convenient to introduce a parameter θ_m, called the maximum burst length. The length of any burst $\hat{\beta}_u$ in the transponder frame must not exceed θ_m:

$$0 < \theta_u = (\hat{t}_{u'} - \hat{t}_u + 1)\, \delta \leq \theta'_m \quad . \tag{12}$$

In units of voice channels (or bits), the maximum burst length, θ_m, can be written as

$$\theta_m = \theta'_m \frac{C_u}{\tau} \quad . \tag{13}$$

The maximum burst length constraint is relevant to the minimization of earth station equipment requirements, since the equipment of a receiving station must usually be engaged continuously from the burst preamble until the data for a given station are received, even if the burst content during the intervening time is of no interest to that station.

d. Digital speech interpolation: If certain parts of the traffic can be transmitted using digital speech interpolation (DSI) [20], the resulting TDMA/DSI burst sizes are smaller than the original (TDMA without DSI) burst sizes by a factor $\zeta(\simeq 1/2)$:

$$\theta_u^{(D)} = \zeta \theta_u \tag{14}$$

where the superscript D denotes the burst length after DSI gain. The actual DSI gain increases with the burst size so that in general ζ may be a function of θ_u.

EARTH STATION EQUIPMENT

Excellent descriptions of TDMA earth station equipment are available in the literature [7], [8]. For the present study, it is sufficient to divide this equipment into two categories: the TDMA terminal (or simply terminal) and the hopping converter (or simply converter). The terminal refers to the assembly of hardware used for transmission/reception of TDMA bursts and includes such components as the multiplexer/demultiplexer, differential encoder/decoder, modulator/demodulator, and preamble generator/detector. Note that a single terminal can simultaneously perform the transmit and receive functions. The converter refers to the hardware assembly involved in IF-to-RF conversion and includes the splitter, up/down frequency converter, and local oscillator. For consistency, it will be assumed that one converter will be required for each discrete carrier frequency transmitted or received by a station, although the use of a single hopping converter in conjunction with multiple local oscillators may per-

mit frequency hopping (via transponder hopping) in burst transmission or reception.

To analyze and evaluate scheduling algorithms, the number of terminals and converters needed at various stations to implement schedules obtained from a particular algorithm must usually be determined. For a schedule obtained from a given algorithm subject to the specified constraints, let $Q_{i1}^{(\eta)}$ denote the number of TDMA terminals required at station S_i for time slot t_η. In addition, let $\{Q_{i1}^{(\eta)}\}$ denote the set of N numbers obtained for N time slots $t_1, t_2, \ldots, t_N$ comprising the frame length, τ. The required number of TDMA terminals, Q_{i1}, at station S_i is then given by the maximum value in the specified set

$$Q_{i1} = \max_\eta \{Q_{i1}^{(\eta)}\}, \qquad 1 \leq \eta \leq N \quad . \tag{15}$$

Finally, if $[Q_{i1}]$ denotes the set of TDMA terminals required at S_i for various algorithms, the minimum number of terminals required within the limits of the attempted algorithms (local optimum) is given by

$$\overline{Q}_{i1} = \min_a [Q_{i1}] = \min_a \left[\max_\eta \{Q_{i1}^{(\eta)}\} \right] \tag{16}$$

where the suffix a indicates algorithmic variation.

Similarly, the minimum number of converters, $\overline{Q}_{i2}$, at station S_i can be obtained as follows. Let $Q_{i2}^{(\eta)}$ represent the number of bursts in time slot t_η which have mutually distinct carrier frequencies and S_i as the transmitting or receiving station. Let $\{Q_{i2}^{(\eta)}\}$ denote the set resulting from considering these bursts for the set of all time slots $\{t_\eta, 1 \leq \eta \leq N\}$. Then,

$$\overline{Q}_{i2} = \min_a \left[\max_\eta \{Q_{i2}^{(\eta)}\} \right] \quad . \tag{17}$$

The preceding criterion of optimality, known as the mini-max criterion (see for example, Reference 21), can obviously be generalized for the whole network by linear summation of Q_{ip} ($p = 1, 2$). More generally, it is possible to attempt to minimize the objective function f_Q,

$$f_Q = \sum_{i=1}^{\lambda} \sum_{p=1}^{2} w_p(\overline{Q}_{ip} - Q_{ip}) \tag{18}$$

where w_p is a weighting factor related to the cost and/or tradeoff priority

of equipment of type p, and Q'_{ip} is the absolute minimum amount of p-type equipment required at station S_i. Q'_{ip} can be determined on the basis of any suitable criterion; for instance, Q'_{i1} is given by the total traffic of station S_i divided by the reference capacity of a transponder (or the average reference capacity for transponders of varying capacities). A search for a global optimum solution can be conducted to obtain the minimum value of the objective function. In addition, the objective function can be further generalized to incorporate the space segment (transponder utilization). It is useful, however, to have independent measures of transponder utilization efficiency as well.

Transponder utilization efficiency

Due to system constraints, it may not always be possible to utilize all of the satellite capacity available for TDMA bursts,

$$v_T = \sum_{k=1}^{\nu} C_k$$

where v_T is referred to as the satellite reference capacity. If the capacity used by burst $\hat{\beta}_s$ is denoted as v_s, the total capacity utilized by all bursts is

$$v_\beta = \sum_{s=1}^{N_\beta} v_s \quad . \tag{19}$$

A portion of v_β is devoted to overhead, specifically preambles, and the remainder contains traffic data Γ_{ij} from station S_i to station S_j $(i, j = 1, 2, \ldots, \lambda)$. The overhead capacity is given by

$$v_\pi = v_o N_\beta \tag{20}$$

where v_o is the capacity devoted to a unit preamble associated with each burst. The difference

$$v'_\Gamma = v_\beta - v_\pi \tag{21}$$

is the so-called configured capacity. Since some of the slots may be only partially filled with traffic due to the discrete nature of the time slots, it is useful to introduce the concept of assigned capacity, $v_\Gamma(\leq v'_\Gamma)$, representing the actual amount of traffic assigned over the transponders. (If all traffic is scheduled, $v_\Gamma = \sum_{i=1}^{\lambda} \sum_{j=1}^{\lambda} \Gamma_{ij}$.)

Now the following ratios are defined as measures of transponder utilization performance or efficiency:

$$\phi_o = \frac{v'_\Gamma}{v_\beta} \tag{22a}$$

$$\phi_\beta = \frac{v_\beta}{v_T} \tag{22b}$$

$$\phi_\Gamma = \frac{v_\Gamma}{v_T} \tag{22c}$$

$$\phi'_\Gamma = \phi_o \phi_\beta = \frac{v'_\Gamma}{v_T} \tag{22d}$$

where ϕ_o is the transmission throughput, Φ_Γ is the overall TDMA scheduling efficiency, and ϕ'_Γ is the overall transponder fill factor.

TDMA scheduling algorithms

General discussion

The burst assignment and scheduling problem results from the multiplicity of available transponders (which in turn results from the multiplicity of frequency bands and antenna beams) and time slots in the transponders. The assignment and scheduling process consists of selecting one of many possible choices for placing a particular traffic element in the time-frequency (or, more generally, transponder-time) space. On the basis of practical considerations, it is useful to introduce certain hierarchical schemes based on the user input and system parameters.

Another aspect of the scheduling algorithm is a priority scheme for determining the sequential order in which the traffic elements are processed. The vacancies in the time-frequency space continually decrease as more and more traffic elements are scheduled. Indeed, the number of vacancies at the time a particular element is scheduled depends on the number and size of traffic elements already scheduled, which are in turn a function of the priority scheme. Thus, the priority scheme affects the outcome of the scheduling in a nonlinear manner. Clearly the element size plays an important role. Consequently, the degree of flexibility in a scheduling process using the same algorithm applied to different traffic data can best be enhanced by allowing the priority scheme to be a direct and simple function of characteristic traffic size.

With these factors in mind, the related algorithms can be discussed more specifically. First, sequential assignment of transponder-time space will be described based on the criterion that traffic elements are processed in a sequential order. (The exact sequence is governed by a priority scheme.) In addition, a transponder is selected for a particular element by sequential examination of the transponder set. (The exact sequence is governed by the beam and transponder hierarchical scheme, and the element is placed in the first available member of the sequence.) Finally, within a transponder, it is assumed that the time slots are utilized in a sequential order corresponding to their serial numbers, since, except for burst overlap characteristics, all time slots are intrinsically equivalent. The generalization of this algorithm to account for the burst overlap criteria as well as other system constraints will be considered later in this paper.

Sequential assignment

For the present study, it is useful to treat the problem of TDMA scheduling as a multistage decision process. Therefore, without loss of generality, suppose that, at any stage of the process, a subset of traffic elements Γ' $= \{\Gamma'_{ij}, i, j = 1, 2, \ldots, \lambda'\}$, belonging to a subset of stations $S' = \{S'_1, S'_2, \ldots, S'_{\lambda'}\}$, is to be assigned to appropriate regions of the transponder-time space. It is assumed that Γ', which will be called a partial traffic matrix, is a subset* of the full traffic matrix Γ such that all relevant transmitting stations are in beam region B_u and all relevant receiving stations are in beam region B_v. Suppose that a subset of transponders $T' = \{T'_k, k = 1, 2, \ldots, \nu'\}$ connects to B_u and B_v for up- and down-link transmission, respectively, and that C'_k is the reference capacity of T'_k. Then the TDMA assignment and scheduling problem consists of distributing the members of the subset of traffic elements Γ' over the time frames of the members of the subset of transponders T' with appropriate time ordering. Several feasible solutions may exist, and the solution which optimizes the parameters of system utilization efficiency is the optimum solution.

For convenience, the traffic assignment or distribution is described in terms of the total transmit traffic, σ'_i, for station S'_i belonging to subset S',

$$\sigma'_i = \sum_j \Gamma'_{ij}, \qquad j = 1, 2, \ldots, \lambda' \quad . \tag{23}$$

If the traffic transmitted from station S'_i and assigned to transponder T'_k

---

*The prime denotes specified subsets of the respective original sets.

is n_{ik} $(i = 1, 2, \ldots, \lambda'; k = 1, 2, \ldots, \nu')$, the burst assignment problem involves determining the values of n_{ik}, subject to the constraints

$$\sum_{k=1}^{\nu'} n_{ik} = \sigma'_i, \qquad i = 1, 2, \ldots, \lambda' \tag{24a}$$

$$\sum_{i=1}^{\lambda'} n_{ik} \leq C'_k, \qquad k = 1, 2, \ldots, \nu' \tag{24b}$$

and

$$n_{ik} \geq 0 \quad . \tag{24c}$$

In addition, the time ordering of the n_{ik}'s within the transponder frames must be determined.

The problem is represented in tabular form in Figure 1. The partial traffic matrix containing only the set of elements $\{\Gamma'_{ij}\}$ is shown on the left, with the row sums σ'_i as indicated. The transponder labels T'_k appear in the top row on the right and the respective transponder capacity C'_k in the bottom row, with the requisite numbers n_{ik} between them. The time frame for each transponder is shown as an independent system coordinate on a slant scale at the bottom, with slots 1 through N marked serially along the bottom right side.

The assignment scheme can be explained by arbitrarily starting with the transmitting station at the top, S'_1, and transponder T'_1 of the first column and determining the corresponding "north-west" element n_{11}. Then, for the subsequent transmitting stations $S'_2, S'_3, \ldots$, the subsequent transponders $T'_2, T'_3, \ldots$, as necessary, are considered sequentially. It should be noted that the time frames of each transponder should be filled as much as possible before going to the next transponder. Since T'_1 cannot contain more traffic than its capacity, C'_1,

$$n_{11} = \min \{\sigma'_1, \overline{C}_1\} \tag{25}$$

where $\overline{C}_1 = C'_1$ if any link can be split in two or more parts. Otherwise, $\overline{C}_1$ represents the sum in equation (23) excluding the traffic elements that cause the sum to exceed C'_1.

Once the first north-west element n_{11} has been determined, all other elements in either the top row or the first column (but not both) can be determined as follows:

a. If $\sigma'_1 < C'_1$, set $n_{11} = \sigma'_1$, and set all the remaining elements of the top row equal to zero. Also, replace the value of column sum C'_1 with that of $C'_1 - \sigma'_1$.

b. If $\sigma'_1 > C'_1$, set $n_{11} = \bar{C}_1$, and set all the remaining elements of the first column equal to zero. Also replace the value of row sum σ'_1 with that of $\sigma'_1 - \bar{C}_1$.

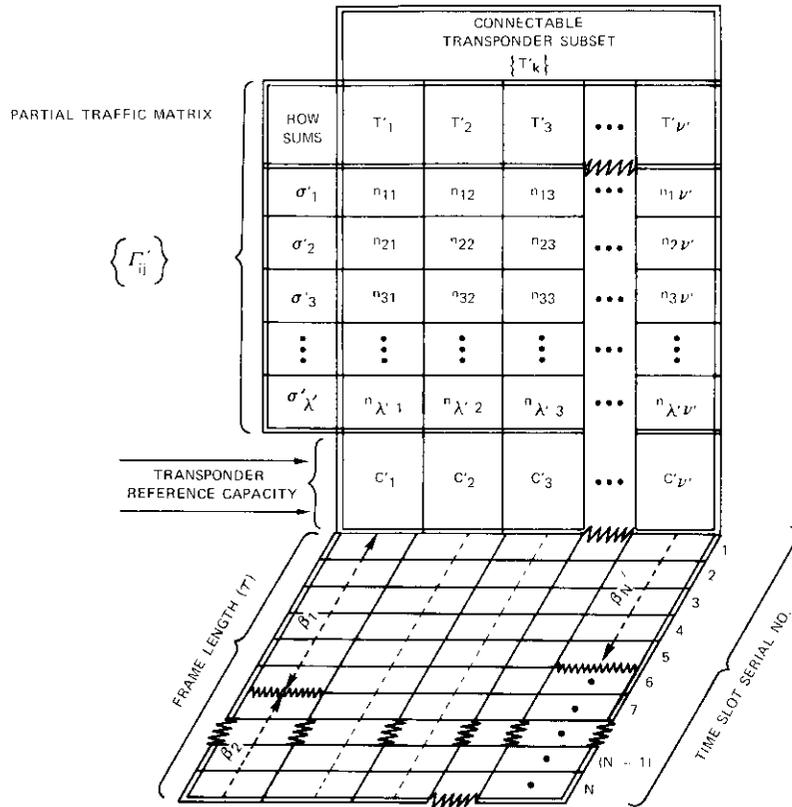*c.* If $\sigma'_1 = \bar{C}_1$, follow either step a or step b (but not both).

Figure 1. *Schematics of Transponder Assignment and Scheduling of the Traffic Elements of Subset* $\{\Gamma'_{ij}\}$

In Figure 1, if condition *a* is assumed, this procedure simply implies that the outgoing traffic σ'_1 of station S'_1 is less than the capacity of trans-

ponder T'_1. Therefore, all this traffic is transmitted through transponder T'_1, and the residual capacity of transponder T'_1 equals the difference $C'_1 - \sigma'_1$. Hence, all the traffic elements Γ'_{11}, Γ'_{12}, Γ'_{13}, ..., are placed in time slots $1, 2, \ldots,$ of T'_1, which constitute the first burst, β_1; the respective time slots of T'_1 are marked accordingly as shown in Figure 1.

Similarly, the next north-west element n_{21} is determined by comparing the residual values of the corresponding row and column sums, that is, by setting

$$n_{21} = \min \{\sigma'_2, C'_1 - \sigma'_1\} \tag{26}$$

and following steps analogous to *a*, *b*, or *c* as applicable. This process is continued until all traffic elements of Γ' are appropriately assigned over the transponder subset T' and the respective time slots, leading to the formation of a total of N' bursts.

Subject to the assumption that the net capacity exceeds the net traffic,

$$\sum_k C'_k > \sum_i \sigma'_i \tag{27}$$

there may still be some vacant spaces (indicated by the dark circles in Figure 1, for instance) in the transponder-time space toward the end of the sequence. Conversely, if the net traffic exceeds the net capacity, some traffic remains to be scheduled at the end of the process. It should be noted that this procedure is analogous to the so-called "north-west corner rule" algorithm for obtaining a "basic feasible solution" of the theory of transportation problem in the literature of operations research and mathematical programming (see, for example, References 21–24).

An optimal solution can be obtained by starting from this type of basic feasible solution and using the so-called "perturbation method" of the transportation problem. The optimization consists of methodically modifying the values of the n_{ik}'s obtained above to minimize a cost-effective objective function,

$$f_o = \sum_i \sum_k n_{ik} c_{ik} \tag{28}$$

where c_{ik} is the cost of assigning one voice channel of traffic data from S'_i to transponder T'_k.

Before these optimization considerations are described in detail, the method of generalizing the aforementioned procedure to incorporate

various operational constraints will be described. First, the subset Γ' = $\{\Gamma'_{ij}, i, j = 1, 2, \ldots, \lambda'\}$ is chosen according to any arbitrary priority scheme, with or without regard to beam coverage. In general, there is now a multiplicity of beam coverages for the transmitting and receiving stations of the resulting partial traffic matrix. Hence, the entire transponder set $T = \{T_k, k = 1, 2, \ldots, \nu\}$ (rather than a subset thereof) must normally be considered in connection with the assignment/scheduling of its traffic elements, and a systematic or hierarchial search for appropriate transponders must be conducted for each traffic element.

Other system constraints mentioned previously are also explicitly superimposed on the preceding assignment scheme. For example, a preamble must be provided to accompany each burst; the burst length must be subject to an upper limit, *i.e.*, the maximum burst length; TDMA/DSI compression must be provided; and excessive burst overlap must be avoided to minimize the equipment requirements of transmitting and receiving stations. The exact sequential order in which the transponders and the time slots are used may be (and generally is) modified due to the presence of these constraints. Hence, the scheduling must be performed dynamically. In other words, the scheduling decisions at any stage can be made only by examining the result of all the relevant prior decisions; these decisions will then affect all subsequent decisions.

For instance, the total preamble or overhead requirement involved cannot be determined *a priori*, but will depend on how often a new burst must be started for the transmitting station to comply with the maximum burst length constraint or the minimum burst overlap condition. The latter, in turn, will require that certain time slots in a given transponder be skipped in the scheduling of a given traffic element, leaving gaps in the transponder frame which may or may not be filled by another traffic element at a later stage of the scheduling process. Similar to the total overhead, the total number of vacancies or the amount of unutilized capacity resulting from the nonoverlap condition cannot be determined *a priori*; instead, it may be a sensitive function of the input data structure and the dynamic decision procedures involved.

Certain other special system considerations have also been incorporated and will be mentioned subsequently. At this point it is sufficient to note that the generalized north-west corner rule type of algorithm is conveniently used in the framework of the mathematical model described previously to obtain a basic feasible solution of the TDMA scheduling problem. However, a completely analytical method for optimizing the solution, such as the perturbation method of the theory of transportation

problem, is impractical due to the inherent size of the problem and the complexity introduced by the system constraints. This limitation is compounded by the fact that the cost factors c_{ik} in equation (28) usually cannot be determined systematically or objectively. In view of these difficulties, a search for the optimum solution is made heuristically by using certain control parameters governing the priority scheme and burst configuration. These parameters will be described in the following subsection, along with the computer implementation of the preceding mathematical model and algorithm.

Computer model

The following is a brief summary of the important steps and options implemented in a computer program for TDMA scheduling recently developed by the author. The scheduling is performed by first reordering the rows and columns of the traffic matrix (without altering its symmetric character) on the basis of beam coverage and traffic size so that narrower beams and stations with a larger total traffic requirement are given a higher hierarchical order. The rearranged traffic matrix (hereafter denoted as Γ) is sorted into four partial traffic matrices

$$\Gamma = \Gamma^{(1)} + \Gamma^{(2)} + \Gamma^{(3)} + \Gamma^{(4)} \tag{29a}$$

$$\Gamma_{ij} = \sum_{n=1}^{4} \Gamma_{ij}^{(n)}, \qquad i, j = 1, 2, \ldots, \lambda . \tag{29b}$$

In equation (29a), $\Gamma^{(1)}$ includes, in part, those elements of Γ which are large enough to equal or exceed the capacity of a transponder (minus the necessary overhead). A part of such an element is extracted to form the corresponding element of $\Gamma^{(1)}$ so that this element (plus the necessary overhead) completely fills one or more transponders. The remaining elements of $\Gamma^{(1)}$ are zero. Similarly, the nonzero elements of $\Gamma^{(2)}$ are those elements of $\Gamma - \Gamma^{(1)}$ whose size exceeds the user-specified maximum burst length, θ_m. $\Gamma^{(3)}$ consists of those elements of $\Gamma - \Gamma^{(1)} - \Gamma^{(2)}$ which correspond to stations with net 1-way traffic less than a user-specified lower limit, θ_n. Finally, the elements of $\Gamma^{(4)}$ are simply the remaining elements [i.e., $\Gamma^{(4)} = \Gamma - \Gamma^{(1)} - \Gamma^{(2)} - \Gamma^{(3)}$]. Note that, except for the possible splitting of links to form the nonzero elements of $\Gamma^{(1)}$, each link specified in Γ appears in one and only one partial traffic matrix.

This classification of traffic data is useful for exercising flexible control over the priority scheme for various traffic elements in the scheduling

process. The elements of $\Gamma^{(1)}$ are assigned and scheduled first, followed by those of $\Gamma^{(2)}$, $\Gamma^{(3)}$, and $\Gamma^{(4)}$, in that order. The elements of a particular partial traffic matrix are scheduled according to their row numbers referred to the rearranged matrix Γ. The elements of the same row are scheduled according to their column numbers (referred to Γ) or, optionally, according to their relative sizes, with the largest first. At this stage each element is considered to be indivisible and is scheduled integrally, thereby minimizing the number of up-chains. A particular element is assigned/scheduled by searching for the appropriate up- and down-link beams, transponders, and vacant time segments (consisting of a sufficiently large number of contiguous empty time slots), in that order, subject to user-specified beam coverage, transponder connectivity, and hierarchy. After a suitable vacancy is found, the final assignment is made by checking for burst overlap and, if necessary, moving the burst (or a single link component thereof, called the sub-burst) to another suitable vacancy to avoid overlap. A provision for necessary preambles is included throughout, as is a provision for evaluating various parameters of transponder utilization efficiency and for determining the required earth station equipment inventory.

This scheme tends to optimize the utilization of transponders, terminals (transmitters), and up-converters by minimizing overhead (preamble), subject to the maximum burst length constraint. Furthermore, this scheme assigns higher priority to larger traffic elements and stations with large traffic requirements, and at the same time tends to prevent excessive equipment requirements for stations with very small traffic requirements. Drastic variations in the implicit priority scheme described above can be effected simply by varying the input values of the parameters θ_m and $\bar{\theta}_m$. Almost any arbitrary distribution of elements within $\Gamma^{(2)}$, $\Gamma^{(3)}$, and $\Gamma^{(4)}$ can be achieved by varying the θ_m and $\bar{\theta}_m$ values over a suitable range. In particular, a very small value of $\bar{\theta}_m$ will cause all remaining nonzero elements to be excluded from $\Gamma^{(3)}$. The converse applies to the choice of θ_m for inclusion or exclusion of elements with respect to $\Gamma^{(2)}$. A very small value of θ_m is not recommended, however, since it will lead to excessive overhead requirements associated with the small value of the implied maximum burst length.

Several options for choosing the overall priority in the scheduling process are provided in the program. For example, the elements can be sequentially packed in transponders without checking for burst overlaps or leaving gaps in the transponder frames, thereby optimizing the fill factor. Alternatively, as mentioned previously, the relevant transponder-time

space may be scanned before each traffic element is assigned to avoid excessive burst overlaps, thereby minimizing equipment (terminal) requirements for transmitting or receiving stations. If excessive equipment requirements are unavoidable, it is possible either to accept the increased requirements and proceed with scheduling, or to omit the specific traffic element subject to the specified minimum equipment constraint. Similarly, if no transponder-time space is available for a particular traffic element, scheduling may be continued with the remaining elements or terminated altogether so that a fresh trial with modified input values of the control parameters can be made.

There are two options for overlap checking. It can be done only for the time span $\Delta\tau$ occupied by the individual link (sub-burst) to be assigned/scheduled at a particular stage, or for the time span $\Delta\tau'$ occupied by the entire burst (including preamble) up to and including the sub-burst in question. These two methods of overlap checking are shown schematically in Figures 2a and 2b, respectively. In the case of a unidestinational burst or a multidestinational burst with TDMA/DSI, the terminal must be engaged for reception only for the sub-burst duration (plus the corresponding preamble), and the overlap checking method shown in Figure 2a is sufficient. In other cases, full burst overlap checking (Figure 2b) must be performed, since the terminal must be fully engaged from the start of the preamble to the end of the sub-burst in question. In Figure 2a for the element $\Gamma_{ij''}$, there are only two overlaps (requiring two terminals) if receiving station $S_{j''}$ processes its traffic in transponders T_{k-1}, T_k, and T_{k+1}. On the other hand, in Figure 2b there are three overlaps (requiring three terminals) for station $S_{j''}$.

It is usually possible to improve the efficiency of transponder utilization and earth station equipment requirements by suitably adjusting the options and the values of the control parameters. The choice is dictated by system parameters such as traffic distribution, satellite configuration, or transponder plan. Hence, it is generally possible to heuristically obtain an optimum or near-optimum schedule.

Capabilities for incorporating user-specified time quantization (slot size), traffic growth, and DSI gain are included in the program. Among the other special optional features provided in the program, the following are noteworthy:

a. *Depolarization isolation:* Climatic conditions at a certain earth station resulting in rain depolarization of its TDMA signal in the centimeter or millimeter wave range may lead to interference with the simultaneously transmitted signals of other earth stations using orthog-
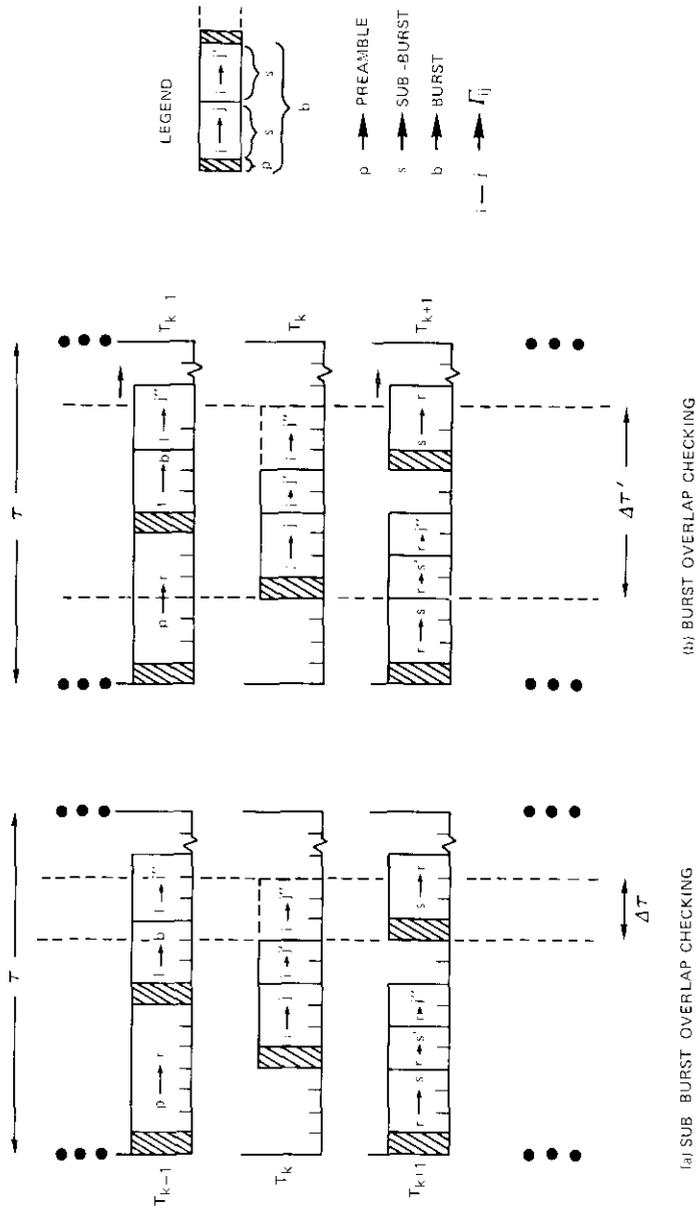
Figure 2. *Two Modes of Overlap Checking in TDMA Scheduling (The dashed arrows in b indicate that the value of $\Delta\tau'$ is modified for the corresponding transponders to include all slots in the burst encountered.)*

---

onal polarization. To avoid such depolarization-induced interference, the signal of the station in question may be isolated in the frame so that the signal of any other stations (or the signal of any member of a certain subset of other stations) is not transmitted during its time slots.

b. Cross-strapping symmetry: If a certain traffic element Γ_{ij} is transmitted through a transponder T_k with characteristic up- and down-link beam frequencies belonging to different frequency pairs Ω_m/Ω_m' and Ω_n/Ω_n' ($\Omega_m \neq \Omega_n$), respectively, then the return traffic Γ_{ji} may be scheduled symmetrically in a cross-strapped transponder T_k' with characteristic up- and down-link beam frequencies belonging to the preceding pairs in inverse order (i.e., Ω_n/Ω_n' and Ω_m/Ω_m', respectively, as shown in Figure 3).
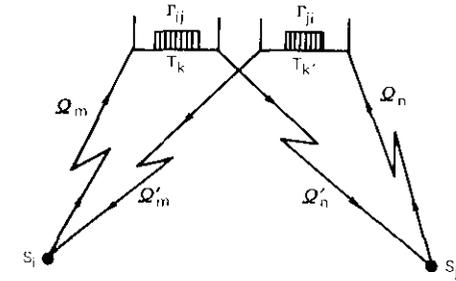

Figure 3. *Cross-strapping Symmetry in Traffic Assignment*

Numerical example

The program described in the preceding subsection has been used to provide TDMA burst schedules using a projected traffic data base and satellite configuration representing the future INTELSAT environment for the Atlantic Ocean region. Schedules can normally be obtained within the limits on available transponders assigned *a priori* on the basis of capacity considerations. With the option of avoiding burst overlap whenever possible, the amount of earth station equipment (specifically, TDMA terminals) required is only marginally (if at all) in excess of the minimum number required on the basis of capacity consideration.

Overall transponder fill factors and TDMA scheduling efficiencies of the order of 80 percent are typically obtained for the system as a whole. With a preamble of six voice channels associated with each burst, the overhead-to-reference capacity is typically less than 5 percent so that the value of the throughput is typically 95 percent. If the size of each time slot is assumed

to be equivalent to the capacity of six voice channels, the ratio of assigned to configured capacity is typically 96 percent; therefore, the capacity degradation due to reasonable quantization of the time scale is only marginal.

Since a specific example involving a large traffic data base is not possible here, a schedule using a small hypothetical input data base is presented. The input traffic matrix is shown in Table 1 and the beam coverage of the earth stations involved is shown in Figure 4. In Figure 5, which summarizes

TABLE 1. EXAMPLE OF TRAFFIC DATA FOR TDMA SCHEDULING

| Station Number and Name	1	2	3	4	5	6	7	8	9	10
1 Argentina	0	104	10	29	56	89	56	23	167	44
2 Brazil	104	0	27	87	113	66	21	66	350	36
3 Canada	10	27	0	18	56	110	9	190	0	26
4 France	29	87	18	0	0	0	0	0	460	38
5 Germany	56	113	56	0	0	0	0	0	678	50
6 Italy	89	66	110	0	0	0	0	0	474	127
7 Spain BU	56	21	9	0	0	0	0	0	230	118
8 U.K.	23	66	190	0	0	0	0	0	1,899	35
9 U.S. Mainland	167	350	0	460	678	474	230	1,899	0	321
10 Venezuela	44	36	26	38	50	127	118	35	321	0
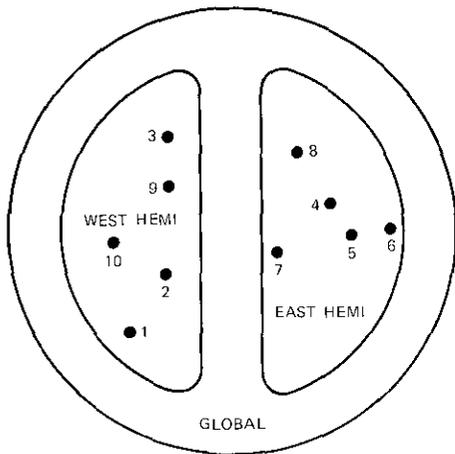
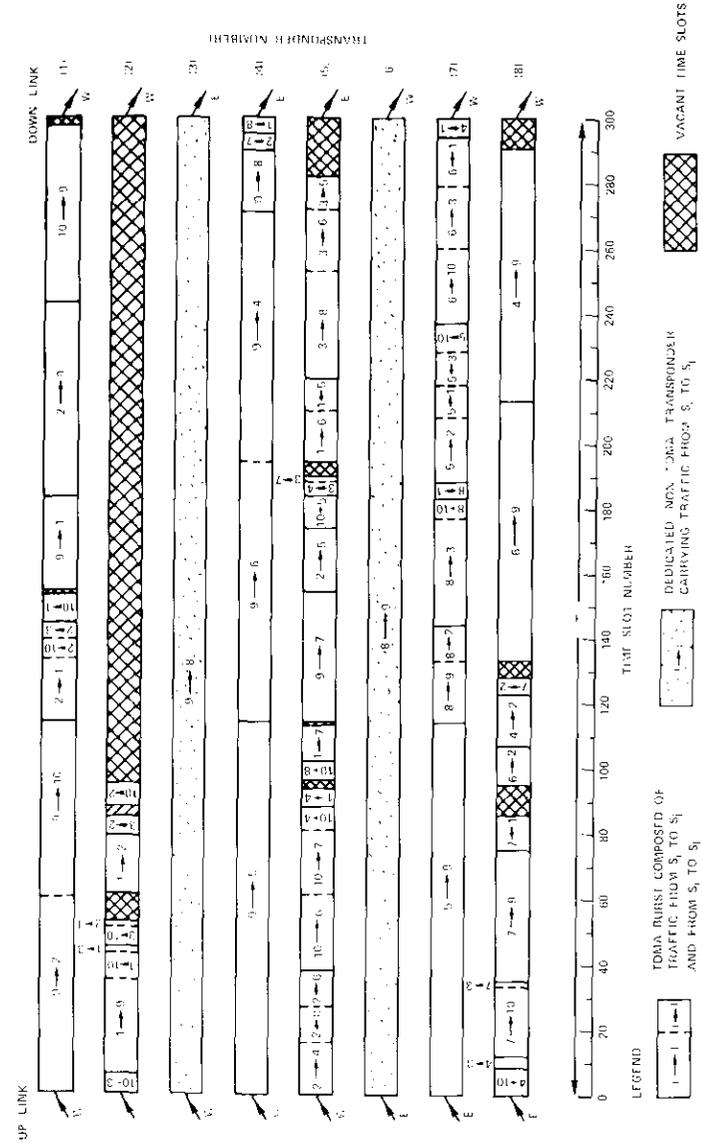Figure 4. *Beam Coverage of the Earth Stations of Table 1*

Figure 5. *Summary of TDMA Burst Assignment and Scheduling for the Traffic Data in Table 1 (transponder connectivities are indicated by the arrows on the left and right sides of each transponder frame; E and W indicate east and west hemispheric beams, respectively)*

the resulting burst configuration for the traffic data of Table 1 and various transponder configurations, it has been assumed that each transponder has a capacity of 1,800 voice channels and that the transponder is divided into 300 time slots. The figures of transponder utilization efficiency are provided in Table 2.

The system performance is generally expected to vary with the number of stations transmitting with TDMA or, equivalently, with the total amount of traffic to be scheduled. This functional dependence is indicated in Figures 6 through 8. The results represented in these figures have been
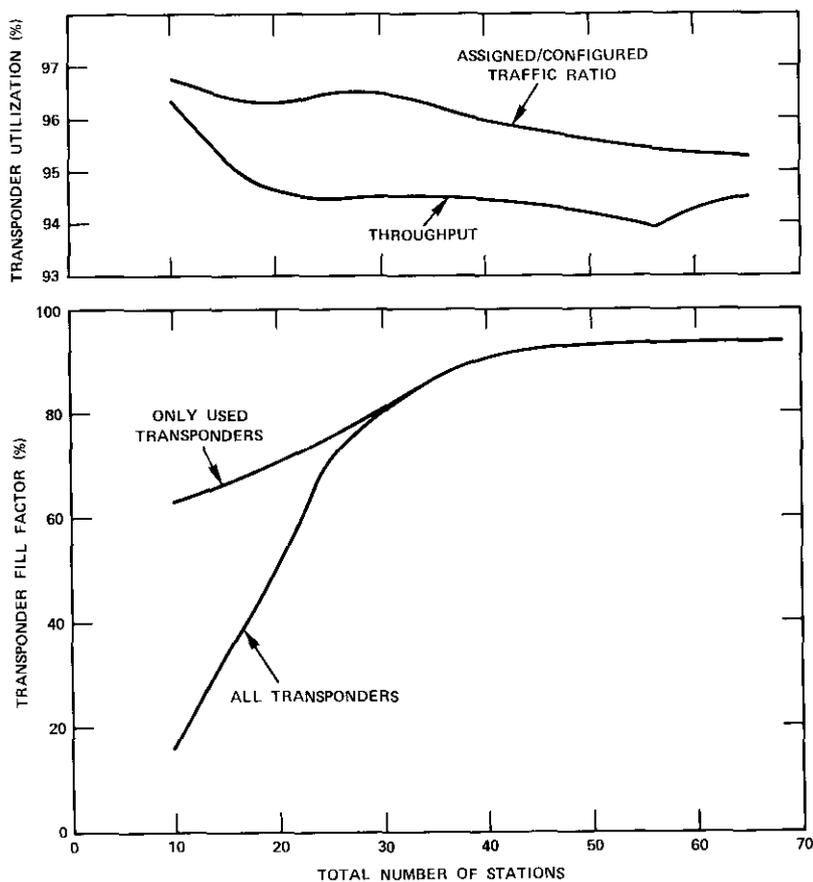

Figure 6. *Variations of Transponder Utilization with Total Number of Stations*

TABLE 2. FIGURES OF TRANSPONDER UTILIZATION EFFICIENCY

ID No.	Up-Link	Down-Link	Reference (R)	Configured (C)	Assigned (A)	No. o Bursts	Preamble Content (P)	(P/R), %	Fill Factor (C/R), %	(A/C), %	Scheduling Efficiency (A/R), %	Throughput [A/(A + P)], %
1	West	West	1,800	1,746	1,720	6	36	2.00	97.00	98.51	95.56	97.95
2	West	West	1,800	474	450	6	36	2.00	26.33	94.94	25.00	92.59
3	West	East	1,800	1,794	1,794	1	6	0.33	99.67	100.00	99.67	99.67
4	West	East	1,800	1,770	1,761	5	30	1.67	98.33	99.49	97.83	98.32
5	West	East	1,800	1,590	1,543	11	66	3.67	88.33	97.04	85.72	95.90
6	East	West	1,800	1,794	1,794	1	6	0.33	99.67	100.00	99.67	99.67
7	East	West	1,800	1,758	1,727	7	42	2.33	97.67	98.24	95.94	97.63
8	East	West	1,800	1,602	1,577	10	60	3.33	89.00	98.44	87.61	96.33
Net System Efficiency			14,400	12,528	12,366	47	282	1.96	87.00*	98.33	85.87	97.26

*The fill factor (including preambles) for all transponders is 88.96 percent.
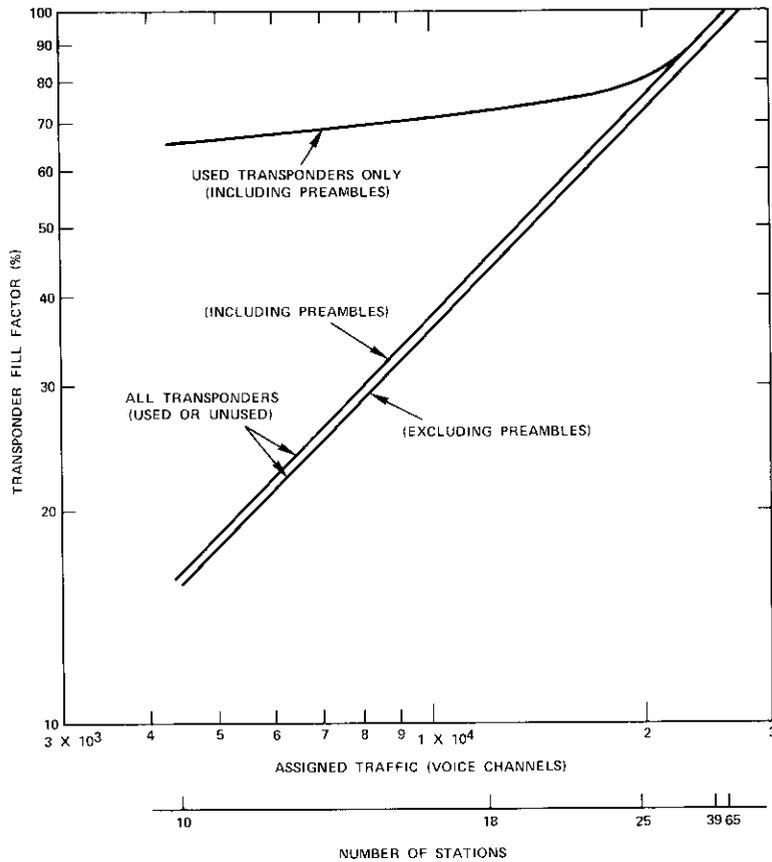
Figure 7. *Variations of Transponder Fill Factor with Assigned Traffic on a Log-Log Plot*
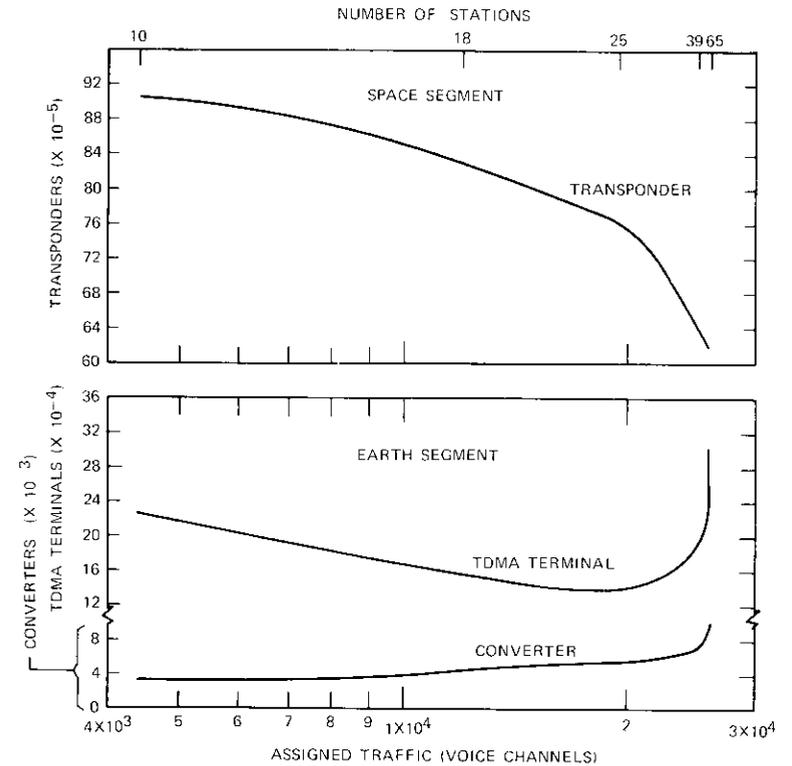

Figure 8. *Amount of Space and Earth Station Equipment Required per Unit (one voice channel) of Assigned Traffic (Vertical scales indicate the amount of equipment required divided by the net assigned traffic.)*

obtained by using a projected sample of the INTELSAT Atlantic Primary traffic data base for the year 1985 and allowing more and more countries (in order of traffic size) to access the TDMA transponders. Of course, only the mutual traffic of the selected set of countries is assigned, and the satellite configuration remains constant from one case to the next.

Figure 6 shows the variation in the figures of transponder utilization efficiency with number of earth stations. The average transponder fill factor, ϕ_{f} [equation (22d)], is indicated at the bottom of the figure for all transponders specified, as well as for the transponders actually utilized.

The fill factor increases almost linearly with increasing TDMA stations, but beyond a certain point the effect of satellite capacity limitations becomes dominant. The loss of efficiency with increased number of earth stations due to overhead and the discrete nature of the time slots is indicated in Figure 6 by the pair of upper curves representing the system throughput, ϕ_{o} [equation (22a)], and the assigned-to-configured traffic ratio, $v_{\Gamma}/v_{I'}$ [see equation (21)], respectively. The variation of these quantities with increasing stations is seen to be only marginal.

The transponder fill factor as a function of the amount of assigned traffic is shown in Figure 7 on a log-log scale. It is interesting to note that the fill factor obtained by considering all the input transponders varies

linearly on this scale, indicating a simple power dependence of the fill factor, ϕ_β, on the assigned traffic, Γ, of the form

$$\phi_\beta = a\Gamma^b \qquad (30)$$

where a and b are constant parameters. For instance, the percentage fill factor shown in Figure 7 can be represented by substituting $a = 10^{-2.5}$ and $b = 1.02$ into the preceding approximate relationship, with Γ expressed in voice channels.

For the different cases used in these examples, the total number of transponders, terminals, and converters required for one unit (voice channel) of transmission in the network (and hence the respective normalized costs) is plotted against the assigned traffic in Figure 8. With proper multiplication factors, these curves can also represent cost trade-offs for space and earth segment utilization with varying traffic size. A curve of total cost vs traffic size (or number of TDMA stations) can also be obtained to determine the optimum level of TDMA operation. More generally, these types of curves can be used to obtain the optimum system design parameters for any specific system. It should be noted, however, that the specific results may be largely dependent on input data and parameters, and gross generalizations are not warranted.

Conclusions

A simple mathematical model of the TDMA burst assignment and scheduling problem for a general satellite communications system has been developed. An algorithm based on sequential assignment of traffic elements in the transponder-time space consistent with system constraints has formed the basis of a computer program for optimizing the solution of the assignment and scheduling problem with respect to the utilization of space and earth segment resources. The feasibility of obtaining an optimum or near-optimum solution by means of this program has been demonstrated.

More specifically, a variety of options are available for obtaining alternate feasible TDMA schedules which are near optimum in terms of the utilization efficiency of the space and earth segment equipment. An attempt has been made to maximize the fill factor of the transponders while minimizing the earth station equipment requirements. Although a simultaneous optimization of the utilization of all types of equipment is not possible, different options can be used to optimize the utilization of

various types of equipment or a weighted combined objective function can be used subject to specified constraints. In addition, parametric variation can be invoked to obtain alternate schedules using the same algorithm. It is expected that a suitable combination of parametric and algorithmic variations will yield a satisfactory scheduling solution for diverse types of space and earth segment configurations and traffic data.

References

[1] J. G. Puente, W. G. Schmidt, and A. M. Werth, "Multiple Access Techniques for Commercial Satellites," *Proc. IEEE*, Vol. 59, No. 2, February 1971, pp. 218–229.

[2] H. Haberle, P. Hein, and S. Dinwiddy, "Combination of Frequency, Time and Space Division Multiple Access in Multi-Transponder Satellite Communications," 2nd International Conference on Digital Satellite Communications, Paris, France, November 1972, *Proc.*, pp. 432–440.

[3] O. G. Gabbard and P. Kaul, "Time Division Multiple Access," IEEE EASCON Conference, Washington, D.C., 1974, *Conference Record*, pp. 179–184.

[4] Y. F. Lum, "A Proposed Time Division Multiple Access (TDMA) Satellite System for ANIK I," IEEE International Conference on Communications, Seattle, Washington, June 1973, *Conference Record*, pp. 8-1–8-5.

[5] W. B. Garner, "A System Comparison of Satellite Switched TDMA and FM-FDMA for Multi-Beam Satellites," 2nd International Conference on Digital Satellite Communications, Paris, France, November 1972, *Proc.*, pp. 441–449.

[6] R. K. Kwan, "A TDMA Application in the TELESAT Satellite System," National Telecommunications Conference, Atlanta, Georgia, November 1973, *Conference Record*, pp. 31E-1–31E-6.

[7] W. G. Schmidt, "The Application of TDMA to the INTELSAT IV Satellite Series," *COMSAT Technical Review*, Vol. 3, No. 2, Fall 1973, pp. 257–274.

[8] J. D. Withers and C. A. Blackwell, "Time Division Multiple Access in the INTELSAT System," *Satellite Multiple Access—The Key to Effective*

Utilization, IEEE International Convention, New York, N.Y., March 1973, pp. 1–7.

[9] S. Browne, "The INTELSAT Global Satellite Communications System," *COMSAT Technical Review*, Vol. 4, No. 2, Fall 1974, pp. 477–487.

[10] W. Schmidt, "An On-Board Switched Multiple Access System for Milli-meter-Wave Satellites," INTELSAT/IEE International Conference on Digital Satellite Communications, London, England, November 1969, *Proc.*, pp. 399–407.

[11] T. Muratani, "Satellite Switched Time-Domain Multiple Access," IEEE EASCON Conference, Washington, D.C., 1974, *Conference Record*, pp. 189–196.

[12] R. Cooperman and W. G. Schmidt, "Satellite-Switched SDMA and TDMA System for Wide Band Multibeam Satellite," IEEE International Conference on Communications, Seattle, Washington, June 1973, *Conference Record*, pp. 12-7–12-12.

[13] W. G. Schmidt, "Satellite Switched TDMA —Transponder Switched or Beam-Switched," AIAA 5th Communications Satellite Systems Conference, Los Angeles, California, April 1974, Paper No. 74-460.

[14] R. A. J. Smith, "TDMA Multiple Transponder Operation by Means of Frequency Hopping," 2nd International Conference on Digital Satellite Communications, Paris, France, November 1972, *Proc.*, pp. 425–431.

[15] N. Shimasaki and R. Rapuano, "Synchronization for a Communications Distribution Center On-Board a Satellite," IEEE International Conference on Communications, 1971, *Conference Record*, pp. 42-20–42-55.

[16] C. Carter and S. S. Haykin, "Precision Synchronization to a Switching Satellite Using PSK Signals," IEEE International Conference on Communications, Minneapolis, Minnesota, June 1974, *Conference Record*, pp. 43D-1–43D-5.

[17] H. Ganssmantel and B. Ekstrom, "TDMA Synchronization for Future Multitransponder Satellite Communications," IEEE International Conference on Communications, Minneapolis, Minnesota, June 1974, *Conference Record*, pp. 43C-1–43C-5.

[18] J. W. Schwartz, J. M. Aein, and J. Kaiser, "Modulation Techniques for Multiple Access to a Hard-Limiting Satellite Repeater," *Proc. IEEE*, Vol. 54, No. 5, May 1966, pp. 763–777.

[19] L. Lundquist, "Optimizing a TDMA Channel Including Synchronization," IEEE International Conference on Communications, Minneapolis, Minnesota, June 1974, *Conference Record*, pp. 43A-1–43A-4.

[20] S. J. Campanella and H. G. Suyderhoud, "Digital Speech Interpolation for Telephone Communications," IEEE EASCON Conference, Washington, D.C., 1974, *Conference Record*, pp. 331–336.

[21] G. B. Dantzig, *Linear Programming and Extensions*, Princeton, New Jersey: Princeton University Press, 1963.

[22] S. I. Gass, *Linear Programming*, New York: McGraw-Hill Book Co., 1964.

[23] M. Simonard, *Linear Programming*, New Jersey: Prentice-Hall, Inc., 1966.

[24] G. S. G. Beveridge and R. S. Schechter, *Optimization: Theory and Practice*, New York: McGraw-Hill Book Co., 1970.

Ashok K. Sinha received a B.Sc. (1963) and an M.Sc. (1965) in physics from Patna University, India, and a Ph.D. in physics from the University of Maryland (1971). As a Research Associate for the U. of Md. (1971–1973) and as a member of the technical staff of Computer Sciences Corporation (1974), he was involved with several NASA-sponsored geophysical research and analysis projects, including ionospheric, upper atmospheric, and geomagnetic field modeling and ionospheric scintillations. He joined COMSAT Laboratories in 1974 as a member of the technical staff of the Transmission Systems Laboratory, where he is currently involved with communications satellite system modeling and long-range planning methodology. He is a member of the American Geophysical Union and the International Association of Geomagnetism and Aeronomy.

Adaptive polarization control for satellite frequency reuse systems

D. F. DiFonzo, W. S. Trachtman, and A. E. Williams

(Manuscript received February 19, 1976)

Abstract

This paper describes networks for the adaptive restoration of polarization orthogonality. These networks may be used to minimize the mutual interference of a dual-polarized satellite communications link operating in the presence of a depolarizing medium. For a properly designed system, the time-varying depolarization due to a variety of sources may necessitate adaptive network control. As described herein, the circuits to accomplish this objective are based on the assumption that the error voltages for adaptive control are derived from narrow-band pilot signals provided within the communications spectrum. In general, four control parameters are required to achieve orthogonality correction, but when differential attenuation may be neglected, a suboptimal control system employing two control parameters leads to a simple and attractive solution.

Introduction

A typical dual-polarized satellite communications link is shown in Figure 1. In such a link the depolarizing effects of the earth station antennas, the satellite antennas, and the propagation medium all combine to

---

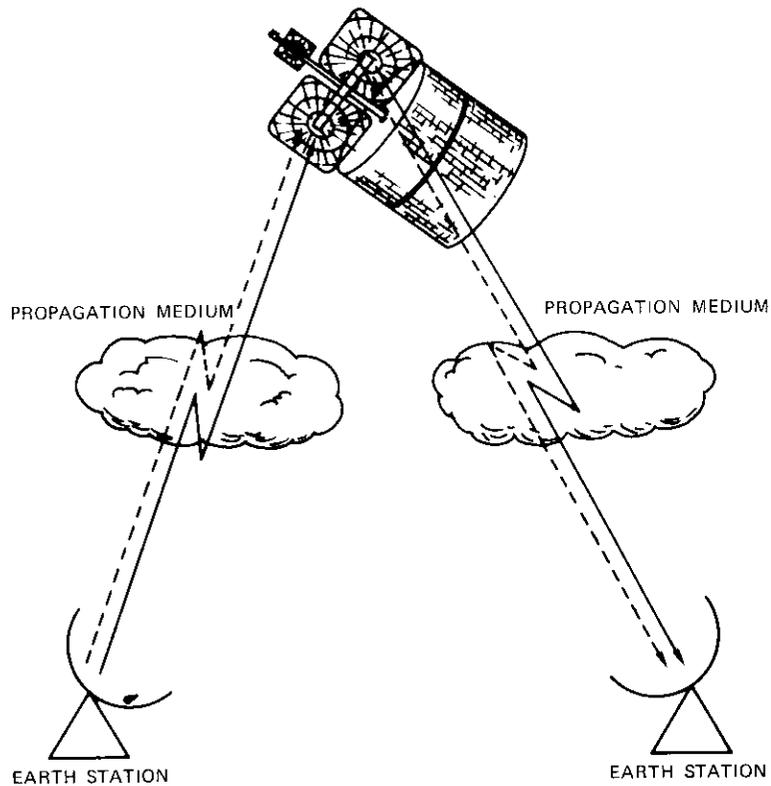limit the isolation between signals in each polarized channel.*

Figure 1. *A Dual-Polarization Satellite Communications Link*

The depolarization effects associated with the spacecraft and earth station antennas and feeds are static; in the absence of other effects, these components can be designed to yield excellent signal-to-interference ratios [1], [2]. To ensure adequate polarization discrimination they are specified to have cross-polarization isolation ratios in the range of 30 dB [3] over

---

* Dual-polarized signals carrying independent information in the same frequency bands are assumed to be uncorrelated. Hence, cross-polarization interference in a given channel appears as a random disturbance and therefore contributes to the total noise budget of the communications link.

both the transmit and receive frequency bands. On the other hand, the depolarization effects of the medium, *e.g.*, rain, are time varying and can produce substantial depolarization due to differential phase shift and attenuation for wave components parallel and perpendicular to the axes of the oblate raindrops [4]-[12]. Two orthogonally polarized waves experiencing differential phase shift will be depolarized, but will maintain their orthogonality. However, differential attenuation actually destroys the orthogonality between the polarization states.

Since rain depolarization is a statistical effect, an adaptive antenna system must be employed to match the polarizations of the input waves and to adaptively maintain this condition with no *a priori* knowledge of propagation conditions [13]. In the presence of a general depolarization medium, four independent parameters must be controlled and hence four independent error voltages must be produced to drive the adaptive system. This control can be most easily achieved by assuming that two pilot signals are assigned within the operational frequency bands (one for each polarization) and that they are detected by phase amplitude phase-lock receivers.

On the basis of this assumption, this paper describes and analyzes three possible networks which might be employed in earth stations at both the receive and transmit frequencies to adaptively compensate for depolarization effects. Signal-to-interference isolations better than 30 dB over a 500-MHz bandwidth can be achieved. Finally, it is shown that at frequencies such as 4 and 6 GHz, where it is possible that differential attenuation may be neglected, a suboptimal adaptive system employing two control parameters leads to a simple and attractive adaptive network.

Depolarization effects in a communications satellite link

Figure 2 shows a general pair of elliptically polarized waves traveling in the positive z direction and having a time variation $e^{j\omega t}$ [14]. The axial ratio, r, is defined as the ratio of the magnitude of the major axis amplitude to that of the minor axis. The sense of rotation of an ellipse is specified by the sign of the axial ratio, *i.e.*, "$+$" for right-hand (clockwise) sense rotation and "$-$" for left-hand (counterclockwise) sense rotation, where the rotation sense is observed in a fixed plane with the wave receding [15]. The tilt angle of each ellipse (α and β) is the angle between the major axis and the x axis. The polarization state of a wave, which is uniquely specified by the tilt angle and axial ratio (including sign), can be expressed as a complex
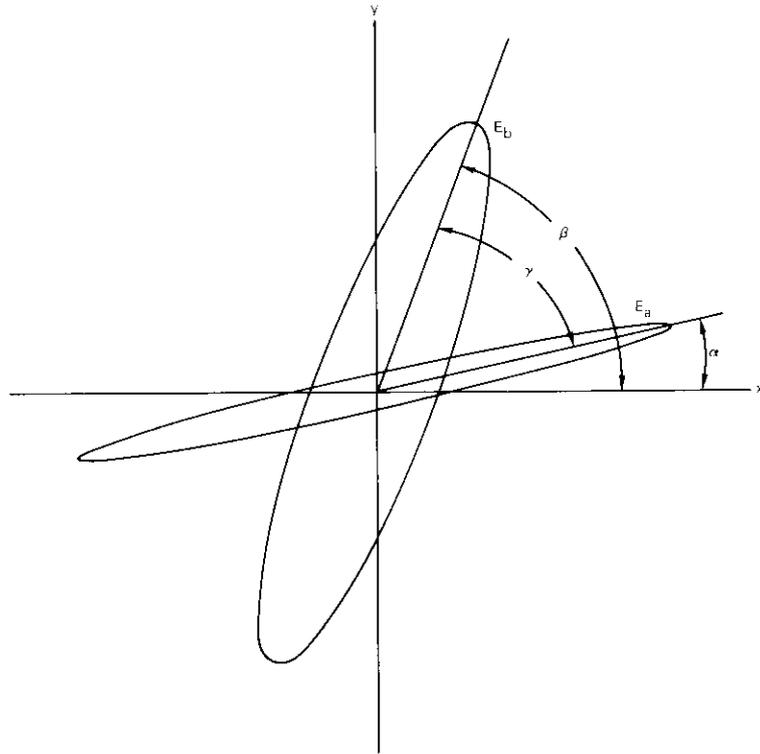
Figure 2. *General Dual-Polarized Field*

vector. Any two such vectors, $\vec{U}$ and $\vec{V}$, are orthogonal if $\vec{U} \cdot \vec{V}^* = 0$, where V^* is the complex conjugate of V. The power associated with a wave is the sum of the power levels in its orthogonal components. The orthogonal vector sets of interest are generally the linear and circular vector cases, although other sets are often used [16].

Transformation from one set of orthogonal vectors to another may be accomplished through the introduction of a lossless phase shift at a given angle in the x-y plane. Phase shift does not affect the orthogonality of two waves, but relates the orthogonality to a different set of reference vectors. However, differential attenuation will change the relative orthogonality of the polarization states. If differential attenuation is applied to a set of perpendicular linear vectors, the result will generally be two linear vectors which are no longer perpendicular.

The device most commonly used in communications satellite systems to couple spatially orthogonal field components from a waveguide into two separate transmission lines is an orthomode transducer (OMT). The OMT ports, which are mutually perpendicular, may be thought of as x and y ports. This method of creating and separating dual-polarized fields will be assumed to be the one with which an adaptive orthogonalization circuit should be compatible.

The depolarizing effects of rainfall have been investigated both analytically and experimentally. Figure 3 shows cross-polarization isolation at 4, 6, and 11 GHz *vs* rain rate over a 5-km path length. These data, adapted from theoretical computations of Oguchi [6], [7], show good agreement with measured data obtained by Taur [9], [10]. The incident waves are assumed to be perfectly circularly polarized (or equivalently, linearly polarized at a canting angle of 45° relative to the raindrops). This condition represents the worst-case degradation of the polarization state due to rain [17]. It is important to note that the effect of differential phase shift is the major source of isolation degradation at these frequencies.

Polarization orthogonality restoration techniques

As described by T. S. Chu [18], the circuit shown in Figure 4 employs a variable rotatable differential phase shifter to linearize the polarization states of two non-orthogonal dual-polarized waves. A variable rotatable differential attenuator then causes the two linear states to be perpendicular. In this system, the reference vectors for orthogonality determination are the linear vectors defined by the two ports of an OMT at ±45° relative to the angle of introduction of the differential attenuation. For the arbitrary dual elliptically polarized field shown in Figure 2, the requirement for simultaneous linearization of both polarization states is the introduction of a differential phase shift, $\Delta\phi$:

$$\Delta\phi = \arctan\left[\frac{2r_b}{(r_b^2 - 1)\sin\left\{\arctan\left[\dfrac{\sin 2\gamma}{\cos 2\gamma - \dfrac{r_a\,(r_b^2 - 1)}{r_b\,(r_a^2 - 1)}}\right]\right\}} \right] \quad (1)$$

which is applied at an angle θ relative to the x axis, where

$$\theta = \left\{ \beta - \frac{1}{2} \arctan \left[\frac{\sin 2\gamma}{\cos 2\gamma - \dfrac{r_a\,(r_b^2 - 1)}{r_b\,(r_a^2 - 1)}} \right] \right\} . \tag{2}$$
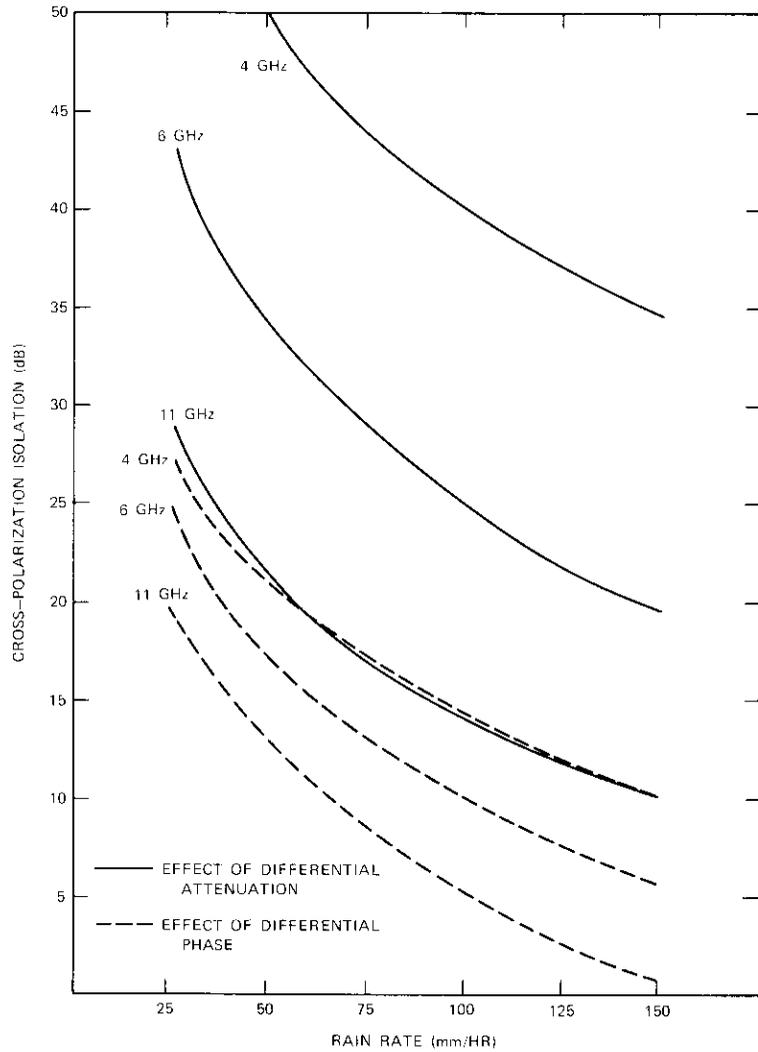

Figure 3. *Phase and Attenuation Components of Theoretical Rain Depolarization Effect*

This will produce two linear polarization vectors oriented at ξ_1 and ξ_2 relative to the x axis, where

$$\xi_1 = \theta \pm \arctan \left[\sqrt{\frac{r_a^2 \tan^2 (\theta - \alpha) + 1}{r_a^2 + \tan^2 (\theta - \alpha)}} \right] \tag{3}$$

$$\xi_2 = \theta \pm \arctan \left[\sqrt{\frac{r_b^2 \tan^2 (\theta - \beta) + 1}{r_b^2 + \tan^2 (\theta - \beta)}} \right] . \tag{4}$$

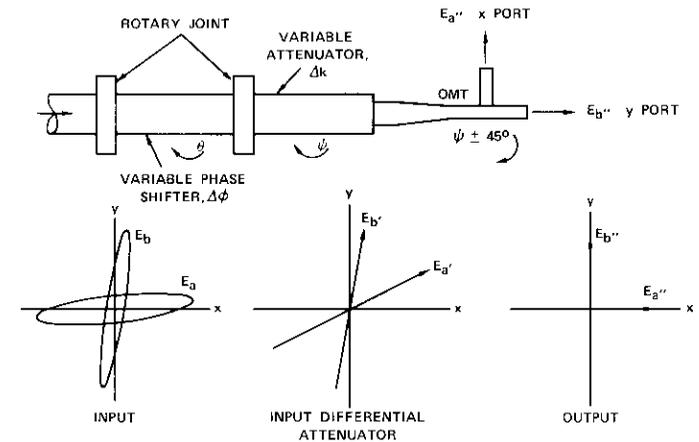The positive solution is chosen if the axial ratio is negative and vice versa.

Figure 4. *Orthogonalization Circuit Employing a Rotatable Differential Phase Shifter and a Rotatable Differential Attenuator*

If $\Delta\xi$, the angle between the two resulting linear vectors, is less than 90°, a differential attenuation of $\Delta k = \tan (\Delta\xi/2)$ applied along the line which bisects the vectors (an angle ψ relative to the x axis) will produce two perpendicular linear polarization vectors. If $\Delta\xi$ is greater than 90°, a differential attenuation of $\cot (\Delta\xi/2)$ applied along the angle ($\psi \pm 90°$) will produce two perpendicular linearly polarized fields. In both cases, the final field vectors are oriented at $\pm 45°$ relative to the line which bisects the nonorthogonal linearly polarized fields. Thus, if an OMT is aligned at 45° relative to the angle of introduction of attenuation, infinite isolation is achieved between the two signals with respect to these OMT ports.

As is evident from Figure 4 and the preceding discussion, four variables

are involved in matching the orthogonality of two arbitrary dual-polarized signals to a set of OMT ports. In this method, these four variables appear as $\Delta\phi$, θ, Δk, and ψ. Two of these variables are associated with phase shift and two with attenuation. This process is one way of correcting for rain depolarization which is caused by differential phase shift and differential attenuation. The rain model, however, predicts that the differential phase and attenuation are introduced into the original field at the same angle. Thus, the rain model has only three degrees of freedom, while the method described herein requires four to achieve orthogonalization. The essential difference is twofold. First, these two general ellipses are not necessarily being returned to the polarization states which they assumed before encountering the rain. Second, the amount of attenuation introduced to correct for the depolarization is generally different from that which caused it. It can be shown that the value of Δk in this correction circuit is the minimum possible to effect orthogonalization [19] and is generally less than the value of attenuation introduced by the rain. Thus, this technique minimizes both Δk and the signal-to-noise degradation at the expense of an additional degree of freedom.

Because the components of the orthogonalization circuit of Figure 4 are realized in waveguide form, it is extremely difficult to implement gain before the circuit. Therefore, the only factors limiting the degree of correction obtained by the circuit are the tolerances on the component values and the angles of introduction of phase shift and attenuation. If it is assumed that the components can be positioned perfectly, the degree of orthogonality of the output polarization states is limited only by the phase error of the polarizer and the amplitude error of the attenuator. While virtually perfect operation may be achieved at a single frequency, broadband operation is limited by the dispersive nature of the waveguide components. From a perturbation analysis, the errors involved will be, to a first order, the same as the errors resulting from the introduction of a small phase shift or attenuation into a set of orthogonal linear vectors. While this depends upon the angles of introduction, the attenuation is always introduced at a 45° angle, the angle at which the phase has its worst effect. Therefore, a worst-case error may be obtained by considering the case of a phase and amplitude error introduced at 45° relative to a set of perpendicular linearly polarized vectors. Figure 5 shows isolation contours as functions of phase and amplitude errors.

An alternate approach to the problem of orthogonalization is the technique of cross coupling [20]. In this implementation a dual-polarized field is separated into its components aligned along two OMT ports. Energy is
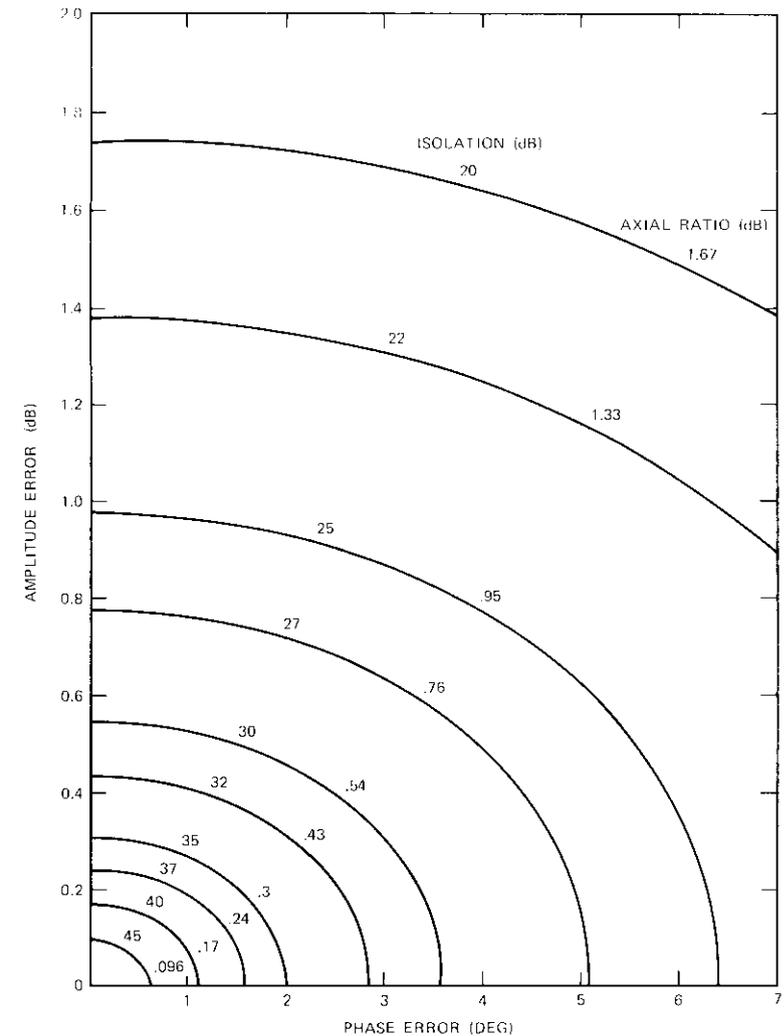
Figure 5. *Polarization Isolation vs Phase and Amplitude Errors*

then coupled between the two signal lines with the proper phase and amplitude to cancel out the undesired signal component in each path. It is assumed, of course, that the undesired component in one path is completely correlated with the desired signal in the other path.
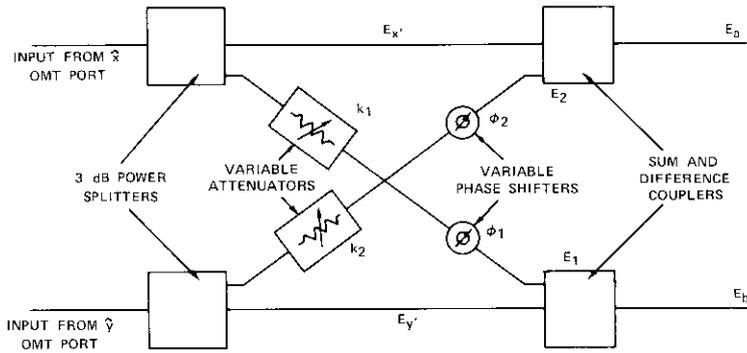
Figure 6. *Cross-coupling Circuit with Decoupled Control Variables*

Figure 6 shows a cross-coupling circuit with fixed 3-dB power dividers and sum and difference couplers in the through paths and a variable attenuator and phase shifter in each of the cross lines. The boxes shown in Figure 6 are variable attenuators for which the voltage coupling through the attenuator is $\sqrt{1 - K^2}$. The $\Delta\phi$ components are variable phase shifters which multiply the signal by $e^{-j\phi}$. Thus, for the dual-polarized field in Figure 1, it can be shown [13] that the output voltages of this circuit are

$$
E_x = \frac{E_a}{2} \left\{ \sqrt{1 - K_2^2}\, e^{-j\phi_2} \left[\sin\alpha + \frac{j}{r_a}\cos\alpha \right] \right.
$$

$$
\left. - \left[\cos\alpha - \frac{j}{r_a}\sin\alpha \right] \right\} + \frac{E_b}{2} \left\{ \sqrt{1 - K_2^2} \right.
$$

$$
\cdot e^{-j\phi_2} \left[\sin\beta + \frac{j}{r_b}\cos\beta \right] - \left[\cos\beta - \frac{j}{r_b}\sin\beta \right] \right\} \tag{5}
$$

$$
E_y = \frac{E_b}{2} \left\{ \sqrt{1 - K_1^2}\, e^{-j\phi_1} \left[\cos\beta - \frac{j}{r_b}\sin\beta \right] \right.
$$

$$
\left. - \left[\sin\beta + \frac{j}{r_b}\cos\beta \right] \right\} + \frac{E_a}{2} \left\{ \sqrt{1 - K_1^2} \right.
$$

$$
\cdot e^{-j\phi_2} \left[\cos\alpha - \frac{j}{r_a}\sin\alpha \right] - \left[\sin\alpha + \frac{j}{r_a}\cos\alpha \right] \right\} \tag{6}
$$

The desired condition is that the E_b component of E_x and the E_a component

of E_y vanish. This yields the following conditions on K_1, K_2, ϕ_1, and ϕ_2:

$$
\sqrt{1 - K_1^2} = \sqrt{\frac{r_a^2 \tan^2\alpha + 1}{r_a^2 + \tan^2\alpha}} \tag{7}
$$

$$
\phi_1 = \arctan\left[\frac{(1/\tan\alpha) - \tan\alpha}{r_a + (1/r_a)} \right] \tag{8}
$$

$$
\sqrt{1 - K_2^2} = \sqrt{\frac{r_b^2 + \tan^2\beta}{r_b^2 \tan^2\beta + 1}} \tag{9}
$$

$$
\phi_2 = \arctan\left[\frac{\tan\beta + (1/\tan\beta)}{r_b - (1/r_b)} \right] . \tag{10}
$$

The cross-coupler circuit shown in Figure 6, which has fixed power splitters and 3-dB hybrids, does not yield the minimum possible noise figure since it is a simplification of a general cross-coupling circuit [13] with four variable couplers. However, if linear amplification is introduced before the circuit, the effect of cross coupling on the circuit noise figure is reduced. Figure 7 indicates the noise temperature degradation of these networks. The general cross-coupler network, having equal values of coupling, yields degradation identical to that of the Chu network. The curve representing the inherent noise temperature degradation is fundamentally related to the departure from orthogonality (α) of two linearly polarized waves. It represents a lower bound on the degradation that would be obtained if high-gain amplifiers preceded the cross-coupler networks. The formulas for computing the noise temperature degradation are presented in Appendix A.

The cancellation ability of the circuit in Figure 6 is limited by the accuracy of the variable phase shifters and attenuators. If they can be driven to perfect positioning at the pilot signal frequencies, the bandwidth of the circuit is then limited by the dispersive nature of the components and the rain depolarization effect itself. This means that, if gain is introduced before the circuit, linear amplifiers with matched amplitude and phase characteristics must be used in both lines. Figure 8 is a plot of achievable isolation improvement vs differential phase and amplitude errors in the cross-coupling circuit and/or in any preceding circuit components such as amplifiers.
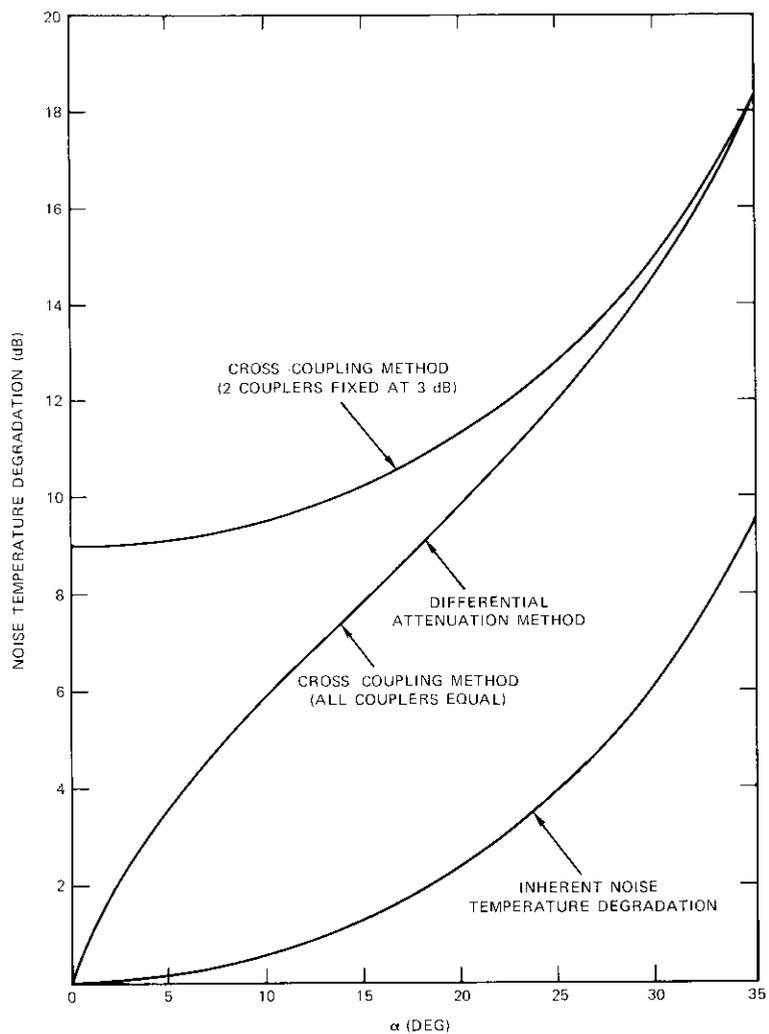
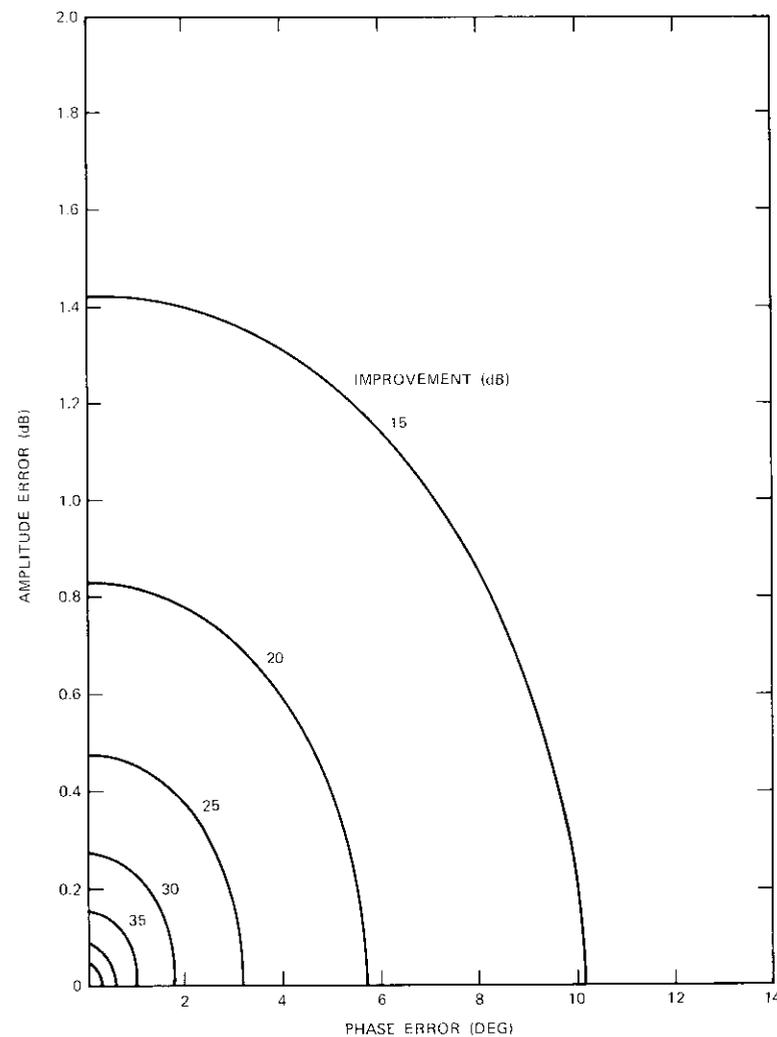Figure 7. *Noise Temperature Degradation vs Degree of Non-orthogonality of Incident Polarization States*

Figure 8. *Improvement in Signal Isolation vs Cross-coupled Signal Phase and Amplitude Error*
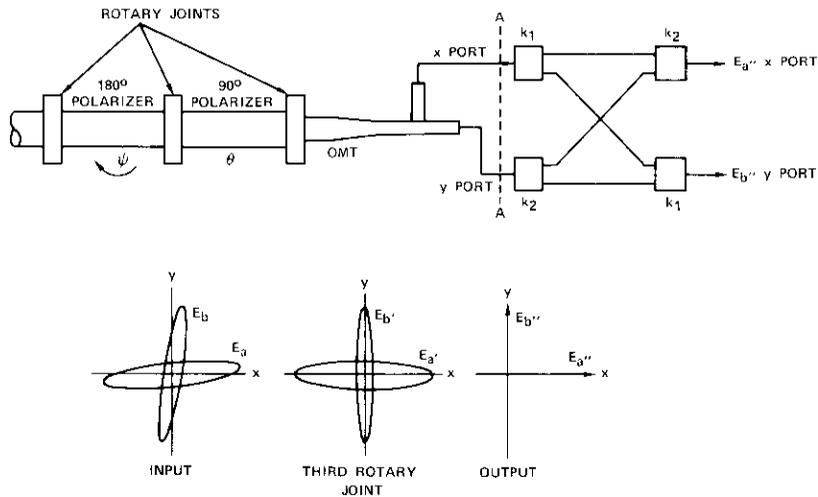
Figure 9. *Orthogonalization Circuit Employing Rotatable Fixed Phase Shifter and Quadrature Cross Coupling*

Polarization orthogonalization can also be achieved by using the circuit shown in Figure 9. The purpose of the variable rotatable phase shifter in the circuit of Figure 4 is to transform the polarization states of the two arbitrary elliptically polarized input waves to linear polarization states. With a fixed polarizer, however, an equally useful transformation is the generation of two elliptically polarized waves, with the same sense of rotation, whose major axes are perpendicular. A fixed phase shift of at least 90° can accomplish this result for an arbitrary set of dual-polarized inputs. A complete analysis of the equations governing this procedure is given in Reference 13. For the general input shown in Figure 2, a rotatable 90° polarizer ($\lambda/4$ plate) should have its phase shift plane at an angle θ relative to the x axis, where

$$\theta = \alpha - \frac{1}{2}\arctan\left[\frac{\cos 2\gamma - \{(r_b/r_a)\,[(1 - r_a^2)/(1 - r_b^2)]\}}{\sin 2\gamma}\right] \quad . \quad (11)$$

This 90° phase shift can also be applied at $\theta \pm 90°$ and $\theta \pm 180°$. The difference in these positions is merely the direction of rotation of both of the resulting polarization ellipses.

The resulting perpendicular ellipses will not generally have their major axes aligned with the OMT ports (the x-y axes in Figure 9). The value of τ_a', the tilt angle of the resulting ellipse E_a' relative to the x axis, is

$$\tau_a' = (\theta - 90°) + \frac{1}{2}\arctan\left[\frac{2r_a}{(r_a^2 - 1)\cos 2(\alpha - \theta)}\right] \quad . \quad (12)$$

The OMT can be rotated by τ_a' or, alternatively, a 180° polarizer ($\lambda/2$ plate) can be rotated by $\tau_a'/2$ to align the major axes of the resulting perpendicular ellipses with the OMT ports. The component of E_a' in the x port will differ from the component of E_a' in the y port by a fixed phase difference of 90°. If the cross-coupling circuit has this fixed 90° phase difference incorporated in the cross paths, only two variable values, K_1 and K_2, are needed to complete the orthogonalization. These values, which are a simple case of the general cross-coupling circuit solutions, are given by

$$K_1 = \sqrt{\frac{1 + (r_a/r_b) \pm \sqrt{[(r_a/r_b) - 1]^2 + 4r_a^2}}{2r_a[(1/r_b) - r_a]}} \quad (13)$$

$$K_2 = K_1\sqrt{\frac{r_a}{r_b}} \quad . \quad (14)$$

This circuit has the advantage of simple implementation in waveguide as well as a small signal-to-noise degradation. In addition, the dissipative variable couplers can be incorporated after amplification to minimize their effect on system noise temperature.

Control techniques for orthogonalization circuits

To determine the suitability of the circuit shown in Figure 4 for closed-loop adaptive control, complex output voltages E_x and E_y can be expressed in terms of complex input voltages E_a and E_b; the variable differential phase shift, ϕ; its orientation, θ; the variable differential attenuation, K; and its orientation, ψ. The equations for E_x and E_y are included in Appendix B. It is important to note that these equations are nonlinear and transcendental for some variables. In addition, there are intrinsic interrelationships between the variables which prevent any change in a measurable quantity of E_x and E_y from being attributed to any single control variable.

A practical control technique is measurement of the phase and amplitude of the E_a and E_b components of the signals in the x and y OMT ports. This information can be used to calculate the axial ratios and tilt angles of the polarization ellipses at the OMT. The effects of the variable rotatable attenuator and phase shifter can be applied in reverse to this field, and the polarization states of the waves at the input to the circuit of Figure 4 can be computed. It is then a simple matter to use the equations given previously to position the variables at the solution values calculated to achieve orthogonalization. This approach requires accurate measurement of the variable positions and a computational device. As microprocessors become less expensive and more sophisticated, this approach will become more attractive. A learning algorithm may even be introduced to allow the processor to correct for inaccurate readings of the control variable positions.

For the cross-coupling technique shown in Figure 6, adaptive polarization control becomes a matter of classic linear feedback. Consider the phase and amplitude of the E_b component of E_2 and the E_b component of $E_x{'}$:

$$\angle E_{b2} = \arctan\left[\frac{1}{r_b \tan \beta}\right] - \phi_2 \tag{15}$$

$$\angle E_{bx}{'} = \arctan\left[\frac{-\tan \beta}{r_b}\right] \tag{16}$$

$$|E_{b2}| = \sqrt{\frac{1 - K_2^2}{2}} \sqrt{\sin^2 \beta + \frac{\cos^2 \beta}{r_b^2}} \tag{17}$$

$$|E_{bx}{'}| = \frac{1}{\sqrt{2}} \sqrt{\cos^2 \beta + \frac{\sin^2 \beta}{r_b^2}} \ . \tag{18}$$

If the phase and amplitude of E_{b2} must equal the phase and amplitude of $E_{bx}{'}$:

$$\phi_2 = \arctan\left[\frac{1}{r_b \tan \beta}\right] - \arctan\left[\frac{-\tan \beta}{r_b}\right] \tag{19}$$

$$\sqrt{1 - K_2^2} = \sqrt{\frac{\cos^2 \beta + (\sin^2 \beta)/r_b^2}{\sin^2 \beta + (\cos^2 \beta)/r_b^2}} \tag{20}$$

which can be shown to be the same as the requirements expressed by equations (9) and (10). Similarly, requiring the magnitude and angle of the E_a components of E_1 and $E_{y'}$ to be equal yields specifications which are the same as equations (7) and (8). This then suggests the control system shown in Figure 10 as a means of achieving orthogonalization using a closed-loop system. Some of the energy in the E_1 and $E_{y'}$ lines is coupled off and fed into a phase amplitude receiver. This receiver produces voltages which are proportional to the phase and amplitude differences between the two input signals. These voltages are then fed back to control the coupled line attenuator and phase shifter. The result is a system which equalizes the phase and amplitudes of the pilot signals in the E_1 and $E_{y'}$ paths. If appropriate offsets are incorporated in either the control loop or the lines following the signal couplers, it is possible to compensate for any static phase or amplitude difference in the couplers or in the signal path lengths of the circuit of Figure 6. This control system is repeated for the x channel.

For the orthogonalization circuit shown in Figure 9, the relationships between the output voltages and the control parameters θ and ψ are given in Appendix C. It has been shown in equation (11) that there is a particular
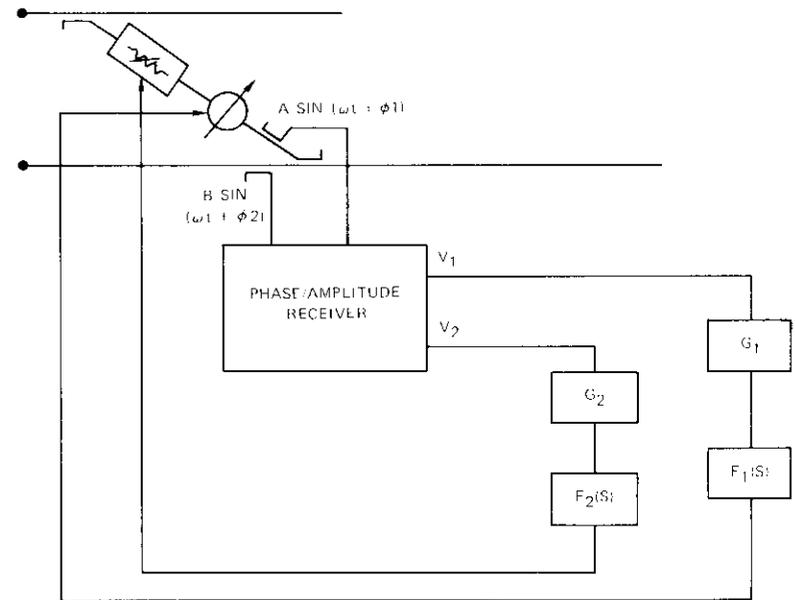

Figure 10. *Control System Block Diagram for One Side of a Cross-coupling Circuit*

solution for θ and ψ which will convert the general input field of Figure 1 to one with spatially perpendicular elliptically polarized signals rotating in the same sense. In terms of the voltages E_x and E_y, this means that the E_a component of E_x leads (or lags) the E_a component of E_y by $90°$. Therefore, the important phase differences are

$$\angle E_a = \angle E_{ax} - \angle E_{ay} \tag{21}$$

$$\angle E_b = \angle E_{bx} - \angle E_{by} \tag{22}$$

which can be expressed as [13]

$$\angle E_a = \arctan \left\{ \frac{\tan(\theta + \alpha - 2\psi)(1 - r_a^2)\sec^2\theta}{(r_a \tan\theta - 1)(r_a + \tan\theta)\tan^2(\theta + \alpha - 2\psi) + (\tan\theta - r_a)(r_a \tan\theta + 1)} \right\} \tag{23}$$

$$\angle E_b = \arctan \left\{ \frac{\tan(\theta + \beta - 2\psi)(1 - r_b^2)\sec^2\theta}{(r_b \tan\theta - 1)(r_b + \tan\theta)\tan^2(\theta + \beta - 2\psi) + (\tan\theta - r_b)(r_b \tan\theta + 1)} \right\} \tag{24}$$

While amplitude coupling may be controlled through a system similar to the amplitude control loop of the cross-coupling circuit described previously, phase control for this circuit is more complicated. To decouple the phase parameters from the amplitude coupling parameters, it seems best to measure the phase in the coupling arms rather than after the cancellation signal has been mixed with the interfering signal. However, there does not appear to be any simple way to directly feed back the phase information to drive the 90° and 180° polarizers without first processing it.

The desired steady-state condition for operation is $\angle E_a = \angle E_b = 90°$, which also means that $\angle E_a - \angle E_b = 0$. Unfortunately, due to the interdependence of θ and ψ, neither of these can be considered to be a sufficient error voltage for closed-loop control. However, through computer simulation it appears that a technique of sequential adjustment of θ and ψ to simultaneously minimize the cross polarization in both channels will usually lead to a solution. Unfortunately, there are local minima in the control surface and the minimum seeking algorithms must be chosen carefully so that the circuit finds the true global minimum and not a local minimum.

This circuit can also be adaptively controlled by using a minicomputer. A processor can be constructed to measure the phase and amplitude information and compute the axial ratios and tilt angles of the polarization ellipses after the 90° and 180° polarizers. The reciprocity theorem can then be used to find, by calculation, the initial incident polarization states at the input to the correction circuit. Finally, the input polarization ellipse information can be applied to the equations given previously to calculate the desired positions of θ and ψ.

A suboptimal adaptive polarization restoration system

It has been noted that four parameters are generally required to provide complete restoration of a dual-polarized frequency reuse system. If rain depolarization is assumed to be the only disturbance, then three parameters suffice since the canting angles of both the phase shift and the differential attenuation are identical. For a 4/6-GHz satellite communications link, a further simplification may be made by neglecting the effect of differential attenuation. From a system point of view, this can be justified by referring to Figure 3. For a 5-km average rain path length, 30-dB isolation is achieved at rain rates of 62 mm/hr* with differential phase correction only. This results in a suboptimal adaptive polarization restoration system which requires only two variable parameters and which can be most easily implemented by a cascaded set of independently rotatable 90° and 180° polarizers (see Figure 9 [to section A-A]). One method of providing adaptive control with this circuit is shown in Figure 11.

Since the error voltage magnitudes and/or phases are not directly related to the angular positions of the two polarizers, simple direct error feedback is not possible. Therefore, some form of error signal processing, together with an algorithm to direct the rotation of the polarizers, is required to minimize these error voltages.

In the particular adaptive control network shown in Figure 11, the output signals from the x port of the OMT will contain the interference from channel 2 and those from the y port of the OMT will contain the interference from channel 1. The interchannel interference is therefore minimized if the error signals are simultaneously minimized. An algorithm which

---

* For example, a rain rate of 62 mm/hr is exceeded only 0.01 percent of the time, i.e., 1 hr/yr in the Washington, D.C., area.
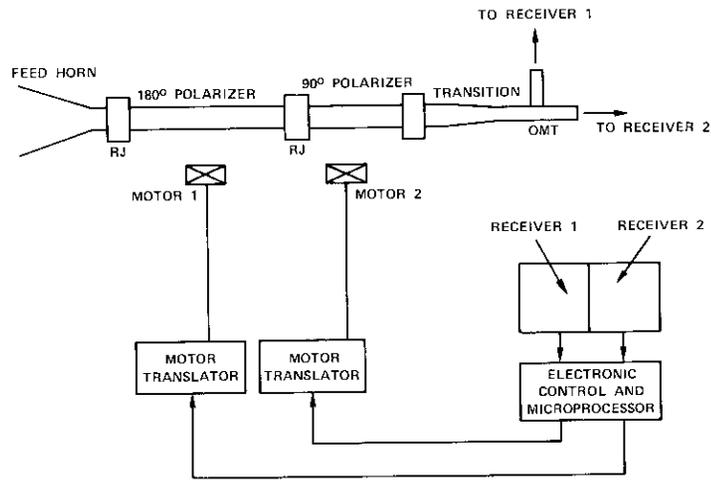
Figure 11. *Adaptive Correction Circuit*

successively rotates the polarizers and minimizes the sum of the error power levels has been modeled on a computer. This procedure is shown in Figure 12 for dual circular polarization and differential phase and atten-
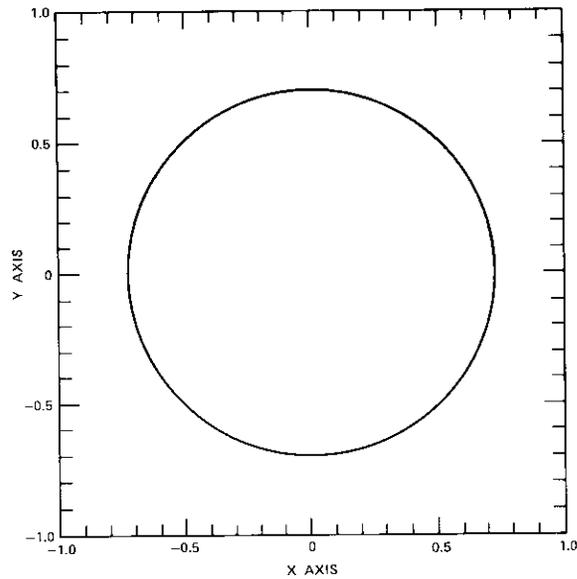

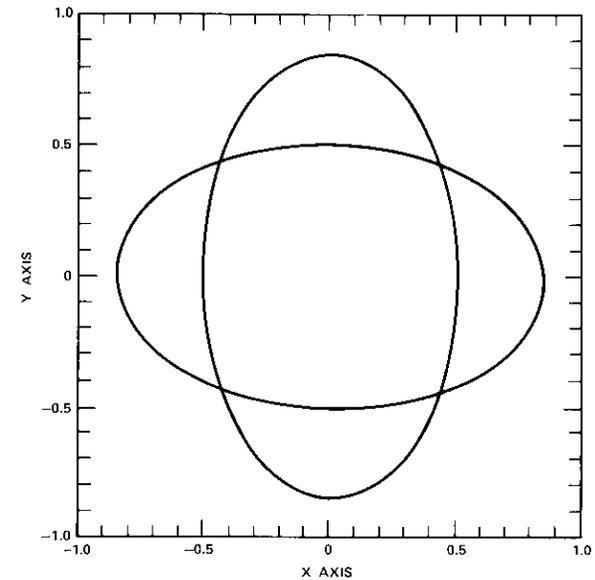Figure 12a. *Circularly Polarized Signals*

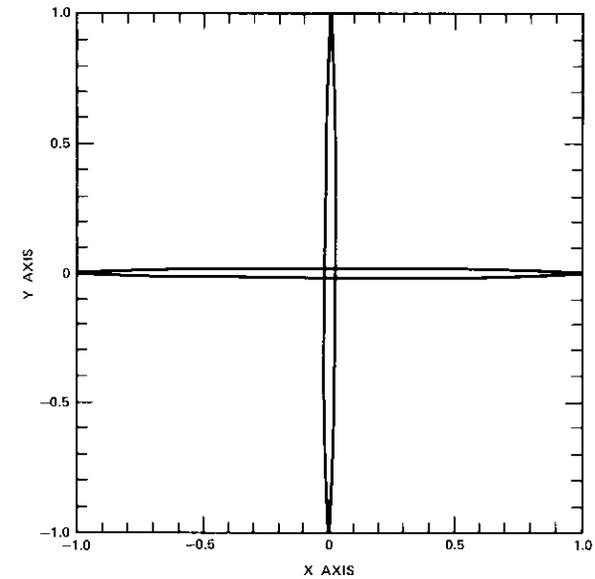Figure 12b. *Circularly Polarized Signals Subjected to a 125-mm/hr Rain Rate at a 45° Canting Angle*

Figure 12c. *Resultant Signals Following Polarizer Correction Network*

uation of 28.5° and 0.245 dB, respectively. (This corresponds to a theo-
retical rain rate of 125 mm/hr at 4 GHz.) The ellipticity of the final vectors
is directly related to the differential attenuation of 0.245 dB, with a resul-
tant isolation of 37 dB (see Figure 3). Error contours for the adaptive
network are shown in Figure 13.

ANGLE OF 180° PLATE

A= −45   E= −25
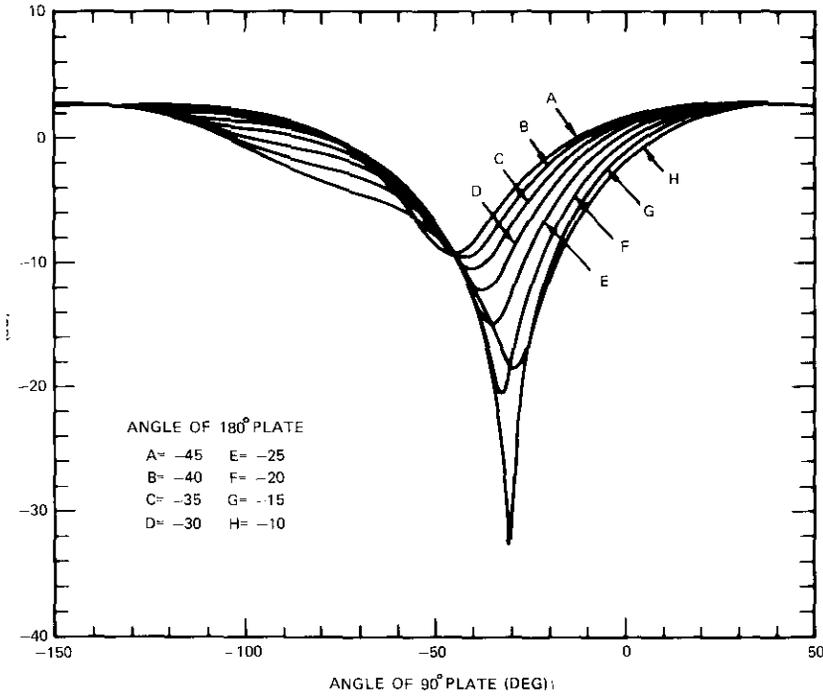B= −40   F= −20
C= −35   G= −15
D= −30   H= −10

Figure 13. *Error Contours Following Polarizer Correction Network*

It is interesting to note that the physical arrangement of the 90° and 180°
polarizers affects the speed of response of the adaptive circuit. In Figure 12,
the 90° polarizer is adjacent to the OMT. These properties are indicated in
Figures 14a and 14b, which show the domain of solutions for the polarizer
angles as a function of the differential phase shift and canting angle of the
rain. Because the range of polarizer angle solutions for the network of
Figure 12 is smaller for input waves that are nearly circularly polarized,
the speed of response for this network can be superior.

The polarization control network was tested in the laboratory using the
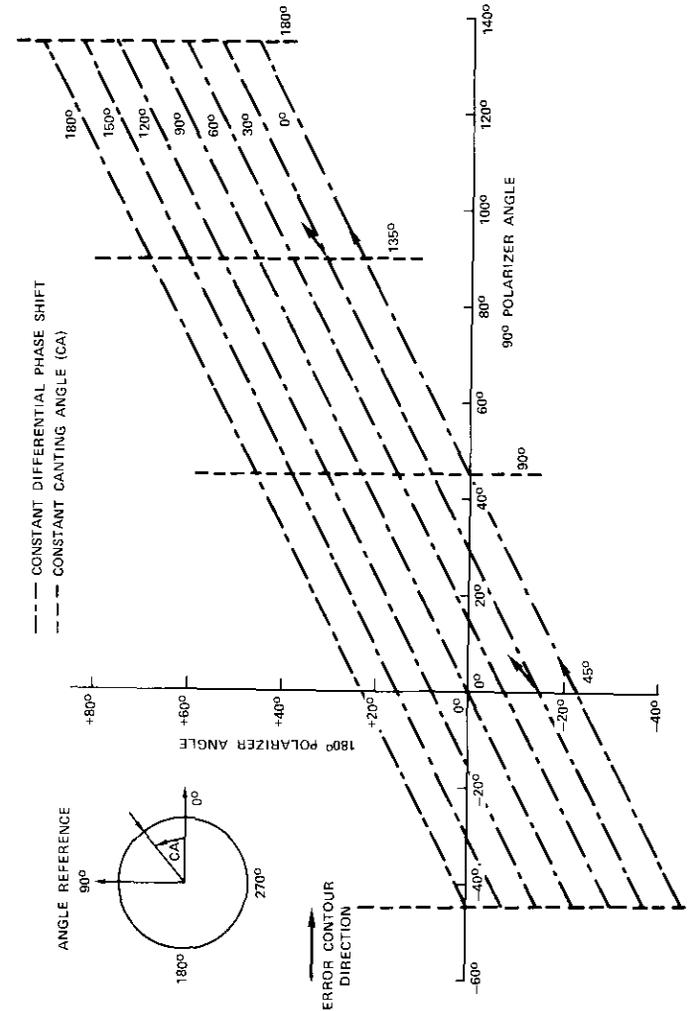4-GHz 90° polarizer [21], rotary joints, and orthomode junction shown in

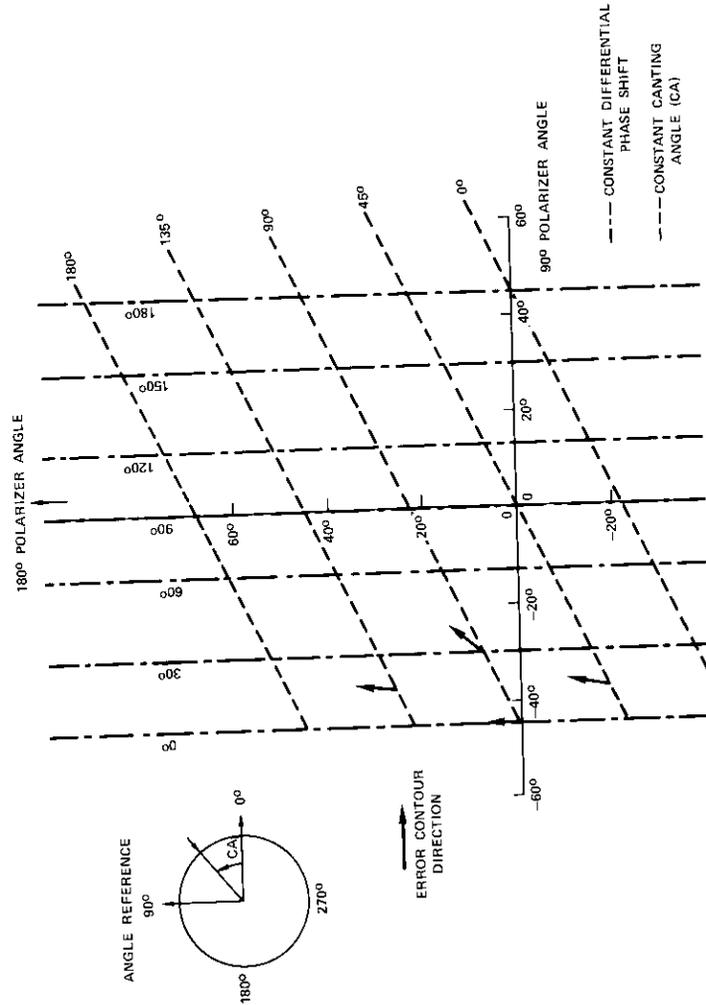Figure 14a. *Solution Domains for Suboptimal Adaptive Circuit*

Figure 14b. *Solution Domains for Suboptimal Adaptive Circuit*

Figure 15. Since the 90° polarizer had an axial ratio better than 0.15 dB and the OMT port-port isolation was greater than 50 dB over a 500-MHz bandwidth, polarization discrimination better than 37 dB was achieved at the pilot signal frequency.
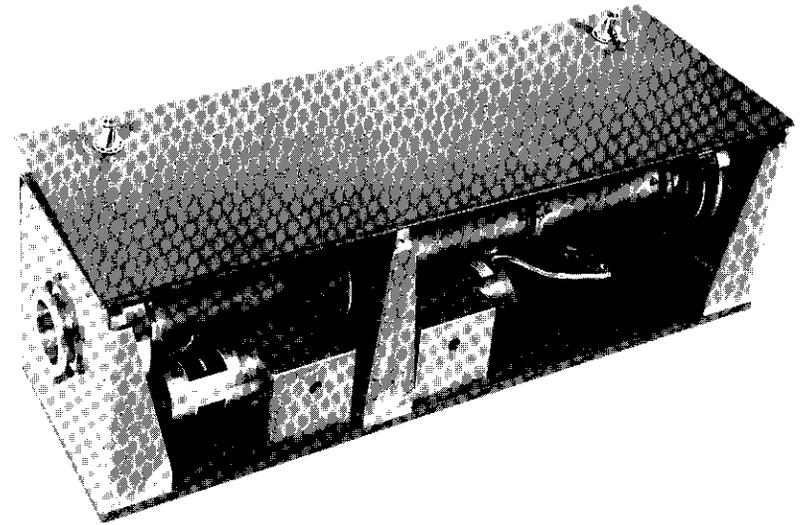
Figure 15. *4-GHz Waveguide Adaptive Circuit*

Conclusions and Discussion

There are several network concepts which can be used to restore the orthogonality of dual-polarized fields on an adaptive basis, i.e., without *a priori* knowledge of the depolarizing properties of the medium. Incorporating pilot signals and suitable receivers for these signals makes it possible to derive error voltages to control the adaptive circuits. Signal-to-interference ratios of at least 30 dB are practical with these techniques.

For the cross-coupling circuit, voltage-controlled attenuators and phase shifters allow rapid response times (*e.g.*, circuit bandwidths $\geq$ 100 Hz). However, to realize minimum system noise figure degradation, these circuits should follow a "matched" pair of low-noise amplifiers whose phase and amplitude tracking must be tightly controlled.

The waveguide correction circuits can be connected to the feed system

of an existing earth station antenna. Minimum degradation in the carrier-to-noise ratio is achieved since the differential phase shift is directly removed by the first rotating polarizer. However, the large inertia of the waveguide components indicates a small control bandwidth or a slow circuit response time of the order of 0.5 second.

A particularly attractive correction circuit for use in the 4/6-GHz satellite bands is the suboptimal realization described in the preceding section. Here, only differential phase shift is removed and only two control parameters must be monitored.

Acknowledgments

References

[1] D. F. DiFonzo and R. W. Kreutel, "Communications Satellite Antennas for Frequency Reuse," *G-AP International Symposium*, September 1971, pp. 287–289.

[2] Nippon Electric Company, Ltd., Tokyo, Japan, "Final Report for Wideband Mode Transducer/Polarizer Assembly," prepared for INTELSAT under contract CSC-IS-473.

[3] W. Sones and M. Grossman, "Polarization Isolation and Spurious Signal Rejection Requirements for a Dual-Polarized Unattended Earth Terminal," *COMSAT Technical Review*, Vol. 5, No. 2, Fall 1975, pp. 367–386.

[4] J. A. Morrison and T. S. Chu, "Perturbation Calculations of Rain-Induced Differential Attenuation and Differential Phase Shift at Microwave Frequencies," *Bell System Technical Journal*, Vol. 52, No. 10, December 1973, pp. 1907–1913.

[5] J. A. Morrison, M. J. Cross, and T. S. Chu, "Rain-Induced Differential Attenuation and Differential Phase Shift at Microwave Frequencies," *Bell System Technical Journal*, Vol. 52, No. 4, April 1973, pp. 599–604.

[6] T. Oguchi, "Attenuation of Electromagnetic Waves Due to Rain with Distorted Raindrops, Part I," *Journal of the Radio Research Labs* (Japan), No. 33, September 1960, pp. 467–485.

[7] T. Oguchi, "Attenuation of Electromagnetic Waves Due to Rain with Distorted Raindrops, Part II," *Journal of the Radio Research Labs* (Japan), No. 53, January 1965, pp. 19–44.

[8] T. Oguchi, "Attenuation and Phase Rotation of Radio Waves Due to Rain: Calculation at 19.3 and 34.8 GHz," *Radio Science*, Vol. 8, No. 1, January 1973, pp. 31–38.

[9] R. Taur, "Rain Depolarization: Theory and Experiment," *COMSAT Technical Review*, Vol. 4, No. 1, Spring 1974, pp. 187–190.

[10] R. Taur, "Rain Depolarization at 4, 6 and 11 GHz," 1974 URSI Meeting, Atlanta, Georgia, June 1974.

[11] M. Saunders, "Cross-Polarization at 18 and 30 GHz Due to Rain," *IEEE Transactions on Antennas and Propagation*, AP-19, No. 2, March 1971, pp. 273–277.

[12] R. A. Semplak, "Simultaneous Measurements of Depolarization by Rain Using Linear and Circular Polarization at 18 GHz," *Bell System Technical Journal*, Vol. 53, No. 2, February 1974, pp. 400–404.

[13] W. S. Trachtman, "Adaptive Orthogonalization of Polarized Signals with Applications to Satellite Communications," M.S. Thesis, M.I.T., June 1975.

[14] V. H. Rumsey et al., "Techniques for Handling Elliptically Polarized Waves with Special Reference to Antennas," *Proc. IRE*, May 1951, pp. 553–556.

[15] International Radio Consultative Committee (C.C.I.R.), *XIIth Plenary Assembly, New Delhi, 1970*, Vol. III, Report 321, Geneva: International Telecommunications Union, 1970.

[16] J. S. Hollis et al., "Microwave Antenna Measurements," Scientific Atlanta, Inc., July 1970.

[17] P. A. Watson and M. Arabi, "Rainfall Cross-Polarization of Linearly and Circularly Polarized Waves at Microwave Frequencies," *Electronic Letters*, Vol. 8, No. 11, June 1972.

[18] T. S. Chu, "Restoring the Orthogonality of Two Polarizations in Radio Communication Systems, I," *Bell System Technical Journal*, Vol. 50, No. 9, November 1971, pp. 3063–3069.

[19] T. S. Chu, "Restoring the Orthogonality of Two Polarizations in Radio Communication Systems, II," *Bell System Technical Journal*, Vol. 52, No. 3, March 1973, pp. 319–327.

[20] D. F. DiFonzo, A. E. Williams, and W. S. Trachtman, "Adaptive Polarization Control for Satellite Frequency Reuse Systems," *IEEE/AP-S Symposium*, June 1975, pp. 209–216.

[21] R. W. Gruner and W. J. English, "A C.P. Reference Standard Antenna System for Satellite or Earth Station Applications," International IEEE-AP-S Symposium, June 1975.

Appendix A. Noise temperature considerations

For the Chu circuit of Figure 4, assume that the proper phase shift has been applied so that the input to the differential attenuator is a field with two non-orthogonal linearly polarized signals spatially separated by an angle ψ and oriented from the x and y ports by an angle α, where $\alpha = (90 - \psi)/2$. The angle α is a measure of the non-orthogonality of the signal states. It is now of interest

to find the amount of noise temperature degradation caused by the operation of the orthogonalization circuit as a function of α. Let the noise temperature of the antenna and propagation path be T_a. If it is assumed that the noise power incident on the attenuator circuit has no preferred spatial orientation, the noise voltage seen at the x and y OMT ports is given by

$$V_{nT_a} = \sqrt{\frac{1 + K^2}{2}} T_a \qquad (A\text{-}1)$$

where K is the value of the attenuator (resistive card).

If it is assumed that the resistive card is a thermal noise generator producing a noise signal oriented parallel to the direction of the attenuation, the contribution to the noise voltage of each of the x and y OMT ports due to the resistive card will be

$$V_{nrc} = \sqrt{\frac{1 - K^2}{2}} T_o \qquad (A\text{-}2)$$

where T_o is the effective noise temperature of the resistive card in Kelvin.

If T_a and T_o are from totally uncorrelated noise sources, the noise power, P_n, in either port will be

$$P_n = \left(\frac{1 + K^2}{2}\right) T_a + \left(\frac{1 - K^2}{2}\right) T_o . \qquad (A\text{-}3)$$

On the other hand, if the signal voltage incident on the circuits is E_s, then the signal power, P_s, is given by

$$P_s = E_s^2 \left[K^2 \cos^2 \frac{\psi}{2} + \sin^2 \frac{\psi}{2} \right] . \qquad (A\text{-}4)$$

Therefore, the equivalent noise temperature of this circuit is

$$T_n = \frac{\left(\frac{1 + K^2}{2}\right) T_a + \left(\frac{1 - K^2}{2}\right) T_o}{K^2 \cos^2 \frac{\psi}{2} + \sin^2 \frac{\psi}{2}} \qquad (A\text{-}5)$$

which reduces to

$$T_n = \frac{T_a}{[\cos 2\alpha]^2} + \frac{T_o \sin 2\alpha}{[\cos 2\alpha]^2} . \qquad (A\text{-}6)$$

The noise temperature is clearly degraded by two separate effects. The T_a component, which is the antenna and propagation path noise, is degraded by the reduction in input signal strength. This can be assumed to be an inherent noise temperature degradation attributable to the process of orthogonalization itself even if a noiseless attenuator ($T_o = 0$) is used in the circuit. The T_o term is recognizable as the noise generated by the attenuator itself.

The inherent noise temperature degradation of Figure 7 is a plot of $-20 \log_{10} [\cos 2\alpha]$, while the differential attenuation curve is a plot of the noise temperature degradation of equation (A-6) for $T_a = 40$ K and $T_o = 290$ K. Similarly, it can be shown that the cross-coupler circuit having equal coupling values exhibits a noise temperature degradation identical to that of Chu's circuit. With fixed 3-dB couplers the noise temperature degradation is inherently worse for small α. If a matched set of high-gain parametric amplifiers precedes the circuit, the noise temperature will approach the lower curve.

Appendix B. Control voltage equations for the Chu correction network

Straightforward analysis of the network of Figure 4 leads to the following expressions for the output voltages in terms of the input parameters:

$$\begin{aligned}
E_x = &\frac{1}{\sqrt{2}} E_a \{ e^{-j\phi}[A \cos \theta + B \sin \theta] [C \cos \theta + D \sin \theta] \\
&+ [B \cos \theta - A \sin \theta] [D \cos \theta - C \sin \theta]\} \\
&+ \frac{1}{\sqrt{2}} E_b \{ e^{-j\phi}[A \cos \theta + B \sin \theta] [E \sin \theta + F \cos \theta] \\
&+ [B \cos \theta - A \sin \theta] [E \cos \theta - F \sin \theta]\} \\
E_y = &\frac{1}{\sqrt{2}} E_a \{ e^{-j\phi}[G \cos \theta + H \sin \theta] [C \cos \theta + D \sin \theta] \\
&+ [H \cos \theta - G \sin \theta] [D \cos \theta - C \sin \theta]\} \\
&+ \frac{1}{\sqrt{2}} \{ E_b\, e^{-j\phi}[G \cos \theta + H \sin \theta] [F \cos \theta + E \sin \theta] \\
&+ [H \cos \theta - G \sin \theta] [E \cos \theta - F \sin \theta]\}
\end{aligned}$$

where

$$\begin{aligned}
A &= K \cos \psi + \sin \psi \\
B &= K \sin \psi - \cos \psi \\
C &= \cos \alpha - \frac{j}{r_a} \sin \alpha \\
D &= \sin \alpha + \frac{j}{r_a} \cos \alpha
\end{aligned}$$

$$E = \sin \beta + \frac{j}{r_b} \cos \beta$$

$$F = \cos \beta - \frac{j}{r_b} \sin \beta$$

$$G = K \cos \psi - \sin \psi$$
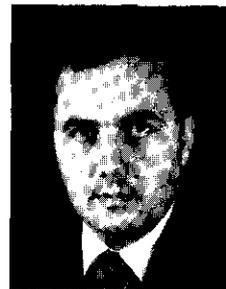
$$H = K \sin \psi + \cos \psi \ .$$

Appendix C. Control voltage equations for the network of Figure 9

Expressing the output voltages for the network of Figure 9 in terms of the input voltages and angular positions of the 180° and 90° polarizers yields

$$
\begin{aligned}
E_x = E_a\{&(A \cos 2\psi + C \sin 2\psi)(j \cos^2 \theta - \sin^2 \theta) \\
&+ (A \sin 2\psi - C \cos 2\psi)(\sin \theta \cos \theta)(1 + j)\} \\
+ E_b\{&(B \cos 2\psi + D \sin 2\psi)(j \cos^2 \theta - \sin^2 \theta) \\
&+ (B \sin 2\psi - D \cos 2\psi)(\sin \theta \cos \theta)(1 + j)\} \\
E_y = E_b\{&(A \sin 2\psi - C \cos 2\psi)(j \sin^2 \theta - \cos^2 \theta) \\
&+ (A \cos 2\psi + C \sin 2\psi)(\sin \theta \cos \theta)(1 + j)\} \\
+ E_b\{&(B \sin 2\psi - D \cos 2\psi)(j \sin^2 \theta - \cos^2 \theta) \\
&+ (B \cos 2\psi + D \sin 2\psi)(\sin \theta \cos \theta)(1 + j)\}
\end{aligned}
$$

where the angle ψ refers to the angular position of the 180° plate, the angle θ refers to the angular position of the 90° plate, and

$$A = \left[\cos \alpha - \frac{j}{r_a} \sin \alpha \right]$$

$$B = \left[\cos \beta - \frac{j}{r_b} \sin \beta \right]$$

$$C = \left[\sin \alpha + \frac{j}{r_a} \cos \alpha \right]$$

$$D = \left[\sin \beta + \frac{j}{r_b} \cos \beta \right] \ .$$

Daniel F. DiFonzo received a B.E.E. from Villanova University in 1962 and an M.S. from California State University in 1972. He is presently Assistant Manager of the Antennas Department of the Microwave Laboratory at COMSAT Laboratories. He is responsible for directing and carrying out research and development related to communications satellite and earth station antenna systems. Specifically, he has performed extensive analytical and experimental studies of techniques for the control of microwave antenna polarization properties and the design of multiple-beam spacecraft antennas. He is a member of IEEE, APS, and G-MTT.

Albert E. Williams was born in Albany, Australia. He received a B.E. in electrical engineering from the University of Western Australia, Perth, in 1962, and a Ph.D. from University College, London, England, in 1966. From 1966 to 1968 he was a lecturer in the Department of Electrical Engineering, University of Western Australia. Currently, he is a technical staff member of the Antennas Department of the Microwave Laboratory at COMSAT Laboratories, where he is working on antenna and filter theory applicable to future communications satellites. Dr. Williams was a joint recipient of the Institute of Electrical Engineers (London) Sylvanus P. Thompson Premium award in 1966.

Warren S. Trachtman received both a B.S.E.E. and an M.S.E.E. from M.I.T. in 1975. In July of 1975, he joined COMSAT Laboratories, where he is presently a member of the technical staff of the Antennas Department of the Microwave Laboratory. His major projects have involved adaptive interference reduction techniques, antenna cross-polarization measurements, and shaped-beam antenna coverage designs.

A contiguous band multiplexer

M. H. Chen, F. Assal, and C. Mahle

(Manuscript received March 18, 1976)

Abstract

A 4-channel waveguide-manifold-type contiguous band multiplexer has been designed, fabricated, and tuned to meet the essential specifications of an INTELSAT V zone output multiplexer. This development allows a substantial weight savings in the satellite transponder in the case of complex antenna systems such as those contemplated for INTELSAT V. Current satellites use multiplexers with guard bands of more than one full channel bandwidth and combine the transmit power of the even and odd channels into two separate output ports, thereby requiring two transmit antennas. A contiguous band multiplexer reduces the guard band to 10 percent of the channel bandwidth and combines all the channels (even and odd) into a single output port, thereby requiring only one transmit antenna. An insertion loss of less than 0.5 dB has been achieved in the experimental model, the feasibility of the concept has clearly been established, and consequently one set of transmit antennas (hemispherical and zonal) can be eliminated in an INTELSAT V application.

Introduction

In communications satellites, the allocated 500-MHz communications band is subdivided into contiguous frequency increments (channels) to

provide almost linear power amplification for multicarrier signals and to maximize the e.i.r.p. per bandwidth increment. For example, as shown in Figure 1, twelve 36-MHz-bandwidth channels are used in INTELSAT IV [1]. The input and output multiplexers channelize and combine a maximum of six nonadjacent channels. For input multiplexers, the resulting 3-dB loss can easily be absorbed. In the case of output multiplexers, where even and odd channels are connected to different low-loss waveguide manifolds, a minimum of one channel bandwidth is used as a guard band between adjacent channels and two transmit antennas are required (see Figure 2).

To efficiently reuse the allocated 500-MHz communications band and thereby increase spacecraft communications capacity, future satellites such as INTELSAT V must provide highly shaped beams with low sidelobe contents and high polarization purity. Since the antennas required are likely to be considerably larger and heavier then those used in INTELSAT IV, the successful design of contiguous band multiplexers (i.e., all channel filters connected to one waveguide manifold) would halve the number of transmit antennas, thereby making the satellite design physically smaller, mechanically less cumbersome, and significantly lighter. Singly terminated filters have been selected for the design of the contiguous band multiplexer to facilitate the requirements for broadband matching.

Designs of contiguous band multiplexers using singly terminated filters have been discussed previously [2]-[4]. The structures assumed in References 2 and 3 require that all the channel filters be connected in series or in shunt at a common point, thereby limiting the number of channels and making this approach difficult to implement with waveguide structures. To accommodate a larger number of channels, a multiplexing technique has been developed [4] using a waveguide manifold. However, the high insertion loss (greater than 6 dB) due to the manifold termination and mismatch severely limits its usefulness.

This paper describes the design and tuning of a contiguous band multiplexer consisting of four channel filters, two reactive structures to match the first and last channels, and a short-circuited waveguide manifold. The filter and the reactive networks are connected to the manifold so that the physical separation between adjacent structures is approximately $\lambda_g/2$. The low insertion loss (less than 0.5 dB) measured in each passband combined with excellent out-of-band rejection satisfy the requirements of INTELSAT V.
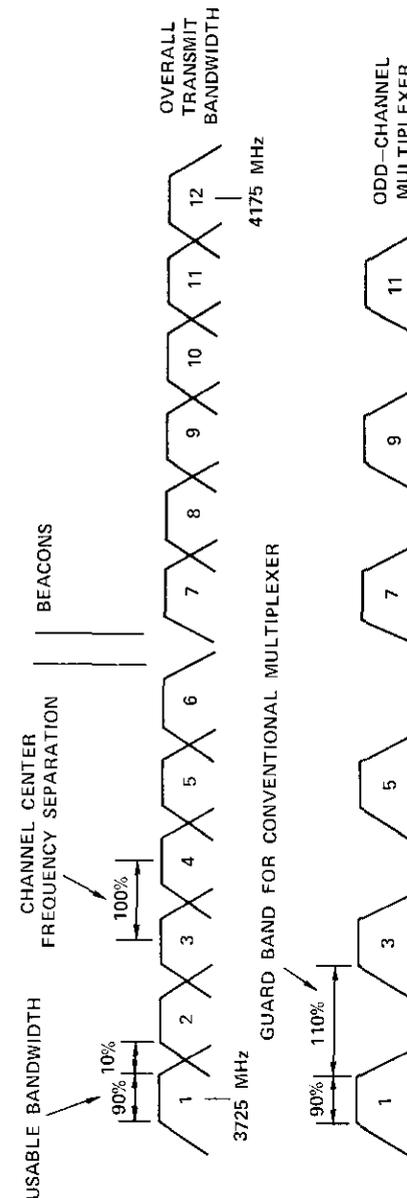
Figure 1. *INTELSAT IV Channel Allocations*
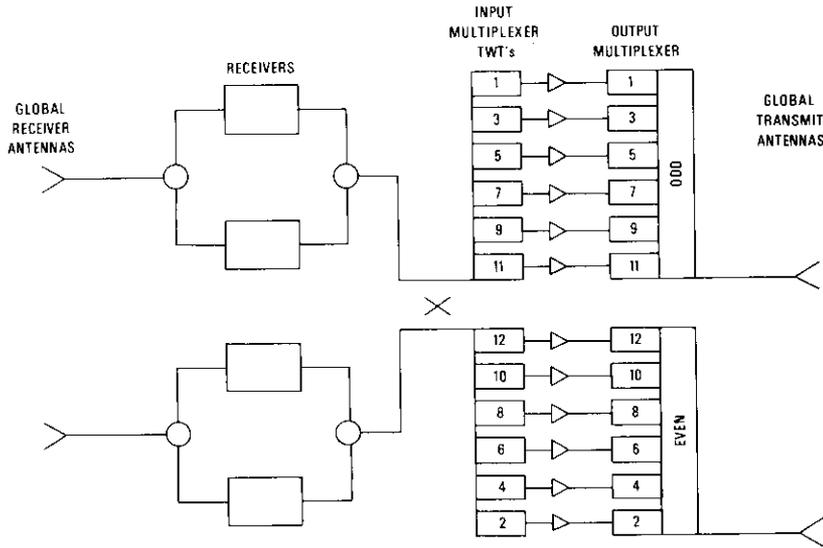
Figure 2. *INTELSAT IV Global Configuration*

Theoretical considerations

Doubly terminated bandpass filters normally used in even and odd multiplexers are not suitable for the design of contiguous band multiplexers, since the input impedance of these filters displays a large and rapidly varying reactance in the adjacent passband that cannot be readily offset. Singly terminated filters that display large and rapidly varying reactances in their passband have been selected in this design to compensate for the reactances of adjacent filters.

To identify the limitations and requirements for the design of contiguous band multiplexers using singly terminated filters, typical input impedance characteristics, $Z_{in}(f) = R(f) + jX(f)$, have been divided into four frequency bands as shown in Figure 3. In this definition, "passband" corresponds to the passband of a channel, "X-band" corresponds to the passband of an adjacent channel, and "R-band" corresponds to the minimum design guard band separating two contiguous channels. In the "passband,"

$$R(f) = R_0 + r(f)$$

$$X(f) = -X_0 + x(f)$$
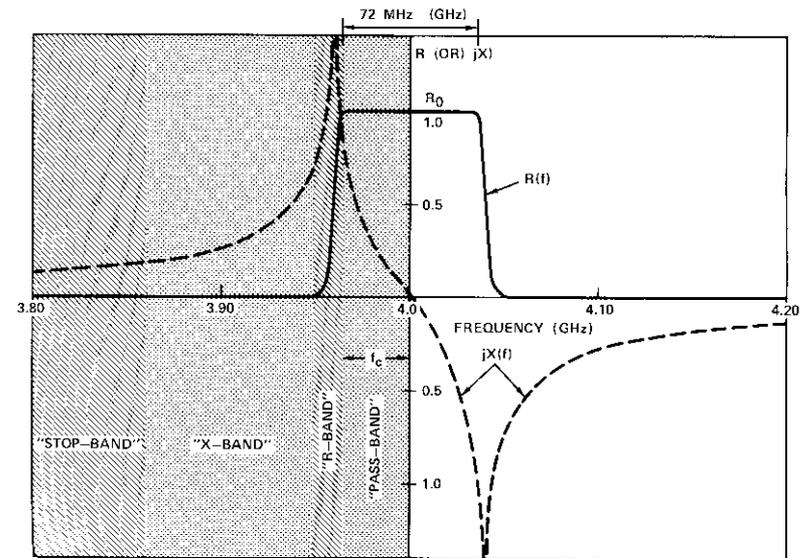
Figure 3. *Typical Impedance Characteristics for Singly Terminated Filters*

where $r(f)$ and $x(f)$ relate to the in-band ripple of the filter. $R(f)$ and $X(f)$ display even and odd symmetry, respectively, and for the filter characteristics of interest,

$$R(f) \cong R_0 \gg r(f)$$

and

$$\frac{\partial X_0(f)}{\partial f} \gg \left| \frac{\partial x(f)}{\partial f} \right| > 0 \quad .$$

In the R-band, $R(f)$ decreases rapidly. The extent of the R-band determines the minimum frequency separation between contiguous passbands since signal energy will be dissipated in two contiguous filters. The R-band can be decreased by increasing the filter selectivity and hence increasing the number of singularities (poles) used to realize the filters. However, it should be noted that the number of poles should not be increased arbitrarily, since the mid-band insertion loss, the overall group-delay variations, and the physical size and weight of the filter will also increase.

In the X-band, $R(f) \cong 0$, $\partial X(f)/(\partial f) > 0$ (Foster reactance theorem), and $X(f)$ varies almost as rapidly as in the "passband." Finally, in the stop band, $R(f) \cong 0$, $X(f)$ increases monotonically, and $\partial X(f)/(\partial f) \ll R_0/2f_c$.

Figure 4 shows the input characteristics of three singly terminated filters with an R-band separating two contiguous filters. Qualitatively, in the "passband" of the second filter, $j[X_1(f) + X_3(f)]$ is almost equal to the complex conjugate of $jX_2(f)$ so that in the second "passband" $R(f) = R_0$ and $[X_1(f) + X_2(f) + X_3(f)] \ll R_0$. This analysis may be repeated in the passband of any filter that is located between two filters. However, the
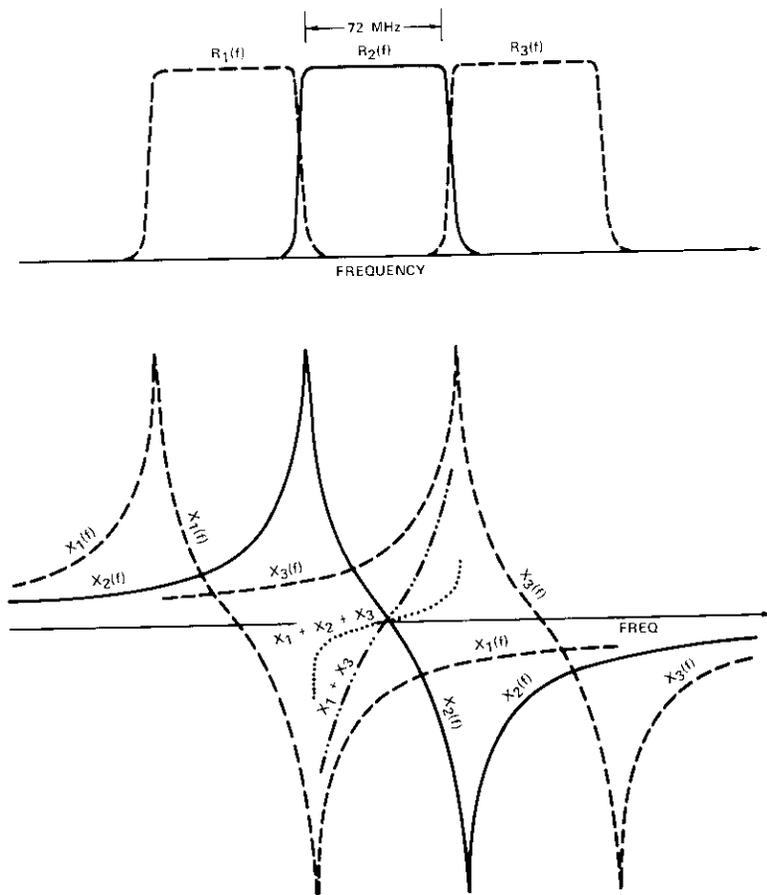
Figure 4. *Contiguous Band Multiplexer with Singly Terminated Filters*

first filter in the structure suffers from a matching imbalance. This problem may be eliminated by assuming that the first filter is extraneous, does not define a channel, and is used only to match the succeeding filter. Similarly, the last filter in the chain must perform the same matching function. A closer look at these characteristics indicates that the passbands are not used for the first and last filters and that they can be replaced by purely lossless reactive networks that have resonances above and below the band edges of the contiguous band multiplexer. The preceding analysis, applicable to filters connected in series or in shunt at a common point [2], is also applicable to the following contiguous band multiplexer design, since the input characteristics of a filter are approximately unchanged if the test ports are moved along a waveguide by $\lambda_g/2$.

In some cases, such as the INTELSAT V channel specifications, both the channel bandwidths and the frequency separations between channels are unequal. Consequently, in addition to the requirement for optimum positioning of channel filters and two reactive networks on a common waveguide manifold, the input impedance magnitude may have to be adjusted to achieve the near-contiguous-band multiplexer requirements. Since the solutions needed to meet or exceed the design specifications are not unique, the multiport broadband matching design described herein has been solved numerically via an interactive computer program [5]. In addition, an optimizing routine that allows the designer to observe the results or participate in the optimization stages has been developed to design a waveguide multiplexer using either calculated or measured filter characteristics.

Feasibility model multiplexer development

A 4-channel contiguous band multiplexer was fabricated to verify the design approach and to exercise the practical tunability of the multiport passive network. INTELSAT V specifications for four contiguous channels assigned to zone beams were assumed for the experimental design. These specifications are presented in Table 1. Significant features of this table are the nonuniform usable bandwidths (72 and 77 MHz), the unequal guard bands (8 and 18 MHz), and the tight gain slopes* at band edges (0.1 dB/MHz). The multiplexer configuration, consisting of four channel filters, two reactive matching networks, and a short-circuited waveguide

---

*The term "gain slopes" is not rigorously correct for a passive device; however, it is commonly used in transponder specifications.

TABLE 1. MULTIPLEXER REQUIREMENTS FOR INTELSAT V

INTELSAT V Channel Definition	1–2	3–4	5–6	7–8
Multiplexer Channel Number	1	2	3	4
Center Frequency, f_k (GHz)	3.7425	3.825	3.905	3.995
Usable Bandwidth, BW_k (MHz)	77.0	72.0	72.0	72.0
Allocated Guard Band (MHz)		8.0	8.0	18.0
Out-of-Band Rejection				
$\quad |f - f_k| = 50$ MHz		≥ 25 dB		
$\quad |f - f_k| \geq 60$ MHz		≥ 30 dB		
In-Band Gain Slopes				
$\quad |f - f_k| \leq 0.9\ BW_k/2$		≤ 0.05 dB/MHz		
$\quad |f - f_k| \leq BW_k/2$		≤ 0.1 dB/MHz		

manifold, is shown in Figure 5. The design of the experimental model followed the procedure indicated in Figure 6.

The equiripple bandwidths for the individual filters were defined to be 2 MHz wider than the channel's usable bandwidth to meet the gain slopes at the band edges and to minimize the overall group-delay variations. The out-of-band rejections needed to limit the R-bands and thereby minimize the gain slope rolloff at the channel band edges exceed the requirements of Table 1. At least 15-dB rejection at the adjacent channel band edge is therefore required to keep the gain slope contribution of the R-band at a negligible level. Hence, either 6-pole elliptic function filters or 8-pole Chebychev filters may be used for this application.

Elliptic function filters are desirable for flight hardware since both the physical dimensions and the weight will be minimized. However, to facilitate the implementation, 8-pole Chebychev filters with 0.1-dB passband ripples have been selected, designed (see Appendix A), and fabricated in separate sections to provide the maximum tuning flexibility. Slots are used for the first and last couplings, and inductive irises with tuning screws are used for all other couplings. The matching networks each consist of two cavities that are slot coupled and designed to approximate the input reactances of the first and fourth channel filters in their respective X-bands.

The multiplexer design consists of determining both the optimum couplings (R_K) of individual filters to the manifold (T-junction), and the optimum position (l_k) of filters and matching networks of the manifold. The interactive computer program described in Reference 5 has been used to obtain an initial design.

As shown in Figure 7, the waveguide manifold is fabricated to permit easy exchange and repositioning of the T-junction coupling slots. The contiguous band multiplexer assembly is shown in Figure 8. The final
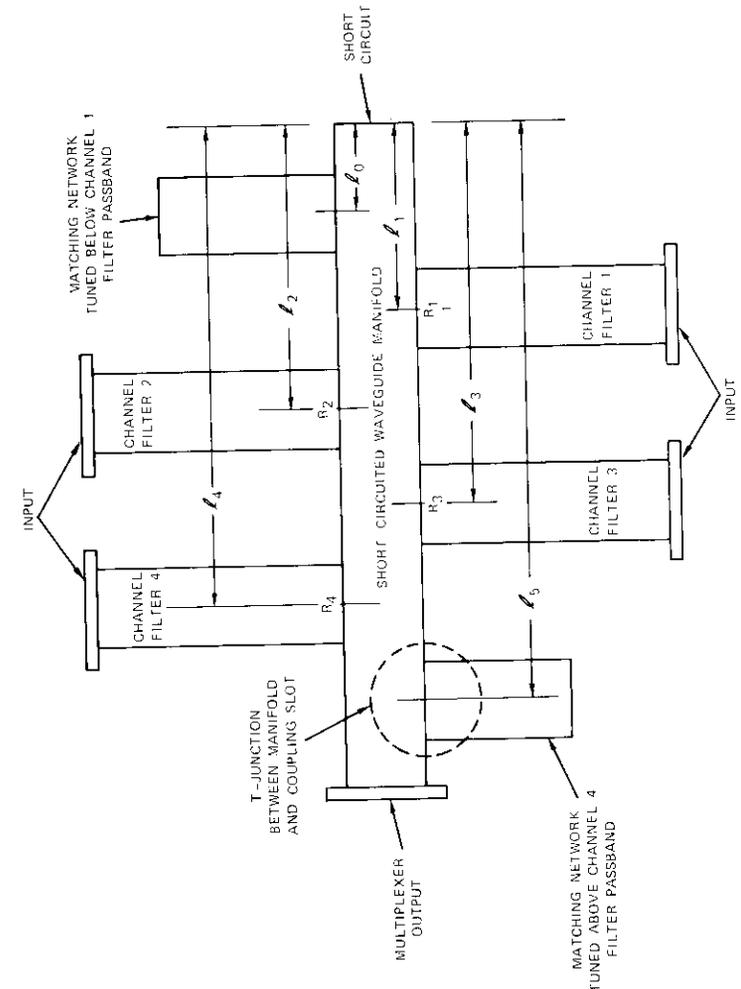
Figure 5. Mechanical Configuration of the Waveguide Multiplexer

```
┌─────────────────────────────────────────────┐
│ 1.  N CHANNEL CONTIGUOUS BAND MULTIPLEXER    │
│     SPECIFICATIONS                           │
└─────────────────────────────────────────────┘
                      │
┌─────────────────────────────────────────────┐
│ 2.  REDEFINITION OF SPECIFICATIONS WITH      │
│     RESPECT TO FILTER DESIGN                 │
└─────────────────────────────────────────────┘
                      │
┌─────────────────────────────────────────────┐
│ 3.  FILTER SELECTION                         │
│     a. SELECT FILTERING FUNCTION (e.g.,      │
│        BUTTERWORTH CHEBYCHEV, AND ELLIPTIC)  │
│     b. DETERMINE FILTER COMPLEXITY TO        │
│        SATISFY IN BAND AND OUT OF BAND       │
│        REQUIREMENTS                          │
│     c. VERIFY THAT "R BAND" WOULD NOT        │
│        SIGNIFICANTLY PERTURB CONTIGUOUS      │
│        "PASSBANDS"                           │
└─────────────────────────────────────────────┘
                      │
┌─────────────────────────────────────────────┐
│ 4.  FILTER DESIGN                            │
│     DERIVE PARAMETERS NEEDED FOR N + 2       │
│     FILTERS TO COMPUTER DESIGN THE           │
│     MULTIPLEXER                              │
└─────────────────────────────────────────────┘
                      │
┌─────────────────────────────────────────────┐
│ 5.  MULTIPLEXER DESIGN                       │
│     a. DETERMINE THE APPROXIMATE POSITIONS   │
│        OF THE N + 2 FILTERS ON A SHORT       │
│        CIRCUITED WAVEGUIDE MANIFOLD          │
│     b. OPTIMIZE THE FILTER POSITIONS AND     │
│        THE FILTER MANIFOLD COUPLINGS OVER    │
│        THE REQUIRED BAND WIDTH (FILTERS 2    │
│        TO N + 1)                             │
│     c. VERIFY THAT COMPUTED PERFORMANCE      │
│        MEETS OR EXCEEDS SPECIFICATIONS       │
└─────────────────────────────────────────────┘
                      │
┌─────────────────────────────────────────────┐
│ 6.  SELECT MATCHING SECTIONS                 │
│     SELECT TWO REACTIVE NETWORKS THAT        │
│     WOULD APPROXIMATE AND REPLACE FILTERS    │
│     1 AND N + 2 OVER THE DESIGN PASSBAND     │
└─────────────────────────────────────────────┘
                      │
┌─────────────────────────────────────────────┐
│ 7.  MECHANICAL DESIGN                        │
└─────────────────────────────────────────────┘
                      │
┌─────────────────────────────────────────────┐
│ 8.  FILTER AND MATCHING SECTION TUNING       │
└─────────────────────────────────────────────┘
                      │
┌─────────────────────────────────────────────┐
│ 9.  MULTIPLEXER ASSEMBLY AND TUNING          │
└─────────────────────────────────────────────┘
```

Figure 6. *Contiguous Band Multiplexer Design Flow Chart*

design parameters, listed in Table 2, were obtained through several tuning iterations performed with computer programs and on the bench. This
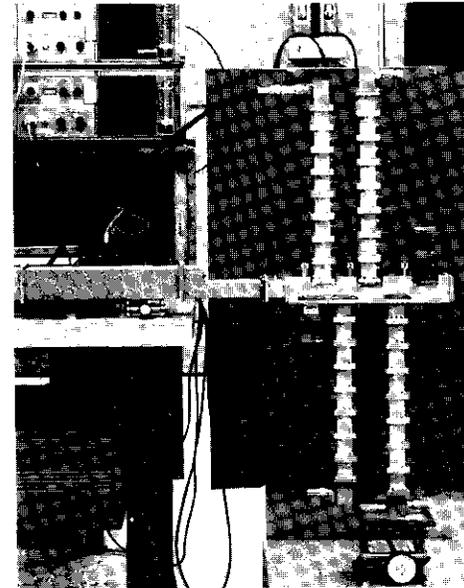
Figure 7. *Photograph of the Waveguide Manifold Used in the Experimental Contiguous Band Multiplexer*
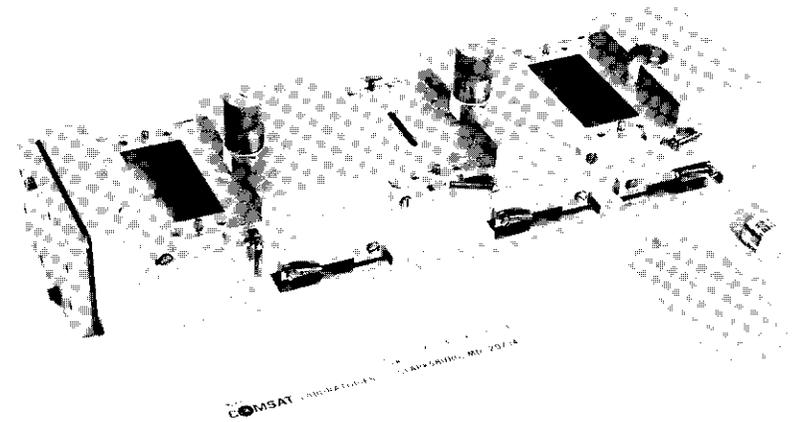
Figure 8. *Contiguous Band Multiplexer in the Test Setup*

TABLE 2. DESIGN PARAMETERS*

Parameter	Matching Network	Channel Number, K				Matching Network
		1	2	3	4	
Normalized R_K	(0.66)	0.66	0.68	0.62	0.59	(0.76)
Channel Center Frequency, GHz	(3.66)	3.7425	3.825	3.905	3.995	(4.095)
Channel Design Bandwidth, MHz	(74.0)	79.0	74.0	74.0	76.0	(74.0)
Filter Location after Tuning (l_k of the manifold of Figure 5), in.	1.037	3.759	5.911	8.004	9.837	10.875

*Numbers in parentheses are for filter functions used only in the multiplexer program of Reference 5.

converging process required both slot repositionings and T-junction coupling slot replacements. It should be noted that the T-junction couplings (R_K) are different for all the channels, since both the specified bandwidths and guard bands are unequal.

Figures 9 and 10 present measurements of the channel characteristics
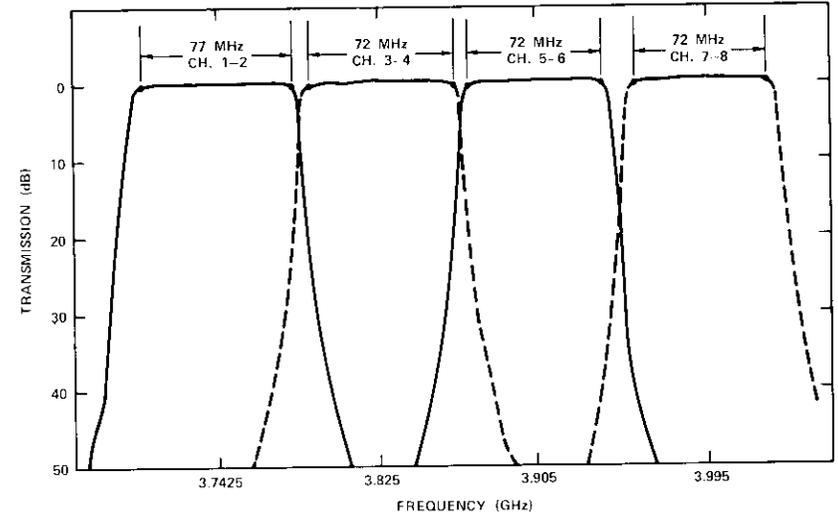
Figure 9. *Measured Transmission Characteristics for the 4-Channel Contiguous Multiplexer That Meets the INTELSAT V Communications Requirements*
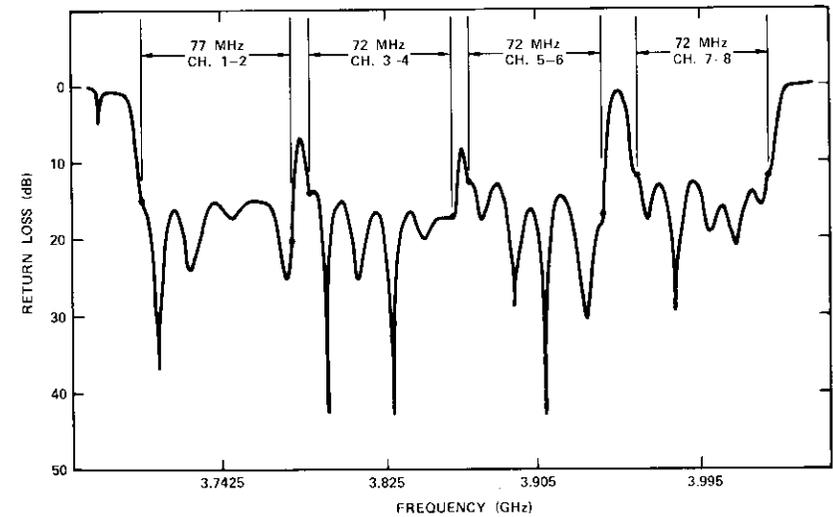
Figure 10. *Measured Return Loss Characteristics*

as seen from the manifold output port. As an example, Figure 11 presents expanded attenuation and group-delay characteristics for channel 1. A detailed evaluation of the measured results shows an insertion loss of less than 0.5 dB at each mid-band and a return loss in excess of 14 dB for most of the passbands. The INTELSAT V specifications for gain slopes at the band edges would be satisfied. Group delay at the band edge exceeds the amount allocated in the INTELSAT V specifications, since much wider filters were originally assumed.
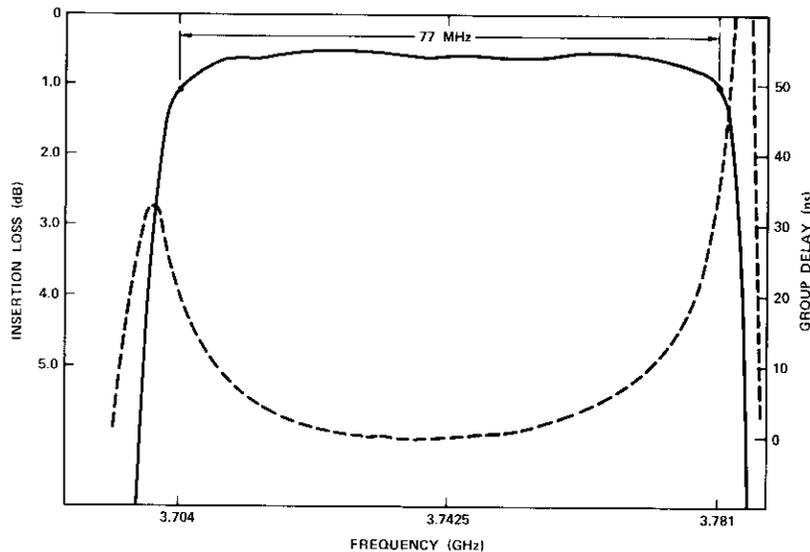
Figure 11. *Measured Group Delay and Expanded Transmission Loss for Channel 1*

Systems considerations

In general, in even/odd channel multiplexing satellite systems, components of an up-link signal that is intended for the kth (e.g., k = odd) channel of an INTELSAT IV or IV-A satellite may also pass through the $(k + 1)$th channel. Depending on spacecraft pointing and receiving earth station location, the resulting down-link signal components recombine in an uncontrolled manner, since the kth and the $(k + 1)$th channels are transmitted through odd and even multiplexers and different antennas. This effect, known as multipath interference [6], is included in the trans-

mission system design and implicitly considered in a 2-antenna satellite specification. In the case of a contiguous band multiplexer, however, the multipath interference is contained within the satellite (between input and output multiplexers) and is a function only of the traveling wave tube amplifier drive levels, which determine the transmission phase shift of the individual tubes. In addition, adjacent channel interference is greatly reduced through a contiguous band multiplexer, since contiguous band filter characteristics do not overlap significantly. This is particularly important in digital transmission when satellite traveling wave tubes will be driven to saturation, resulting in a spread spectrum [7], [8].

Conclusions

A 4-channel contiguous band multiplexer was designed, fabricated, and tuned to meet the essential specifications of an INTELSAT V zone output multiplexer. The feasibility of the concept was clearly established; therefore, one set of transmit antennas (hemispherical and zonal) could be eliminated in an INTELSAT V application. The design approach described herein resulted in practical hardware, although it involved several iterations including experimental steps.

The computer-aided short-circuit tuning procedure presented in Appendix A was used with considerable success to tune the singly terminated filters. This procedure, which is applicable to singly as well as doubly terminated filters, is a systematic step-by-step tuning procedure that is both accurate and timesaving.

The contiguous band multiplexer can be used as a satellite input multiplexer, thereby reducing overall loss and weight (by eliminating several circulators and channel-to-channel attenuation equalizers) at the expense of design and tuning complexity.

Acknowledgments

References

[1] I. Dostis, "Communications Subsystem," *COMSAT Technical Review*, Vol. 2, No. 2, Fall 1972, pp. 277-294.

[2] E. G. Cristal and G. L. Matthaei, "A Technique for the Design of Multi-plexers Having Contiguous Channels," *IEEE Transactions on Microwave Theory and Techniques*, MTT–12, January 1964, pp. 88–93.

[3] G. L. Matthaei, L. Young, and E. M. T. Jones, *Microwave Filters, Impedance-Matching Networks and Coupling Structures*, New York: McGraw-Hill Book Co., 1964.

[4] W. A. Edson and J. Wakabayashi, "Input Manifolds for Microwave Chan-nelizing Filters," *IEEE Transactions on Microwave Theory and Techniques*, MTT–18, May 1970, pp. 270–276.

[5] A. E. Atia, "Computer-Aided Design of Waveguide Multiplexers," *IEEE Transactions on Microwave Theory and Techniques*, MTT–22, March 1974, pp. 322–336.

[6] A. L. Berman, C. Mahle, and M. R. Wachs, "Transmission Modeling," *COMSAT Technical Review*, Vol. 2, No. 2, Fall 1972, p. 495.

[7] G. Robinson, O. Shimbo, and R. Fang, "PSK Signal Power Spectrum Spread Produced by Memoryless Nonlinear TWTs," *COMSAT Technical Review*, Vol. 3, No. 2, Fall 1973, pp. 227–257.

[8] C. Devieux, "Analysis of PSK Signal Power Spectrum Spread with a Markov Chain Model," *COMSAT Technical Review*, Vol. 5, No. 2, Fall 1975, pp. 225–252.

Appendix A.   Filter design and tuning

To design an N-channel contiguous band multiplexer, the computer program of Reference 5 requires the normalized coupling coefficients of $N + 2$ filters, as indicated in Figure 6. These parameters may be derived once the filter func-tion, center frequencies, and bandwidths are selected. The low-pass element values may be derived for an m-pole singly terminated Chebychev filter accord-ing to standard procedures given in Reference 3.

Conventional filter tuning for a doubly terminated filter is based upon the filter's return loss characteristics. Minimization of the reflection in the passband with correct center frequency and bandwidth is the usual criterion for filter tuning. However, since a singly terminated filter is not matched over the entire passband due to the existence of passband reactance, the criterion of minimum reflection may not be used for filter tuning. Instead, a method based upon the short-circuit impedance [A1] may be used. It consists of matching the computed and measured short-circuit impedance of the filter over a wide frequency range and thereby determining the correct tuning screw position. This impedance measurement technique has been used successfully for tuning all the experimen-tal 8-pole Chebychev filters. Figures A-1 and A-2 demonstrate the accuracy of the tuning procedure.

Reference

[A1] A. E. Atia and A. E. Williams, "Measurements of Intercavity Couplings," *IEEE Transactions on Microwave Theory and Techniques*, MTT–23, No. 6, pp. 519–522.
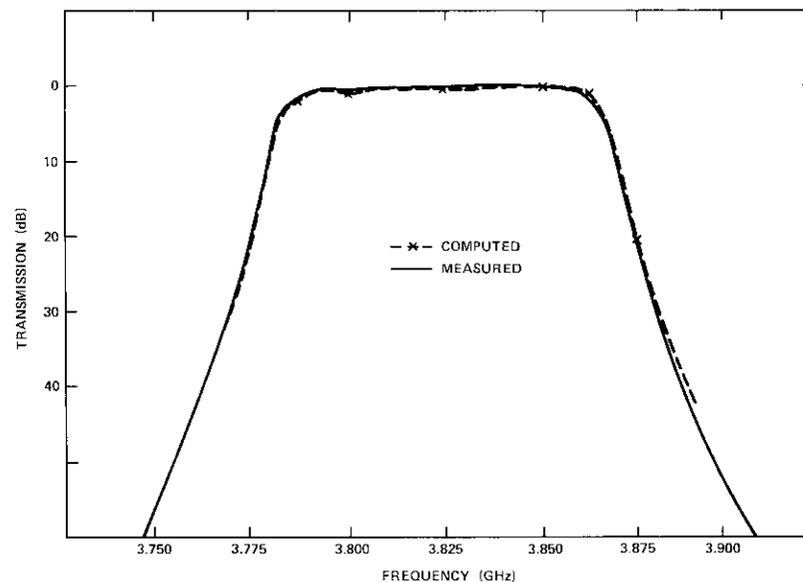
Figure A-1. *Transmission Response for 3.825-GHz Filter*
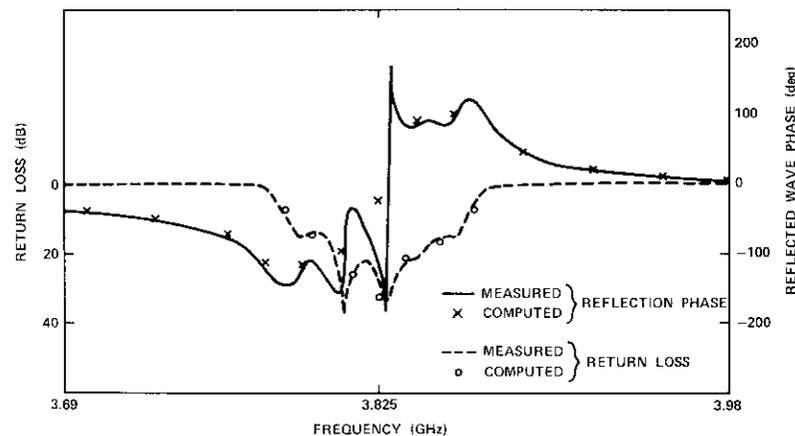
Figure A-2. *Reflection Responses for 3.825-GHz Filter*

Appendix B. Multiplexer simulation and modified T-junction interface

The multiplexer design consists of determining the optimum T-junction couplings to the manifold and the optimum positions of the filters and matching network with respect to the short circuit of the waveguide manifold. The computer program of Reference 5 is an interactive program that requires all $(N + 2)$ filters to have the same coupling matrix. This computer program, which does not include the dispersive effects of the filter coupling slots and does not simulate the perturbations caused by the T-junctions, has been used as a baseline for the initial design (see Figures B-1 and B-2). The computed filter positions are given in Table B-1. Two matching sections have been designed to approximate the reactances of the two additional filters over the full communications band. Four filters and two matching networks have been fabricated, tuned, and measured using the computer-derived positions (Table B-1), and the input characteristics as seen from the manifold port have been measured. A comparison of previous predicted and measured return loss characteristics, shown in Figure B-2 and the solid curve of Figure B-3, respectively, indicates large discrepancies.
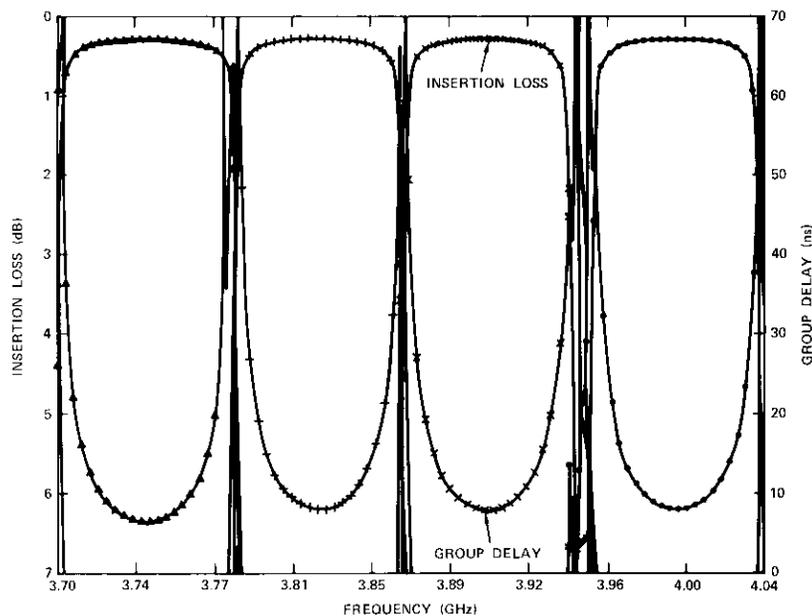
Figure B-1. *Computed Nondispersive Insertion Loss and Group-Delay Characteristics of the 4-Channel Contiguous Band Multiplexer*
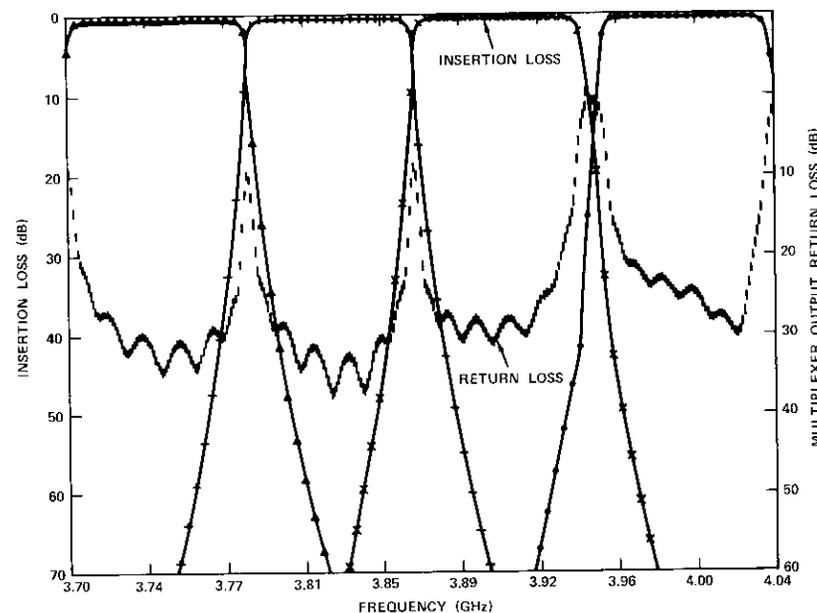
Figure B-2. *Computed Nondispersive Insertion and Return Loss Characteristics of the 4-Channel Contiguous Band Multiplexer*

TABLE B-1. FILTER POSITIONS ON SHORT-CIRCUITED WAVEGUIDE MANIFOLD

Filter Location, l_i (in.)	Matching Network	Channel Number				Matching Network
		1	2	3	4	
Computed (initial design)	1.910	4.210	6.375	8.460	10.115	11.120
Measured (after tuning)	1.037	3.759	5.911	8.004	9.837	10.875
Δl_i	0.873	0.451	0.464	0.456	0.278	0.245

To study these effects, a revised computer program that accepts measured data points of the filters and includes the modified T-junction described below has been developed. The filter-manifold interface may be defined as a modified T-junction (Figure B-4) by comparing equivalent circuits for a slot in a rectangu-
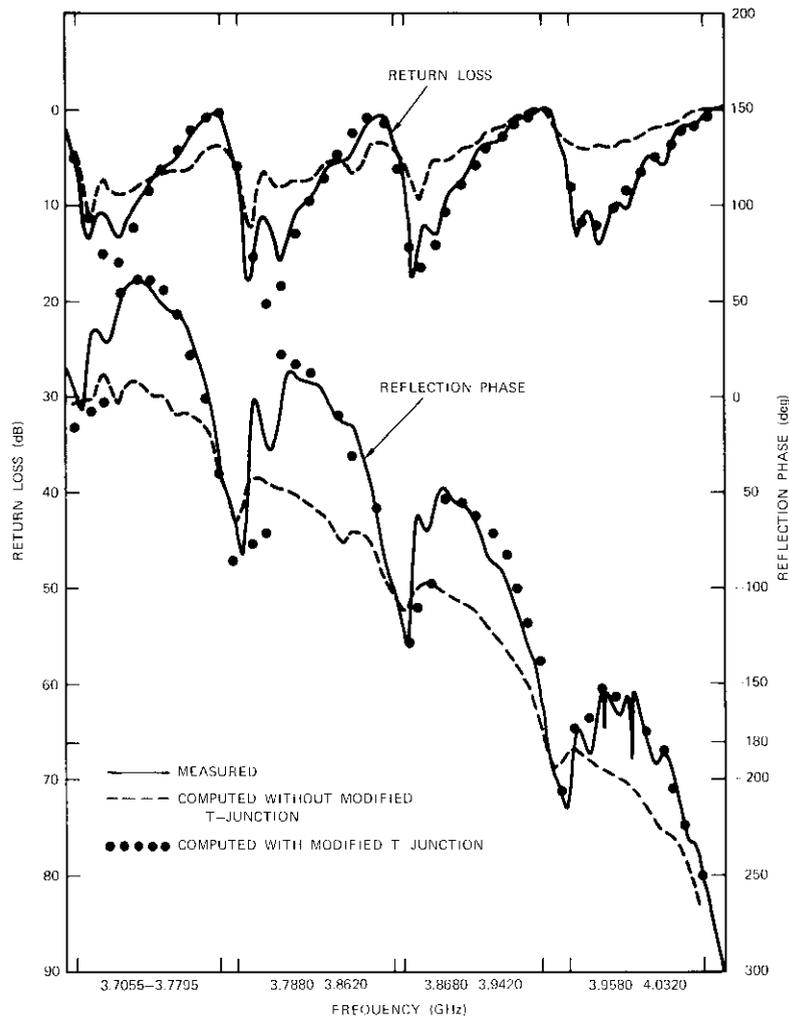
Figure B-3. *Measured Return Loss Characteristics Compared with Computed Characteristics with and without T-Junction Modification*

Figure B-4. *Equivalent Circuit for the Filter-Manifold Interface*

lar waveguide and a T-junction [B1] connecting two rectangular waveguides through a slot, since the input susceptivity of the filter must be subtracted from the susceptance of the T-junction. Furthermore, since the wall thickness and the slot size are not small, the exact values for coefficients C_C and C_M given in Figure B-4 are not known.

Assuming that $C_C = 20$ and $C_M = -1.0$ has resulted in a reasonable match between computation and measurement for a single filter on a manifold. It has been further assumed that these coefficients are approximately the same for all the modified T-junctions. The dashed and dotted curves of Figure B-3 have been computed by using these assumptions, measured filter characteristics, and the previously predicted slot positions as inputs to the revised computer program. Since a good match between the measured return loss (solid curves) and the computed return loss (dotted curves) in Figure B-3 has been obtained, these assumptions for the modified T-junction appear to be satisfactory. Therefore, including this modified T-junction in the computer program of Reference 5

would result in an improved design procedure for the multiplexer. The results presented in Figures 9-11 were obtained by experimentally tuning the multiplexer on the bench. The new filter positions are compared with the previously predicted positions in Table B-1.

Reference

[B-1] N. Marcuvitz, *Waveguide Handbook*, Dover Press, 1965, pp. 238, 248, 363–365.

Ming Hui Chen received a B.S.E.E. from Cheng Kung University, Taiwan, China, in 1960; an M.S. in electronics from Chiao Tung University, Taiwan, China, in 1962; an M.S.E.E. from Utah State University in 1964; and a Ph.D. in electrophysics from the Polytechnic Institute of Brooklyn in 1969. Before joining COMSAT, he was Chief Microwave Engineer at Microtech Co., a principal engineer at Radiation Systems, Inc., and a member of the technical staff at M.I.T. Lincoln Laboratory. In 1970 he was appointed Assistant Professor at George Washington University. He joined COMSAT Laboratories in 1974 as a member of the technical staff in the Transponders Department, Microwave Laboratory, responsible for microwave filter development. His research interests include microwave devices, antennas, and wave propagation in periodic structures.

Francois T. Assal received B.E.E. and M.E.E. degrees from the City College of New York. Prior to joining COMSAT in 1969, he held a National Science Foundation Traineeship and was a lecturer at C.C.N.Y. Since joining COMSAT, where he is currently Manager of the Microwave Systems Department of the Microwave Laboratory, he has participated in and contributed to earth station and transponder design and utilization to maximize analog and/or digital communications capacity in the bandwidth-restricted and power-limited environments of satellites. He is a member of Tau Beta Pi and Eta Kappa Nu.

Dr. Christoph Mahle received Dipl. Ing. and Dr. S.C. Tech. degrees from the Swiss Federal Institute of Technology in 1961 and 1966, respectively. As Manager of the Transponders Department of the Microwave Laboratory at COMSAT Labs, he is responsible for directing research and development in such areas as analysis and laboratory verification of communications systems performance applied to satellite transponders, and advancing the state-of-the-art of communications transponders. He has participated in systems design and evaluation of the communications transponders of the INTELSAT IV, IV-A, and V programs. In addition, he has been responsible for the design, development, and testing of the transponder orbited on ATS-6 as part of the COMSAT Millimeter Wave Propagation Experiment. He is a member of IEEE.

Network topologies to enhance the reliability of communications satellites

F. Assal, C. Mahle, and A. Berman

Abstract

In the past, the reliability of failure-prone devices or subsystems was enhanced by providing 2-for-1 or 3-for-2 redundancy. The overall survival probability of large systems degrades as the number of failure-prone devices is increased. Hence, 24 traveling wave tube amplifiers (TWTAS) were included in the INTELSAT IV satellites to achieve a probability of 0.9 that 9 out of 12 channelized transmitters would survive seven years in space.

This paper presents a network topology that would significantly enhance the end-of-life reliability of communications satellites by pooling together all failure-prone devices such as TWTAS. The resulting topology, identified by a double-ring structure, requires a small number of 4-port 4-position switching elements or R switches to provide interconnect flexibility. Enhancement of the end-of-life (7-year) probability of survival from 0.69 to 0.91 is demonstrated for a double-ring redundancy configuration consisting of 12 TWTAS for eight active units. It is also shown that the double-ring topology permits numerous input-to-output accesses to satisfy the requirements of particular designs. For example, the outputs of two TWTAS can be combined to provide greater power for some paths.

Introduction

INTELSAT IV satellites [1] were launched with one redundant traveling wave tube amplifier (TWTA) for each active TWTA, thereby requiring 24 amplifiers for 12 active channelized transponders. Weight constraints in the INTELSAT IV-A satellites [2] limited the transmitter design redundancy to two for one in four global and three for two in 16 hemispherical channelized transponders. INTELSAT V is expected to have 37 active channelized transponders connected to a total of seven transmit antenna beams, hence requiring a large number of TWTAS. Future communications satellites may employ a large number of narrow-beam antennas, each of which may be connected to at least one active amplifier. In this case, operational constraints require that all beams must function at the end of life of the satellite.

This paper presents a redundant network topology [3] which provides flexible routing of N inputs to N outputs through M ($M > N$) failure-prone devices such as TWTAS to significantly enhance the end-of-life reliability of future communications satellites. In the proposed approach, the M failure-prone devices are pooled together, permitting up to $M - N$ failures without loss of full interconnectivity. In most applications, the proposed topology does not increase the number of switches that are normally required in the 2-for-1 or 3-for-2 redundancy configurations, but obtains its flexibility by using 4-port switches. The redundancy network topology is characterized by two switching layers. Each layer consists of M 4-port switches connected to form a ring, and M failure-prone devices connecting the two rings between N inputs and N outputs.

Definition of R and T switches

A key element required to synthesize the proposed redundant network topology is the 4-port 4-position or R switch defined in Figure 1. However,
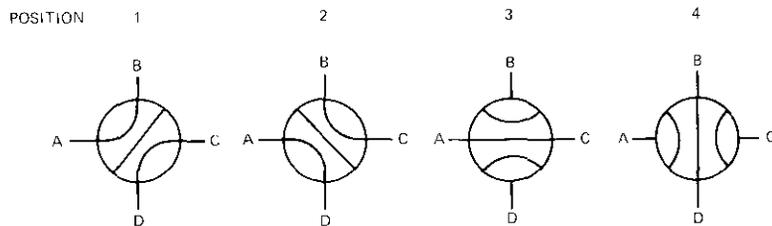
Figure 1. *Definition of an R Switch*

as can be seen by comparing Figures 1 and 2, the 4-pole 3-position or T switch defined in Figure 2 is preferable since it decreases the required switching positions to three and increases the reliability of the redundant network topology by providing additional interconnection paths with minimum switching.
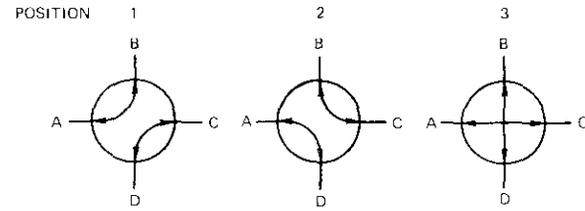
Figure 2. *Definition of a T Switch*

Low-loss R switches are commercially available in a rotary waveguide configuration. These switches would have to be upgraded for space application to minimize their weight and establish their reliability. Low-loss T switches have been developed using PIN diodes, and corresponding coaxial realizations can easily be envisaged using known technology.

Basic operation of the double-ring redundancy configuration

The hypothetical satellite communications subsystem presented in Figure 3 indicates the network topology and its corresponding interconnect flexibility. In this example, 12 TWTAS are provided for eight frequency-channelized transmitters. As can be seen in the figure, all the power amplifiers are pooled together. It is convenient, however, to define as primary the TWTAS (P-1 through P-8) that provide the shortest transmission paths through the input and output channel filters (C-1 through C-8). The four redundant TWTAS (R-1 through R-4) are also indicated in Figure 3.

To easily identify the redundancy configuration, Figure 4 shows part of the transmitter array of Figure 3 as a double ring. The basic building block for this topology is also identified in Figure 4. Four primary amplifier failures (P-3, P-4, P-6, and P-7) are shown to indicate one possible rerouting pattern. It should be noted that other interconnections can be provided. For example, the following rerouting may be commanded: C-2 through R-1, C-3 through P-2, C-4 through R-2, C-5 through P-5, and C-6 through R-3.
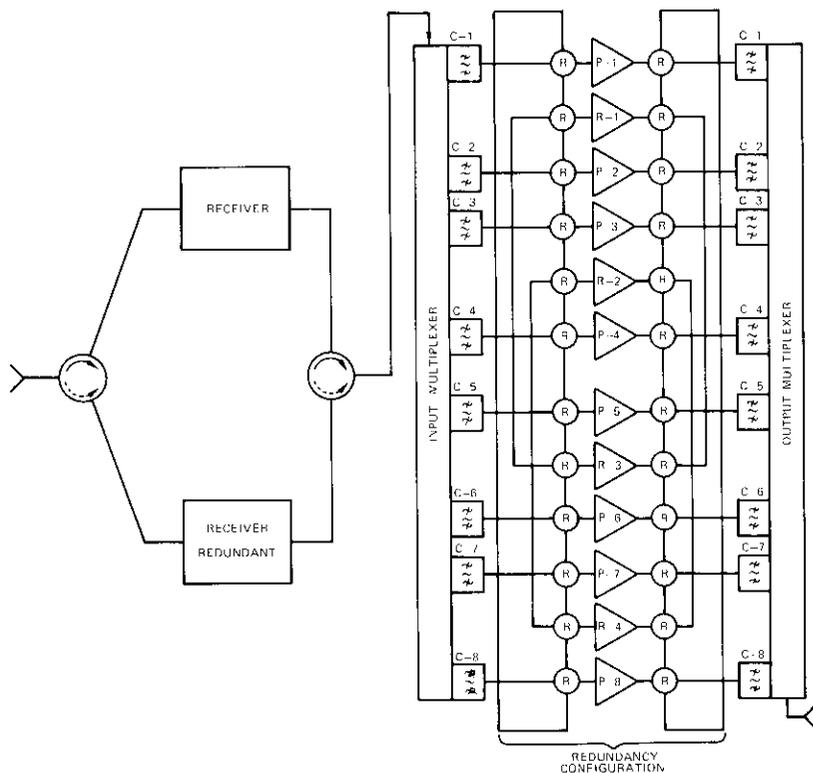
Figure 3. *Satellite Communications Subsystem with 8-Channel 12-TWTA Transmitter*

Figure 4. *Double-Ring Redundancy Network for the Satellite Communications Subsystem of Figure 3 (using R switches)*

Additional switching redundancy can be provided with reduced switching if the 4-position R switches are replaced with 3-position T switches. For example, the routing for channels C-3, C-6, and C-7 would remain as shown in Figure 4, but channel C-4 would be routed directly through R-3 without having to move C-5. To further demonstrate interconnect flexibility with the T switches, the four primary failures in Figure 4 can be rerouted as follows: C-3 through R-1, C-4 through R-2, C-6 through R-3, and C-7 through R-4. It is therefore evident that the routing flexibility provided by R switches is further enhanced by using T switches. As a result, the use of 3-position T switches rather than 4-position R switches would be expected to provide higher reliability.

Reliability computations

In this section, the reliability will be computed for the 12-for-8 double-ring redundancy configuration of Figure 4. In the calculations [4], it will be assumed that only random failures with constant failure rates occur and that the useful life of the components exceeds the satellite mission time of seven years. Therefore, wear-out and early failure mechanisms will be excluded.

It is also assumed that the switches connected to the input and output of a TWTA are coupled mechanically and that a single drive is provided for

the pair. The performance of the 4-position R switch pair and the 3-position T switch pair can be described in terms of the following sets: $\{G, P_1, P_2, P_3, P_4, B\}$ and $\{G, P_1, P_2, P_3, B\}$, respectively, where G indicates that the switches are good or functional upon command; P_1, P_2, P_3, and/or P_4 indicate that the switches have failed and are stuck in one of the stated positions (see Figures 1 and 2); and B indicates that the switches have failed and can no longer be used.

It should be noted that, although the switches may become stuck in one position, full interconnectivity may be provided under certain conditions. For example, if T switches are used in Figure 4, the switches in channel C-5 are stuck in the cross position (i.e., position P_3 as defined in Figure 2), and primary amplifier P-4 fails, then C-4 can be routed directly through redundant amplifier R-3 without having to switch C-5 through R-3 and C-4 through P-5. If R switches are used instead, other channels must be rerouted to accommodate the partial switch failures.

To accurately compute the reliability of the 12-for-8 double-ring redundancy configuration, it would be necessary to assign a probability to each parameter in the aforementioned sets, and to identify all the possible combinations that would result in the successful operation of the satellite transmitter. (Successful operation is assumed if all the channels are operational with either the primary or the redundant amplifiers.) Since the algebraic derivations are rather lengthy, the reliability of the double-ring configuration will be computed by assuming that the switch performance is limited to $\{G, B'\}$, where $B' = \{P_1, P_2, P_3, P_4, B\}$. This switch failure definition should therefore result in conservative estimates of the transmitter reliability.

Let p_S, q_S, p_A, and q_A represent the probabilities of success or failure for a switch pair and a TWTA, respectively. Then, the probability that all eight channels are operational, $P_{12/8}$, is derived:

$$P_{12/8} = p_A^8 p_S^8 [p_A^4 + p_A^3 q_A (4 + 8\, p_S) + p_A^2 q_A^2 (6 + 32\, p_S + 28\, p_S^2)$$
$$+ p_A q_A^3 (4 + 32\, p_S + 120\, p_S^2 + 64\, p_S^3)$$
$$+ q_A^4 (1 + 16\, p_S + 108\, p_S^2 + 248\, p_S^3 + 122\, p_S^4)] \quad .$$

Under certain limited conditions, this equation may be generalized in terms of M and N. However, since numerous exceptions would have to be provided, only the specific case of 12-for-8 redundancy has been presented. If $p_S = 1$ in the above equation, the results will correspond to the most optimistic estimate for transmitter reliability, and the algebraic equation

may be derived directly from the binomial expansion of $(p_A + q_A)^{12}$.

To compare the double-ring configuration with the interconnection techniques used previously in INTELSAT IV-A, i.e., 3-for-2 redundancy, the probability that two channels are operational, $P_{3/2}$, is easily derived:

$$P_{3/2} = p_A^2 p_S^2 (1 + 2\, p_A p_S) \quad .$$

However, since four such 3-for-2 networks are required, the probability that all eight channels are operational is then

$$P_{4(3/2)} = (P_{3/2})^4 \quad .$$

Using the above equations, Figure 5 shows computed results for $P_{12/8}$ ($p_S \neq 1$), $P_{12/8}$ ($p_S = 1$), $P_{4(3/2)}$ ($p_S \neq 1$), and $P_{4(3/2)}$ ($p_S = 1$). In the computations, it was assumed that the TWTA failure rate is 3,000 failures per 10^9 hours and the switch pair failure rate is 100 failures per 10^9 hours. Accurate probability of survival estimates would be between the conservative and the most optimistic solutions presented for each case. These performance characteristics clearly establish the reliability enhancement of the double-ring configuration.

Design of the double-ring redundancy configuration

Appropriate values of M and N are usually obtained from tradeoff studies that take into account operational requirements, communications satellite payload weight and power budgets, and reliability estimates. The M-for-N redundancy configuration is formed by connecting M amplifiers and $2M$ R or T switches to produce a double ring. The input and output ports are obtained by judicious cross-strapping to provide the maximum interconnect flexibility.

As an example, future communications satellites may employ a large number of narrow-beam antennas, each of which may be connected to one active amplifier. To satisfy the operational requirement that all the beams function at the end of life of the satellite, all the TWTAS ($M = 2N$) with equal output power may be pooled together. The overall broadband transmitter reliability can be further enhanced by incorporating the driver stages into the double-ring configuration as shown in Figure 6. As can be seen from Figure 7, up to five consecutive TWTAS may fail in this configuration before flexibility or service is degraded. A set of five S switches, defined in Figure 8, has been included at the output to increase the inter-
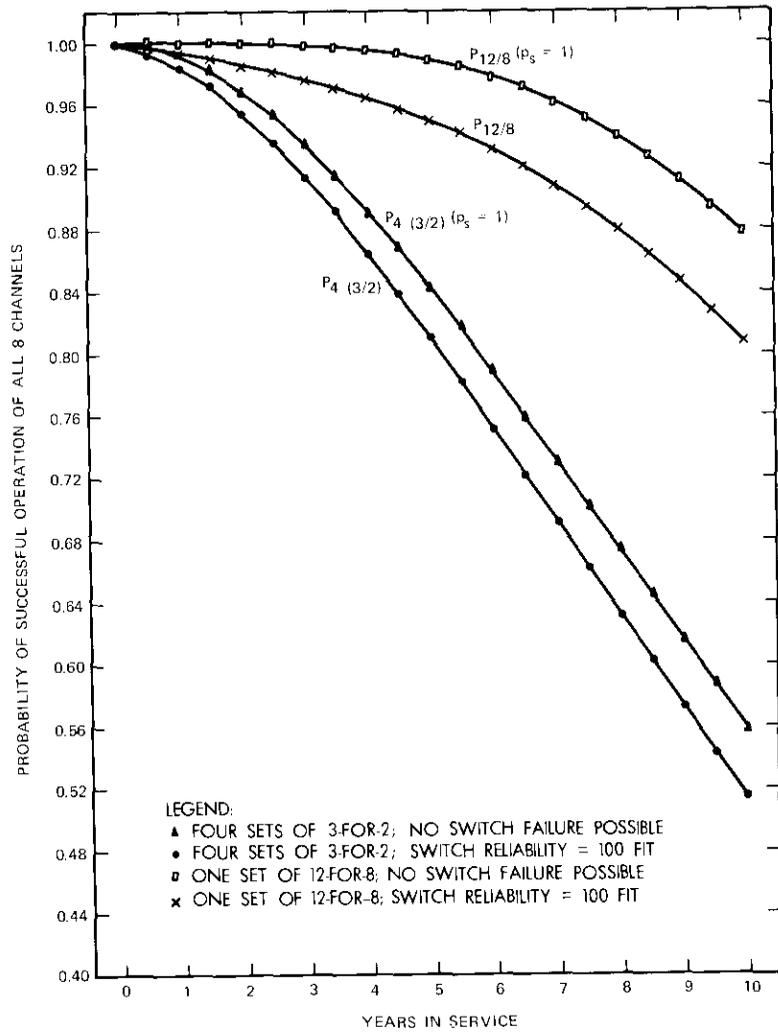
Figure 5. *Channelized Transmitter Reliability (eight active TWTAs and four redundant TWTAs)*
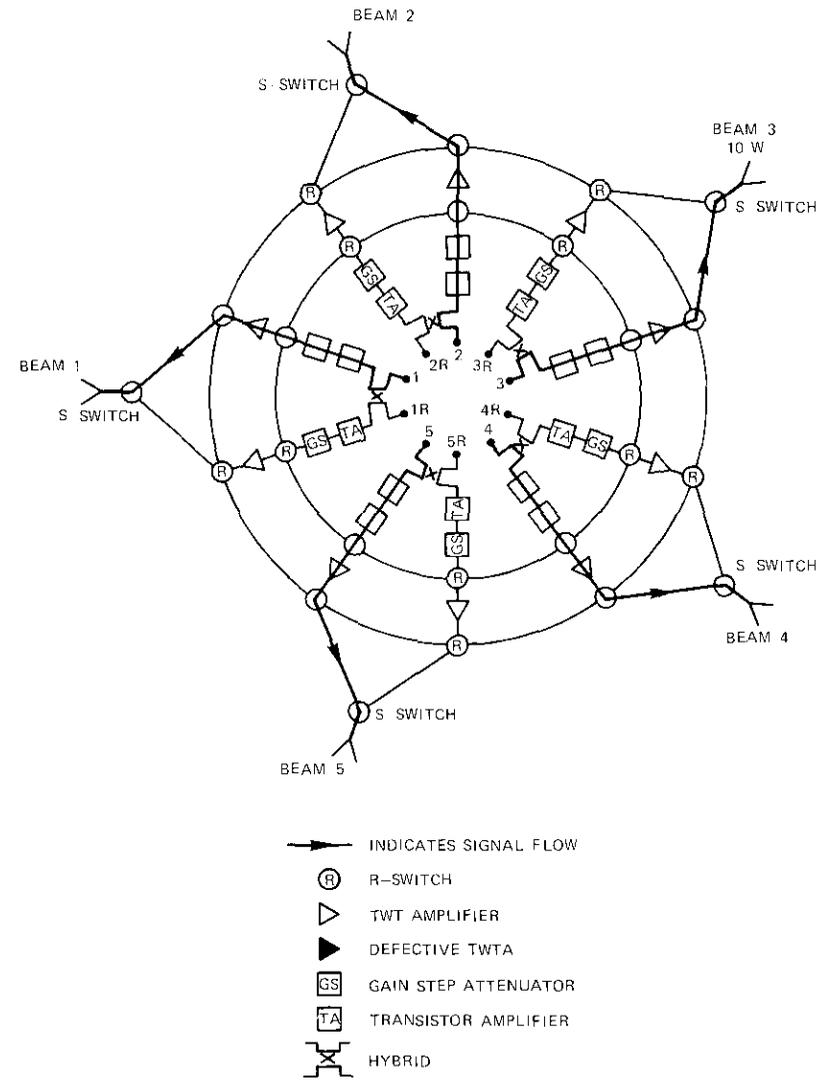
Figure 6. *Broadband Multibeam Transmitter Configuration Using the Double-Ring Redundancy Network*
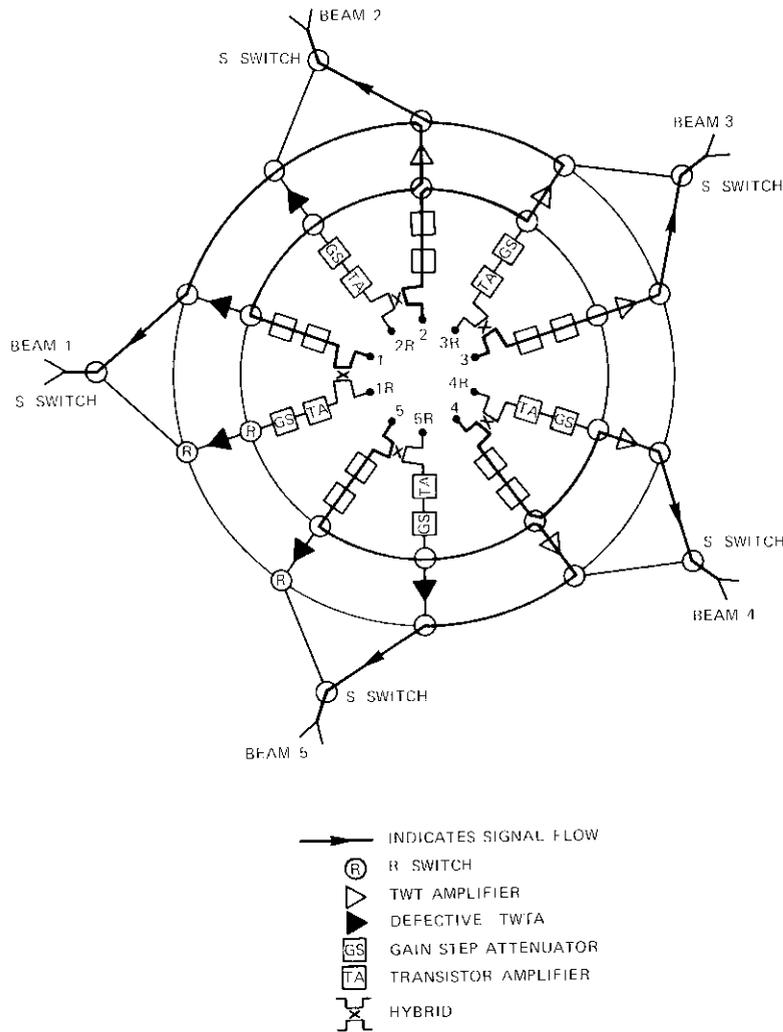
Figure 7. *Worst-Case Failure for the Transmitter of Figure 6*
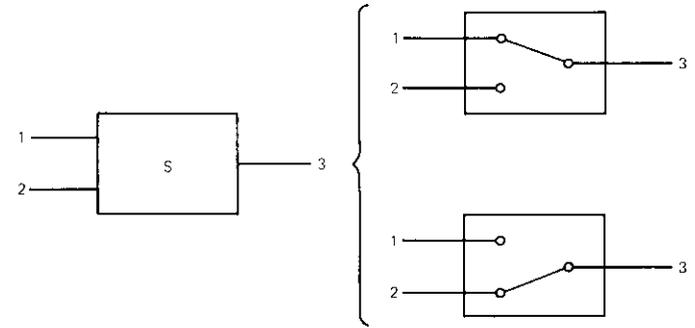
Figure 8. *Definition of an S Switch*

connect flexibility. As shown in Figure 9, it is also possible to combine the outputs of two amplifiers to increase the output power for a specific path.

Conclusions

This paper has presented network topologies that would significantly enhance the reliability of communications satellites. This is particularly critical for multibeam satellite designs, since a transmitter failure may correspond to a beam failure, which in turn may be considered to be a total satellite failure.

The low-loss 4-position R switch is a key element in the realization of double-ring redundancy networks. A waveguide configuration of R switches that are mechanically driven through rotary movement is already available, but their fabrication must be upgraded to minimize their weight and establish their reliability for satellite applications. The reliability of double-ring redundancy networks would be further enhanced if 3-position T switches were used, since fewer switching states would be required and additional interconnect paths would be provided. However, low-loss waveguide T switches are not yet available.

Acknowledgment
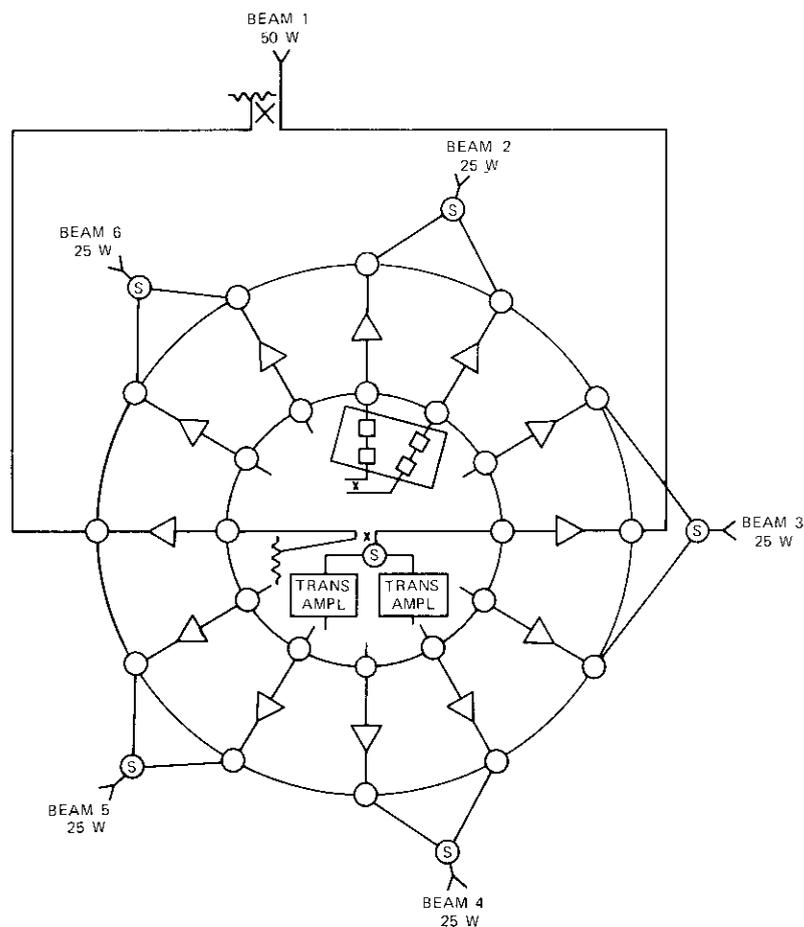
Figure 9. *Broadband Multibeam Transmitter Configuration Using the Double-Ring Redundancy Network (channel 1 output is obtained by combining two TWTAs)*

References

[1] I. Dostis, "Communications Subsystem," *COMSAT Technical Review*, Vol. 2, No. 2, Fall 1972, pp. 277–294.

[2] J. Dicks and M. Brown, Jr., "INTELSAT IV-A Transmission System Design," *COMSAT Technical Review*, Vol. 5, No. 1, Spring 1975, pp. 73–104.

[3] "Redundant Microwave Configuration," U.S. Patent Application Serial No. 670,290, Filed March 25, 1976.

[4] I. Bozovsky, *Reliability Theory and Practice*, New York: Prentice-Hall, 1961.

Francois T. Assal received B.E.E. and M.E.E. degrees from the City College of New York. Prior to joining COMSAT *in 1969, he held a National Science Foundation Traineeship and was a lecturer at C.C.N.Y. Since joining* COMSAT, *where he is currently Manager of the Microwave Systems Department of the Microwave Laboratory, he has participated in and contributed to earth station and transponder design and utilization to maximize analog and/or digital communications capacity in the bandwidth-restricted and power-limited environments of satellites. He is a member of Tau Beta Pi and Eta Kappa Nu.*

Dr. Christoph Mahle received Dipl. Ing. and Dr. S.C. Tech. degrees from the Swiss Federal Institute of Technology in 1961 and 1966, respectively. As Manager of the Transponders Department of the Microwave Laboratory at COMSAT *Labs, he is responsible for directing research and development in such areas as analysis and laboratory verification of communications system performance applied to satellite transponders, and advancing the state-of-the-art of communications transponders. He has participated in systems design and evaluation of the communications transponders of the* INTELSAT IV, IV-A, *and* V *programs. In addition, he has been responsible for the design, development, and testing of the transponder orbited on* ATS-6 *as part of the* COMSAT *Millimeter Wave Propagation Experiment. He is a member of IEEE.*

Arnold L. Berman received a B.S.E.E. from M.I.T. in 1957. Prior to joining COMSAT in 1964, he performed communications system analysis for meteorological and communications satellite systems at the Astro-Electronics Division of RCA, Princeton, New Jersey. At COMSAT he initially participated in the Interim C.C.I.R. Conference (Monaco, 1965). He later participated in the development of the INTELSAT II, III, and IV communications systems and the COMSAT ATS-F Propagation Experiment transponder. He is presently Director of the Microwave Laboratory of COMSAT Laboratories.

PCM/FDMA satellite telephony with 4-dimensionally-coded quadrature amplitude modulation

G. R. WELTI

(Manuscript received March 22, 1976)

Abstract

This paper shows that quadrature amplitude modulation (QAM) with 4-dimensional (4D) coding is more efficient in terms of power and bandwidth utilization than ordinary QAM. Combined with companded PCM, it permits a straightforward increase in FDMA transponder capacity. The shortest code, which contains eight waveforms, provides performance similar to that of 4-phase PSK transmission with rate 3/4 forward error correction and soft decision decoding. For codes with 32 or more waveforms, the bandwidth can be reduced with a minimal increase in power consumption. Power/bandwidth tradeoffs are calculated for 4D-QAM with code lengths from 8 to 1,352. Satellite transponder capacities are found to be 1.5 to 2 times higher with 4D-QAM than with FM.

Introduction

As the terrestrial telephone network progresses toward the use of digital systems, it is appropriate to investigate the transmission of PCM multiplex

signals via satellite. A possible evolutionary approach to the introduction of digital signals into the satellite network is the insertion of digital carriers a few at a time into FDMA transponders, perhaps to replace a few analog carriers. For this approach, it is advantageous to select an efficient modulation system which can be adjusted to maximize transmission capacity over a wide range of carrier-to-noise ratios.

This paper describes an enhanced form of quadrature amplitude modulation (QAM) obtained by introducing coding in a 4-dimensional (4D) signal space. In addition, it explores the application of 4D-QAM to telephony over satellite transponders to show that 4D-QAM gives 1.5 to 2 times greater capacity than FM and provides a similar degree of adjustment flexibility.

Description of 4D-QAM codes

Quadrature amplitude modulation is widely used in data modems for the transmission of digital data over telephone voice channels. It is also employed in the form of 4-phase PSK for PCM transmission of telephone voice channels over INTELSAT SPADE [1] transponders and over the Canadian Anik satellites [2], [3].

The advantages of QAM for digital transmission have been pointed out by a number of investigators, some [4]–[8] concentrating on 2-dimensional (2D) signal mappings, and others [9], [10] considering multidimensional codes. Although codes based on triangular lattices have been shown to provide the densest packing in two dimensions, practical modem implementations have generally incorporated non-optimum square rectangular grids [8]. Welti and Lee [11] have presented a general method for constructing compact 4D codes, and have calculated the performance of certain 4D codes with a 4D peak energy constraint. In principle, 4D codes can be applied to a large variety of transmission systems, including, for example, systems with two or four separate RF carriers operating with different polarizations or at different frequencies. In practice, however, the simplest configuration is obtained when QAM is applied to two successive pulses operating at the same carrier frequency.

The present paper examines the construction and performance of 4D codes with a more practical 2D peak energy constraint. Specifically, the same peak energy constraint is applied to both pulses, i.e., to the sum of the in-phase and quadrature components of each pulse. The codes are designed to retain optimality with square lattices in the 2D projection spaces so that implementation is quite straightforward.

A 2-pulse code can be specified as a set of distinct vectors or codewords:

$$C_i = (x_{i1}, y_{i1}, x_{i2}, y_{i2}), \qquad i = 1, 2, \ldots, M$$

where M is the length of the code, C_i represents the ith codeword, and the four vector components are the in-phase and quadrature (x and y) components of the first and second pulse, respectively, as indicated by the second subscript. The normalized average 2D energy, E_2, is

$$E_2 = \frac{1}{Md^2} \sum_{i=1}^{M} |C_i|^2$$

where d is the minimum Euclidian distance between codewords:

$$d = \min_{\substack{i, m \\ i \neq m}} \sqrt{(x_{i1} - x_{m1})^2 + (y_{i1} - y_{m1})^2 + (x_{i2} - x_{m2})^2 + (y_{i2} - y_{m2})^2} .$$

A useful class of codes is obtained by setting

$$x_{i1} = \frac{x_0 + e_i}{\sqrt{2}}$$

$$y_{i1} = \frac{y_0 + e_i + 2f_i + k_i}{\sqrt{2}}$$

$$x_{i2} = \frac{x_0 + g_i}{\sqrt{2}}$$

$$x_{i2} = \frac{y_0 + g_i + 2h_i + k_i}{\sqrt{2}}$$

where $e_i, f_i, g_i,$ and h_i are suitably chosen integers; k_i is either 0 or 1; and the biases x_0, y_0 depend on the subclass:

$$(x_0, y_0) = \begin{cases} (0, 0), & \text{subclass 0} \\ (0, 0.5), & \text{subclass 1} \\ (0.5, 0.5), & \text{subclass 2} \end{cases} .$$

These three subclasses correspond to signal constellations with different symmetries in the x, y planes.

Note that the four components of each codeword are made interdependent by the integers e_i, g_i, and k_i. The difference vector between distinct codewords i and m, $(e_i - e_m, e_i - e_m + 2f_i - 2f_m + k_i - k_m,$

$g_i - g_m, g_i - g_m + 2h_i - 2h_m + k_i - k_m)/\sqrt{2}$, is contained in the sub-class 0 code. The number of components in which two adjacent codewords differ is exactly two. For example, if $e_i = e_m, f_i = f_m, g_i = g_m$, and $h_i = h_m$, but $k_i \neq k_m$, then the difference vector is $(0, \pm 1, 0, \pm 1)/\sqrt{2}$ and has unit length. Since there are six possible combinations of two non-zero components among the four components, and each combination can produce four sign patterns, a codeword can have no more than 24 adjacent codewords. These codes have $d = 1$. Signal constellations for some codes in this class are shown in Figure 1, where the white points correspond to codewords with $k = 0$, and the black points correspond to codewords
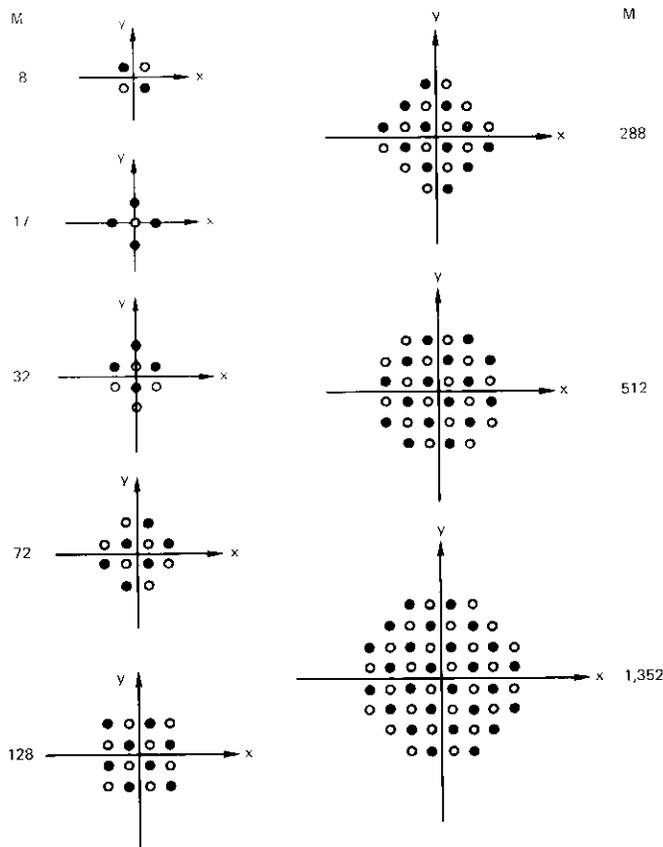
Figure 1. *Signal Constellations*

with $k = 1$. Codes for other values of M can be derived from these codes by expurgation.

In Figure 1 note that the signal constellation for $M = 8$ is identical to a 4-phase PSK constellation. This is also true of the $M = 17$ constellation if the point at the origin is expurgated. In fact, in this case, the expurgation eliminates the interdependence of codeword components, and the 4D-QAM system degenerates to a conventional 4-phase PSK system. In the former case, the 4D-QAM $M = 8$ system is indistinguishable from 4-phase PSK with rate 3/4 forward error correction block coding.

Modem implementation

The general implementation approach for 4D-QAM modems is the same for all values of M, but the specific component parameters such as register length depend on the specific code. The following description uses the parameters applying specifically to the subclass 2, $M = 128$ code, but illustrates the general approach for other codes as well.

Figure 2 is a typical modulator implementation block diagram. The modulator inputs are a sequence of binary digits and a sequence of clock pulses. The input bits are read serially into an encoder, which simultaneously stores seven bits identified as $b_1, b_2, \ldots, b_7$. The encoder output consists of four parallel bits. The first output word is b_1, b_2, b_3, b_4; the second is b_5, b_6, b_7, b_8, where b_8 is the modulo-2 sum of b_1 through b_7. The next pair of output words has the same format as the first pair, but the individual bits are taken from a new group of seven input digits. The clock rate is multiplied by $2/\log_2(M) = 2/7$.

The detailed encoding process is shown in Table 1 and Figure 3 for $M = 128$. In Figure 3, the binary inputs (b_1, b_2, b_3, b_4 for the first pulse, or
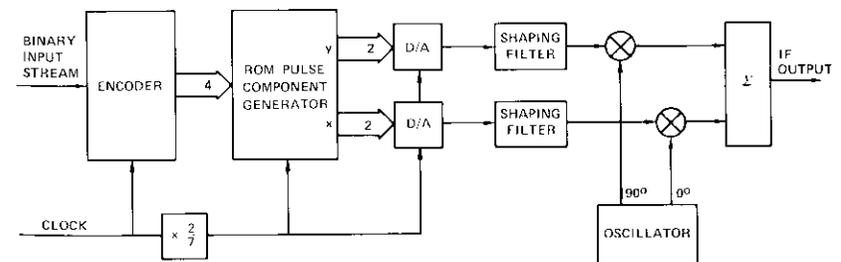
Figure 2. *Modulator Block Diagram*

TABLE 1. DETAILED ENCODING PROCESS FOR $M = 128$

Binary Input								Code-word Index, i	Integers					Codeword, C_i			
b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8		e_i	f_i	g_i	h_i	k_i	x_{i1}	y_{i1}	x_{i2}	y_{i2}
0	0	0	0	0	0	0	0	1	1	0	1	0	1	$\frac{3}{2\sqrt{2}}$	$\frac{3}{2\sqrt{2}}$	$\frac{3}{2\sqrt{2}}$	$\frac{3}{2\sqrt{2}}$
0	0	0	0	0	0	1	1	2	1	0	1	-1	1	$\frac{3}{2\sqrt{2}}$	$\frac{3}{2\sqrt{2}}$	$\frac{3}{2\sqrt{2}}$	$\frac{-1}{2\sqrt{2}}$
0	0	0	0	0	1	0	1	3	1	0	0	0	1	$\frac{3}{2\sqrt{2}}$	$\frac{3}{2\sqrt{2}}$	$\frac{1}{2\sqrt{2}}$	$\frac{1}{2\sqrt{2}}$
			.					.			.			.	.	.	.
			.					.			.			.	.	.	.
			.					.			.			.	.	.	.
1	1	1	1	1	1	0	0	127	-2	-1	-1	0	1	$\frac{-1}{2\sqrt{2}}$	$\frac{-1}{2\sqrt{2}}$	$\frac{-1}{2\sqrt{2}}$	$\frac{3}{2\sqrt{2}}$
1	1	1	1	1	1	1	1	128	-2	-1	-2	-1	1	$\frac{-1}{2\sqrt{2}}$	$\frac{-1}{2\sqrt{2}}$	$\frac{-1}{2\sqrt{2}}$	$\frac{-1}{2\sqrt{2}}$
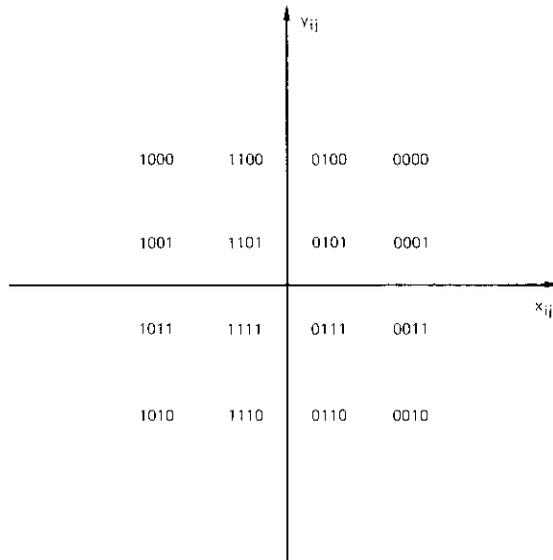
Figure 3. *Signal Mapping for $M = 128$*

b_5, b_6, b_7, b_8 for the second pulse) are juxtaposed with the corresponding signal loci. Note that the binary digits form a 2D Gray code.

The read-only memory (ROM) pulse component generator emits the x and y components of the first and second pulses in digital form. Since the subclass 2, $M = 128$ code employs four x amplitudes and four y amplitudes, the x and y outputs consist of two bits each.

The first or second pulse components are fed to the x and y digital-to-analog (D/A) converters. The converters emit rectangular voltage pulses which are then shaped by shaping filters to produce pulses whose squared Fourier transforms satisfy the Nyquist criteria for zero intersymbol interference. The filter outputs are fed to a pair of balanced mixers with quadrature reference inputs, and the mixer outputs are combined to form an IF (or RF) output.

Figure 4 is a possible demodulator block diagram. The acquisition and tracking logic develops appropriate control inputs to the level controller, the local oscillator, and the clock to ensure that the signals at the output of the matched filters are properly aligned and sampled by the analog-to-digital (A/D) converters at the optimum time. These functions can be implemented by using an approach similar to that used in the experimental 2-pulse amplitude and phase modulation (APM) modem developed for INTELSAT [12].
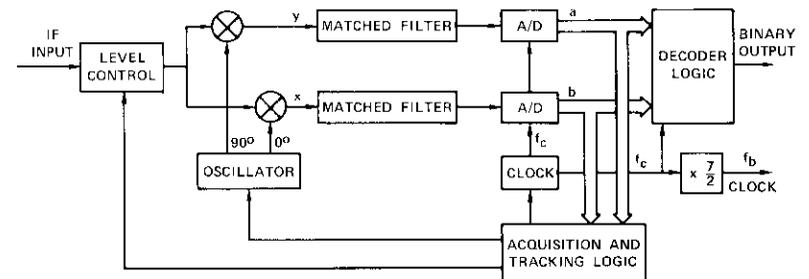
Figure 4. *Demodulator Block Diagram*

After level adjustment, the IF (or RF) signal is converted to zero frequency, and its components are separated by means of a pair of balanced mixers with quadrature reference inputs. The signals are passed through matched filters to a pair of 6-bit A/D converters whose outputs feed both the acquisition and tracking logic and decoder logic. The decoder logic makes a maximum likelihood decision following each received pulse pair

and generates a digital output which corresponds to the subscript i of C_i, the codeword most likely to have been transmitted.*

Performance of 4D-QAM

The quality of a digital communications link is measured in terms of its error statistics. For many applications the salient statistic is the number of decoding errors (or hits) per source bit, P_B. In conventional 2- and 4-phase PSK systems, each bit is decoded independently, and the distinction between decoding errors and bit errors is lost. A decoding error usually produces two bit errors in differentially coded 4-phase PSK systems, and two or three bit errors in 4D-QAM systems. Nevertheless, in telephony only a single error event is actually perceived by the user.

The probability, P, of erroneously decoding a codeword with a maximum likelihood detector has a very close union upper bound. For channels with additive Gaussian noise, this bound takes the form

$$P = \frac{N}{\sqrt{2\pi}} \int_{d/2\sigma}^{\infty} \exp\left(\frac{-t^2}{2}\right) dt$$

where σ^2 is the noise variance per dimension, N is the average number of adjacent codewords:

$$N = \frac{1}{M} \sum_{i=1}^{M} \sum_{k=1}^{M} \delta(|C_i - C_k| - d)$$

and

$$\delta(x) = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0 \end{cases} \quad .$$

The average carrier-to-noise-power ratio is

$$\gamma = \frac{E_2 d^2}{2\sigma^2}$$

where E_2 is the normalized average energy and the number of decoding

---

*A detailed description of a maximum likelihood decoder implementation is expected to be presented in a future paper.

errors per source bit is

$$P_B = \frac{P}{\log_2(M)} \quad .$$

Therefore,

$$P_B = \frac{N}{\sqrt{2\pi} \log_2(M)} \int_{\sqrt{\gamma/2E_2}}^{\infty} \exp\left(\frac{-t^2}{2}\right) dt \quad .$$

Values of N and E_2 for $M = 8$ to 1,352 are listed in Table 2.

TABLE 2. CODE CHARACTERISTICS

Code Length, M	Number of Adjacent Codewords, N	Average Energy, E_2
8	6.000	0.250
17	5.647	0.471
18	8.111	0.458
32	9.063	0.625
72	11.778	0.917
97	10.392	1.196
128	13.500	1.250
225	14.436	1.653
242	14.091	1.773
288	14.667	1.917
338	15.485	2.010
512	16.313	2.500
697	16.390	2.967
722	16.889	2.928
800	16.563	3.089
968	16.711	3.523
1,017	17.400	3.575
1,352	17.763	4.096

The relationship between P_B and γ is shown in Figure 5 for eight representative codes. Note that for 4D-QAM these curves are spaced about 1.5 dB apart from each doubling of M as opposed to 2D-QAM, where the spacing is 3 dB, and PSK, where the spacing is just under 6 dB. The relatively closer spacing of 4D-QAM means that a nearly optimum code can be

found for any combination of γ and P_B in the region to the right of the $M = 8$ curve in Figure 5. In this respect, 4D-QAM provides almost as much flexibility for digital carriers as FM provides for analog carriers. The $M = 17$ curve indicates essentially the same performance as 4-phase PSK, which is identical to 4D-QAM with $M = 17$ and the (0, 0, 0, 0) codeword expurgated. The $M = 8$ curve represents the performance of a 4-phase PSK system with rate $3/4$ block encoding and soft decision decoding.
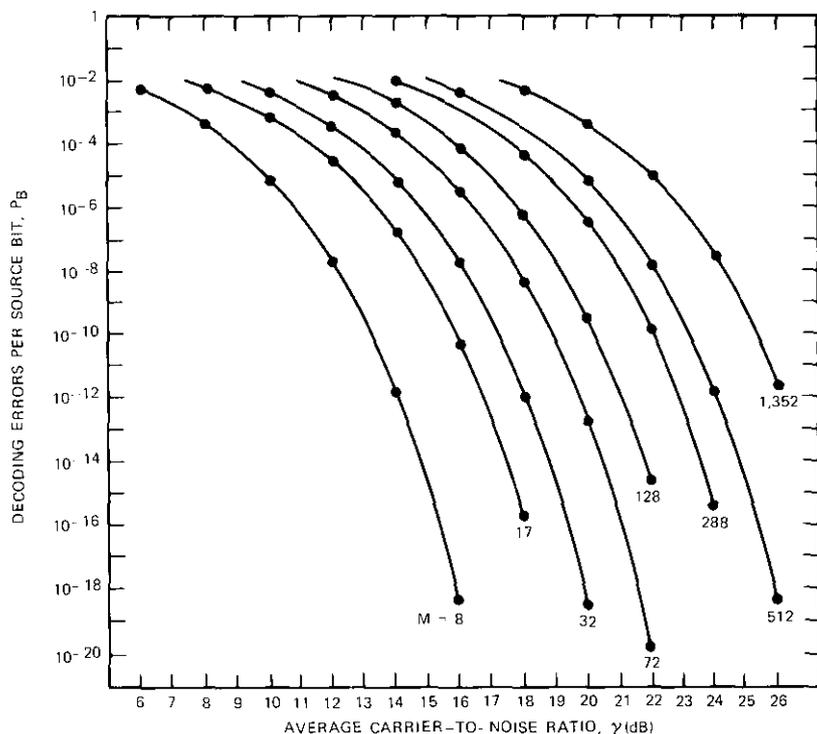
Figure 5. *4D-QAM Performance*

The information transmission rate is $R = \log_2(M)/2$ bits per cycle for a given bandwidth. The relationship between R and γ is shown for two values of P_B in Figure 6, which also includes the Shannon bound for error-free transmission, $R = \log_2(1 + \gamma)$. Note that, for $P_B = 10^{-4}$, the 4D-QAM
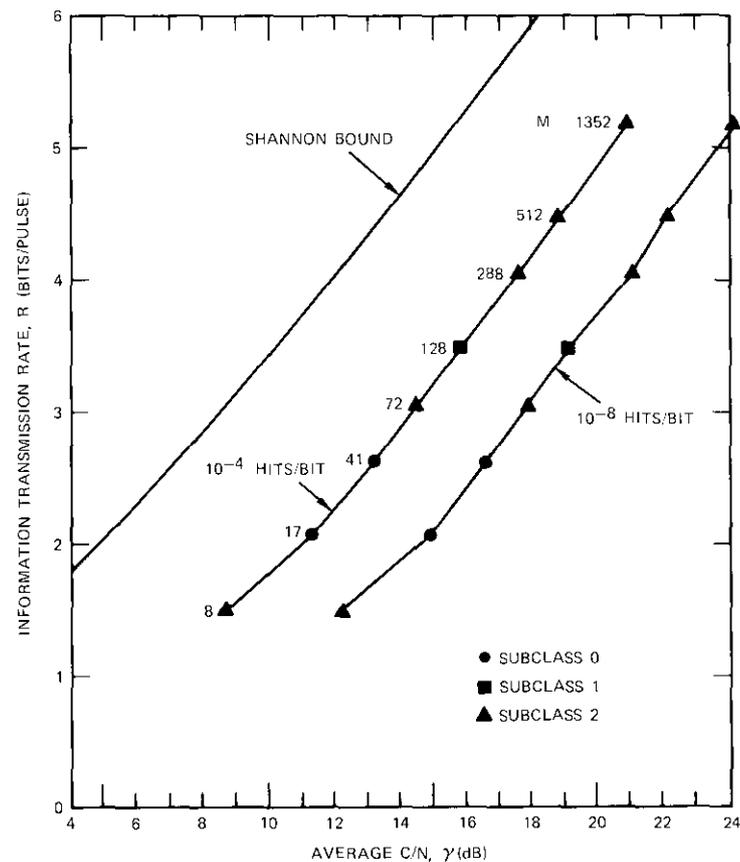
Figure 6. *Average Power Performance (ideal case)*

performance lies 1.5 to 1.8 bits/pulse below this bound.*

The bit-error rate can be derived from P_B by analyzing the bit-error patterns that can be generated by decision errors. The bit-error patterns

---

*For a strictly fair comparison with Shannon's bound, the channel capacities of the 4D-QAM systems should be used rather than the information transmission rates, R. A lower bound on the channel capacity of a 4D-QAM system is

$$\text{channel capacity} > R[1 - P_B \log_2(24)] = R - 4.58 \, RP_B \text{ bits/pulse}$$

since the number of most likely error patterns cannot exceed 24. Hence, the difference between R and the channel capacity is less than 4.58 RP_B, or about 0.001 for $P_B = 10^{-4}$ in Figure 6.

are determined by the specific assignment scheme used to associate the input bits $b_1, b_2, \ldots, b_n, n = \log_2(M)$, with transmitted codewords. In general, the assignment schemes are selected for specific applications, and can be chosen to optimize various criteria, such as high overall signal-to-noise ratio, low bit-error rate, ambiguity resolution, implementation economy, and low radiated power. The minimum bit-error rate is achieved with an assignment scheme employing 2D Gray coding; in this case, the bit-error rate is less than $2P_B$. For a random assignment scheme, the bit-error rate is $(n + 1)P_B/2$, which can range from $2P_B$ to about $5P_B$. However, since the bit-error distribution will be known, forward error correction can be applied, if desired, with much greater economy than for Poisson distributed bit errors.

FDMA transponder capacities and power requirements for 4D-QAM

The ideal power requirements for 4D-QAM are plotted in terms of average carrier-to-noise-power ratio in Figure 6 for two different circuit qualities. The practical power requirements are somewhat greater to allow for amplitude, phase, and timing errors in the tracking circuits, as well as for quantization errors and imperfect filtering in the demodulator. Calculations supported by direct experience with 2-phase APM modems indicate that a total implementation margin of 1.5 dB is quite safe.

A more difficult problem than the implementation margin, however, is that of specifying the circuit quality. The C.C.I.R. and C.C.I.T.T. have not made any recommendations on error statistics for PCM circuits. Terrestrial digital networks are presently providing at least 3,596 error-free seconds per hour (which corresponds to $P_b = 1.7 \times 10^{-8}$ at 64 kbps) for 95 percent of the time [13]. At the same time, maximum bit-error rates of 10^{-4} are being specified for PCM voice transmission [14]. The former circuit quality is, of course, essential for data transmission, while the latter must be regarded as acceptable only for voice transmission.

The channel bandwidth assigned to a QAM carrier is

$$B = \frac{(1 + \rho) R_b}{R}$$

where ρ is the roll-off factor of the Nyquist filters used in the modem, R_b is the number of bits per second transmitted by the carrier, and

$$R = \frac{\log_2(M)}{2}$$

is the number of bits per cycle associated with a 4D code of length M. However,

$$R_b = \frac{64,000 N_{CH}}{1 - \epsilon}$$

where N_{CH} is the number of voice channels and ϵ is the fraction of service digits in the digital stream. To obtain the capacity of a transponder with a useful bandwidth of B, the three preceding equations can be combined to yield

$$N_{CH} = \frac{(1 - \epsilon) B \log_2(M)}{128,000(1 + \rho)} \quad .$$

For $B = 36 \times 10^6$ Hz, and $(1 - \epsilon)/(1 + \rho) = 0.782$,

$$N_{CH} \approx 220 \log_2(M)$$

per 36-MHz transponder. For ITU digital carriers, ϵ lies between 0.005 and 0.012, thus permitting a ρ value as high as 0.26 in the worst case. Since ρ values of 0.15 are state-of-the-art, the capacity formula is conservative.

Figure 7 shows transponder capacity versus C/N for $M = 8$ to 1,024. Capacities for FM carriers, derived from Reference 15, and for 2-pulse APM, derived from References 12 and 16, are also shown for comparison. This figure indicates the benefits of 4D coding over FM, which uses only 1-dimensional coding. For voice transmission, 4D-QAM performs nearly as well as 2-pulse APM. Although 2-pulse APM also uses 4D coding, it is designed to match the source distribution of analog FDM basebands. For each doubling of M, the required C/N for 4D-QAM increases by about 1.5 dB, and the transponder capacity increases by 220 channels.

Conclusion

4D-QAM appears promising for PCM/FDMA transmission. It provides almost the same flexibility as FM in terms of adaptability to a wide variation in carrier-to-noise ratios. In terms of transponder capacity, 4D-QAM is superior to FM and almost as good as 2-pulse APM. Development work is needed to demonstrate that the ideal performance indicated in this paper

can be closely approached in practice. The performance of practical systems must be carefully evaluated in a realistic environment.
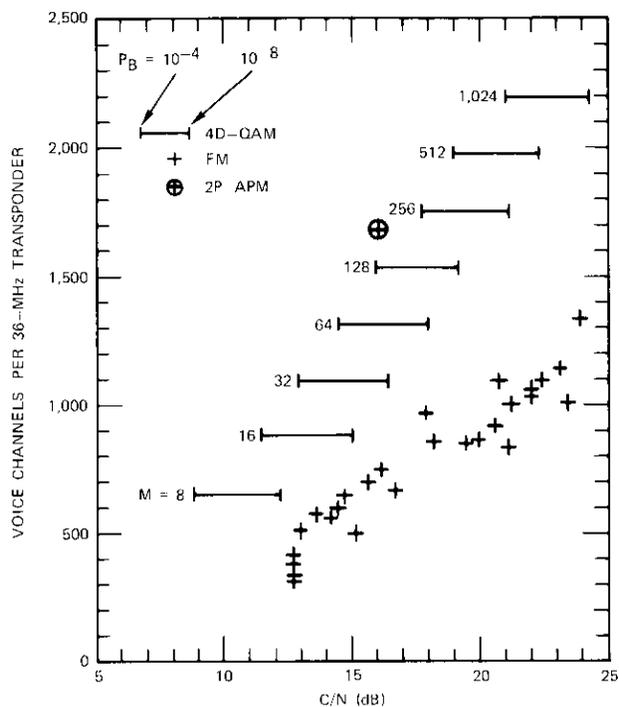
Figure 7. *Transponder Capacity*

Glossary

B Channel bandwidth

b_i ith binary digit of input word

C_i ith codeword

d Minimum Euclidian distance between codewords

E_2 Normalized 2D average energy

e_i, f_i, g_i, h_i, k_i Integers specifying codewords in a subclass of codes

$\exp(t)$ The function e^t

M Code length (number of codewords)

N Average number of adjacent codewords

N_{CH} Number of voice channels

P Probability of erroneously decoding a codeword

P_B Decoding errors per source bit

R Information transmission rate (bits/pulse)

R_b Number of bits per second transmitted by a digital carrier

x_0, y_0 Biases associated with subclasses of codes

x_{ij}, y_{ij} Components of C_i transmitted on the jth pulse

γ Average carrier-to-noise-power ratio

$\delta(...)$ Kronecker delta function

ϵ Fraction of service digits in a digital stream

σ Gaussian noise deviation per dimension

ρ Roll-off factor of a Nyquist filter

References

[1] C. J. Wolejsza, Jr., A. M. Walker, and A. M. Werth, "PSK Modems for Satellite Communications," *INTELSAT/IEE International Conference on Digital Satellite Communications*, November 25–27, 1969, pp. 127–141.

[2] S. Yokogama et al., "The Design of a PSK Modem for the Telesat TDMA System," *IEEE 1975 International Conference on Communications*, June 16–18, 1975, Vol. III, pp. 44-11–44-15.

[3] R. K. Kwan, "The Telesat TDMA System," *IEEE 1975 International Conference on Communications*, June 16–18, 1975, Vol. III, pp. 44-1–44-5.

[4] C. R. Cahn, "Combined Digital Phase and Amplitude Modulation Communications Systems," *IRE Transactions on Communication Systems*, CS-8, September 1960, pp. 150–154.

[5] J. C. Hancock and R. W. Lucky, "Performance of Combined Amplitude and Phase-Modulated Communications Systems," *IRE Transactions on Communication Systems*, CS-8, December 1960, pp. 232–237.

[6] C. N. Campopiano and B. G. Glazer, "A Coherent Digital Amplitude and Phase Modulation Scheme," *IRE Transactions on Communication Systems*, CS-10, March 1962, pp. 90–95.

[7] R. W. Lucky and J. C. Hancock, "On the Optimum Performance of N-ary Systems Having Two Degrees of Freedom," *IRE Transactions on Communication Systems*, CS-10, June 1962, pp. 185–192.

[8] M. K. Simon and J. G. Smith, "Hexagonal Multiple Phase-and-Amplitude-Shift-Keyed Signal Sets," *IEEE Transactions on Communication Systems*, CS-21, October 1973, pp. 1108–1115.

[9] D. Slepian, "Permutation Modulation," *Proc. IEEE*, Vol. 53, March 1965, pp. 228–236.

[10] R. Ottoson, "Group Codes for Phase- and Amplitude-Modulated Signals on a Gaussian Channel," *IEEE Transactions on Information Theory*, IT-17, May 1971, pp. 315–321.

[11] G. R. Welti and J. S. Lee, "Digital Transmission with Coherent Four-Dimensional Modulation," *IEEE Transactions on Information Theory*, IT-20, No. 4, July 1974, pp. 497–502.

[12] T. Furuya, M. Sugiyama, et al., "Hybrid Modem for 120 Channels FDM Telephony Signal Transmission," *Third International Conference on Digital Satellite Communications*, November 1975, pp. 91–98.

[13] C. A. Buzzard, "Error Performance of Digital Data System Channels Between Philadelphia and New York," National Telecommunications Conference, December 2–4, 1974, *NTC '74 Record*, pp. 818–823.

[14] International Radio Consultative Committee (C.C.I.R.), "Fixed Service Using Communication Satellites (Study Group 4)," *XIIIth Plenary Assembly, Geneva, 1974*, Vol. IV, Geneva: International Telecommunications Union, 1975, p. 120.

[15] J. L. Dicks and M. P. Brown, Jr., "Frequency Division Multiple Access (FDMA) for Satellite Communication Systems," *EASCON '74 Record*, October 7–9, 1974, pp. 167–178.

[16] G. R. Welti, "Application of Hybrid Modulation to FDMA Telephony via Satellite," *COMSAT Technical Review*, Vol. 3, No. 2, Fall 1973, pp. 419–430.

George R. Welti received B.S. and M.S. degrees in Mechanical Engineering in 1947 and 1948, respectively, and an M.S. degree in Electrical Engineering in 1955 from M.I.T. He was Director of the Advanced Studies Laboratory from 1968 to 1973, and is presently a Senior Scientist at COMSAT *Laboratories. He was previously employed by Westinghouse Electric Corporation; DECO Electronics, Inc.; Raytheon Co.; M.I.T.; and Sperry Gyroscope Company. He holds a number of patents in fire control, radar, and communications.*

A strategy for delta modulation in speech reconstruction

J. C. Su, H. G. Suyderhoud and S. J. Campanella

(Manuscript received April 23, 1976)

Abstract

An adaptive slope control strategy has been developed and incorporated into a linear delta modulation system operating at 32 kbps for the reconstruction of high-quality speech signals. The linear delta modulation system consists of a bistate comparator in the forward loop and a 2-pole, 1-zero linear low-pass filter in the feedback loop. The adaptive loop for slope control utilizes a variable step size function $Q(I)$, which, at each sampling instant, determines the step size based on the four most recent bits obtained by sampling the comparator.

The signal-to-distortion ratio has been used to objectively evaluate and compare the performance of the adaptive slope control strategy with that of conventional companded PCM. In addition, critical listener preference tests have been conducted. The source speech used in these tests consists of five phonetically balanced test sentences spoken by two male and two female talkers.

The objective performance characteristics are presented for a speech input level dynamic range of 0 to -50 dBm0. The effects of bit-error rates in the range of 10^{-3} to 10^{-6} have also been investigated. Finally, a graphic display of original and reconstructed speech is presented.

Introduction

Digital transmission of speech is becoming increasingly important for both terrestrial and satellite communications applications. The present

standard of digital transmission requires bit rates of 56 or 64 kbps for a voice channel. The search for techniques which will lower the bit rate and hence increase efficiency without significant loss of quality has yielded several alternatives. This is particularly important for satellite applications.

One alternative is delta modulation, a simple and therefore attractive technique of digital transmission which can yield good-quality speech at lower than standard bit rates. Bit rates as low as 19 kbps have yielded reasonable speech quality, although perhaps not good enough for commercial telephone usage. However, at the bit rate of 32 kbps, certain implementations of delta modulation promise to approach commercial quality speech transmission. Hence, efforts by researchers to perfect these techniques are continuing.

COMSAT Labs has developed and implemented a delta modulation strategy which yields high-quality speech. This paper will discuss the details of the development of the strategy and results from computer-simulated implementation of recorded speech passages.

Background

The principle of delta modulation (DM) was described for the first time in a French patent [1]. More literature followed, and detailed descriptions of various aspects were published in 1952 [2]–[4]. Since 1952, extensive studies of adaptive delta modulation have emerged [5]–[9].

As is well known, due to its autocorrelative properties, the voltage produced by continuously sampled speech can be estimated at each sampling instant as a function of its past values. In linear delta modulation (LDM), also known as "constant slope" DM, the estimated value is obtained from a constant amplitude impulse driving a linear filter. In adaptive delta modulation (ADM), also known as variable slope DM, the estimated value is obtained from a variable amplitude impulse driving a linear filter. The adaptivity of the step size is generally designed so that the error signal is smaller for small signals and larger at large signal levels. Several adaptive strategies have been developed [8], [9].

The LDM technique suffers from idle noise, quantizing noise, and slope overload distortion. Furthermore, because of the fixed step size, acceptable performance in terms of signal-to-noise (S/N) ratio can be realized only for a limited dynamic range of input speech levels. As a result of these shortcomings, ADM techniques have been the subject of intensive interest.

Adaptive or variable slope DM systems are generally designed so that their slope can be adjusted to compensate for the change in magnitude

(or slope) of the speech signals. Thus, each value of the reconstructed signal sample is adaptively predicted by means of a time-varying function of its past history. Because ADM maintains a minimum sum of slope overload and quantization distortion, it yields a higher signal-to-distortion (S/N_e) ratio with a wider dynamic range.

This paper is intended to present the results of a variable step size DM strategy which yields high-quality telephone speech. The step size is determined by the most recent four bits of the channel bit stream using a functional relationship designated as $Q(I)$. The function $Q(I)$ is selected so that the error between the incoming speech signal and the remade signal is minimized. The reconstruction loop employs a 2-pole, 1-zero filter with parameters chosen to optimize the telephone speech quality.

Both objective and subjective evaluations have been conducted, the former by using a signal-to-distortion criterion and the latter by comparing critical-listener tests of processed and unprocessed speech. A detailed description of the ADM strategy, the choice of filter parameters, and the results of computer simulation including the effects of bit errors will be presented.

Description of the delta modulation system

Figure 1 is a schematic block diagram of the adaptive DM system, including both the transmitter and the receiver. The transmitter is a dual closed-loop processor, and the receiver is an open-loop processor which has exactly the same structure as the complete feedback loop of the
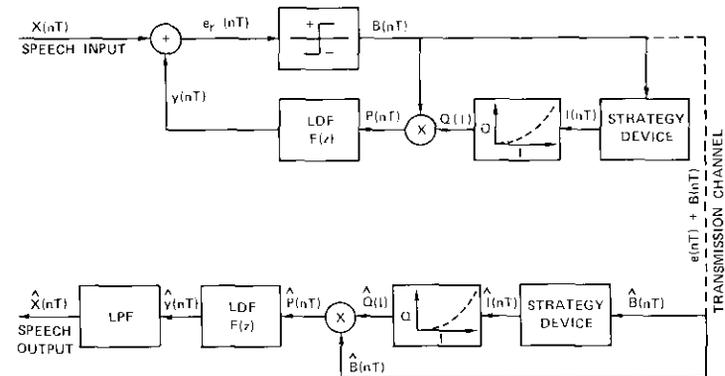
Figure 1. *Adaptive Delta Modulation Systems: Transmitter and Receiver*

transmitter. The operation of the system can be described as follows. The sampled speech signal, $x(nT)$, is compared with the reconstructed speech signal, $y(nT)$, where T is the sampling period and n the sample index. The difference, $e_r(nT)$, between $x(nT)$ and $y(nT)$ is called the "error signal." Note that $y(nT)$ is the result of all x samples up to and including $x[(n-1)T]$, but excluding the present sample, $x(nT)$.

The error signal, $e_r(nT)$, is fed to a bistate comparator which produces a bit stream, $B(nT)$, of either $+1$ or -1, indicating the sign of the error. This bit stream is the signal available for transmission to a receiver through a transmission channel, where bit errors may be introduced. In the transmitting processor, the bit stream simultaneously feeds a strategy device and a bit sign multiplier. The strategy device, by continuously observing the four most recent bits of $B(nT)$, generates an index I which commands the Q function to supply a step size $Q(I)$, which is then fed to the multiplier. The output of the multiplier, $B(nT) \times Q(I)$, is subsequently passed through the digital filter with transfer function $F(z)$, which will be discussed in a subsequent section. The output of the digital filter is the reconstructed speech sample, $y(nT)$, which is an estimate of the present speech sample, $x(nT)$, based on its past history.

As pointed out previously, the structure of the receiver is essentially the same as that of the feedback loop of the transmitter (see Figure 1). However, to eliminate unwanted distortion components, a low-pass filter (LPF) is added at the output of the digital filter. Obviously, if no bit errors are introduced in the transmission channel, $\hat{B}(nT)$ will always be equal to $B(nT)$, and thus $\hat{y}(nT)$ will be equal to $y(nT)$. Consequently, $\hat{x}(nT)$ becomes a close replica of the input speech signal, $x(nT)$. The mean square difference, $E[|x - \hat{x}|^2]$, between these signals is defined as the total distortion power; its value is used to quantify the performance resulting from the processing. This system reduces to a linear DM system when the step size is fixed for all n.

Q function strategy

At each sampling instant the strategy device generates a command to increase or decrease the previous value of I, resulting in a new value of $Q(I)$. The objective is to enable $y(nT)$ to closely follow a rapid increase in $x(nT)$ (steep slope increase control), while the inherent decay time of the feedback filter shown in Figure 1 makes it possible to follow the decreasing values of $|x(nT)|$. The relationship between I and $Q(I)$ is shown in Figure 2. It can be seen that the $Q(I)$ function provides a very large range of quantized correction step size and hence slope rate which can be applied to
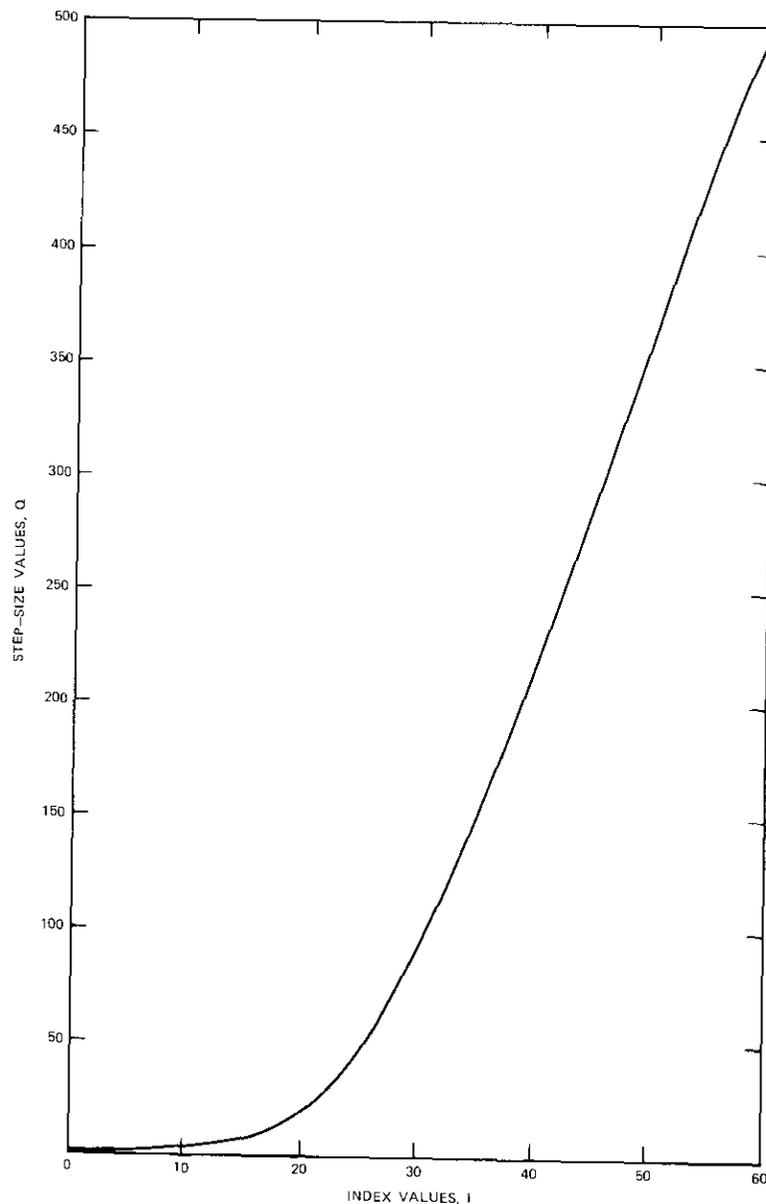
Figure 2. *Step Size Q Function*

the feedback correction loop. As a result, distortion induced by the DM processing is significantly decreased. For input signals having a low rate of change, the quantizing step size should be small to minimize quantizing noise. For such signals the strategy unit must generate low values of index I.

For functions having a high rate of change, the quantizing step size must be much greater and the strategy unit must accordingly generate high values of index I. The unit must also be able to change quickly from low to high index values to follow the abrupt changes in signal slope which are characteristically encountered in speech. The periodic exponentially decaying sinusoid characteristic of larynx excited sounds and the burst and fricative consonants are typical examples of this type of speech function. Changes from high to low index values need not be as rapid because of the exponential decay characteristic of the typical pitch impulse function. It can be seen that the $Q(I)$ function has a nonlinear relationship with I, which is the result of experimental trimming to produce the minimum mean square error for a set of speech samples comprising sentences spoken by men and women.

The rules for increasing the quantizing step size are based on the occurrence of runs of the same bit value. Such runs indicate the occurrence of slope overload, which can be overcome only by increasing the quantizing step size. It will be seen that the logic is organized to provide increasingly rapid step size augmentation to keep pace with abrupt increases in signal rate of change as the run length increases. The following rules apply to the logic for increasing quantizing step size:

a. rule 1: If two consecutive bits have the same sign, the Q value is increased by one step size ($\Delta I = 1$);

b. rule 2: if three consecutive bits have the same sign, the Q value is increased by two steps ($\Delta I = 2$);

c. rule 3: if four consecutive bits have the same sign, the Q value is increased by three steps ($\Delta I = 3$);

d. rule 4: if more than four consecutive bits have the same sign, the Q value is increased by two steps ($\Delta I = 2$).

Because decreases in quantizing step size are required only to follow the exponential decay of the speech pitch period, they need not be as rapid. Hence, there is only one rule for decreasing the quantizing step size. This rule prevails when the bit stream is characterized by randomly alternating bits, a condition which occurs when the quantization step size is equal to or greater than that needed to fulfill the rate of change requirement. The rule is stated as follows:

rule 5: If the sign of the current bit is the opposite of the sign of its immediate predecessor, the Q value is decreased by one step (i.e., $\Delta I = -1$); this rule overrides all others.

These rules have been devised with the following general objectives:

a. Only one rule should apply to any situation.

b. When steep slopes are initially encountered, three consecutive bits with the same sign will occur and rule 2 will ensure a large increase in $Q(I)$.

c. When very steep slopes occur, four consecutive bits with the same sign will result and rule 3 will be invoked to ensure very large $Q(I)$ values.

d. Rule 4 will prevent $Q(I)$ from becoming excessive during a short interval.

e. When the reconstructed signal exceeds the input signal, rule 5 will moderately reduce the step size. It will also produce the minimum output when no signal is present.

The nonlinear relationship between I and $Q(I)$ results in a rapidly increasing quantizing step size to cope with rapid increases in the signal rate of change. This significantly reduces slope overload by causing the quantizing step size to accurately follow the rms signal slope. Thus, the adaptive DM strategy is a form of nonuniform nonlinear processing similar to companding.

The Q function shown in Figure 2 has been developed with due consideration of the statistical properties of the speech signal. The maximum value of $Q(I)$ is approximately $2.5\sigma_x$, where σ_x is the rms value of the maximum speech signal. Other parameters such as maximum S/N_e, sampling frequency, and input power level have also been taken into consideration. Furthermore, if a curve to the right of the particular Q function shown in Figure 2 had been chosen, it would have resulted in greater slope overload noise, but less granular noise. On the other hand, a curve to the left would have resulted in more granular noise and less slope overload noise.

Digital feedback filter

It has been shown by deJager [3] that, for LDM systems, the reconstructed signal can be improved by using a double integrator in the feedback loop. However, a pure double-pole filter used in conjunction with a hard limiter (bistate comparator) can cause oscillation and lack stability under certain conditions [10]. To avoid this instability, a zero must be introduced into

the filter transfer function. The resulting transfer function can be written in the general form

$$F(s) = \frac{F_0(s + \omega_3)}{(s + \omega_1)(s + \omega_2)} \frac{1}{T} \tag{1}$$

where F_0 is a gain constant (scaling factor), $\omega_i = 2\pi f_i$, and T is the sampling period. The $1/T$ term in equation (1) eliminates variation in the response function gain caused by change in the sampling rate. Since the telephone signal has a frequency band of 300 to 3,500 Hz, f_1 and f_2 have been chosen to be 300 and 3,500 Hz, respectively. For maximizing S/N_e, the optimum value of f_3 has been found to lie between 4 and 6 kHz; hence, 5 kHz has been selected.

To examine the frequency response of the feedback filter, equation (1) is rewritten as

$$F(s) = \left(\frac{A}{s + 2\pi f_1} - \frac{B}{s + 2\pi f_2} \right) \frac{1}{T} \tag{2}$$

where

$$A = F_0 \frac{f_3 - f_1}{f_2 - f_1} \tag{3}$$

$$B = F_0 \frac{f_3 - f_2}{f_2 - f_1} \quad . \tag{4}$$

The corresponding expression in the z domain for digital filters is

$$F(z) = \frac{A}{1 - e^{-\omega_1 T} z^{-1}} - \frac{B}{1 - e^{-\omega_2 T} z^{-1}} \quad . \tag{5}$$

For $f_3 = 5$ kHz, and f_1 and f_2 as above,

$$A = +0.55078$$

$$B = +0.17578 \quad .$$

The resulting amplitude and phase characteristics of a filter using a sampling rate of 32 kHz are shown in Figure 3 over the frequency range of interest (0–5 kHz). The ratio of the 16-kHz folding frequency (due to the sampling rate) to the upper frequency of interest, i.e., 5 kHz, is sufficiently

high to result in negligible distortion of the frequency and phase characteristics.
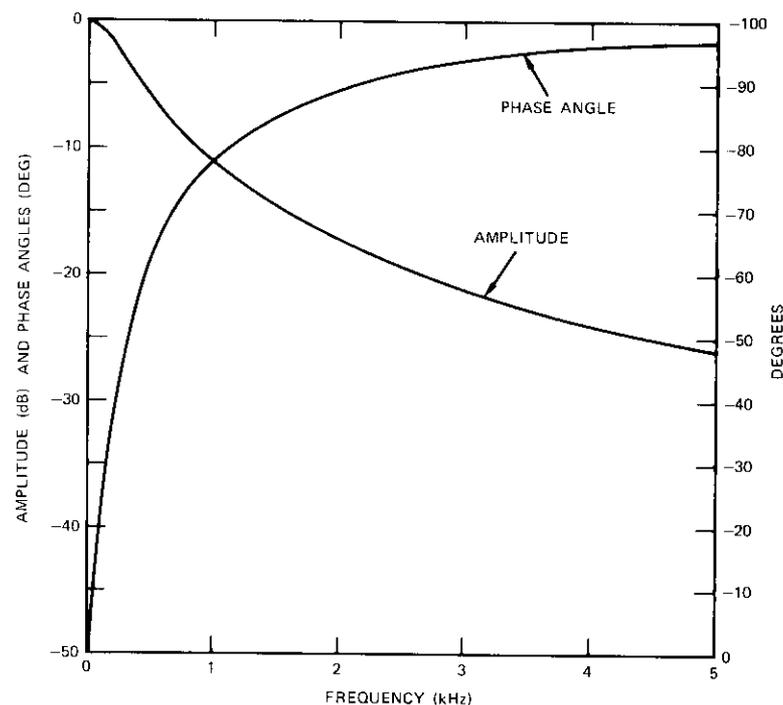
Figure 3. *Frequency Responses of 2-Pole, 1-Zero Feedback Linear Filter*

Performance criteria

The distortion introduced by the DM processing is assessed by calculating the rms difference between the original speech signal and the reconstructed speech signal. This must be done so that inconsequential phase differences between the original and reconstructed speech signal samples are essentially eliminated. To accomplish this in the evaluations reported here, the original speech was processed via the same low-pass filter characteristic that was used to smooth the output of the DM processor. As shown in Figure 4, which presents examples of the original speech, the direct delta modulator output, the low-passed original speech, and the low-passed delta modulator output, this resulted in relatively precise phase corre-

spondence between the reconstructed speech and the original low-passed speech. Figure 4 also shows the corresponding bit stream, the Q function index, I, and the step size, $Q(I)$, with bit sign.
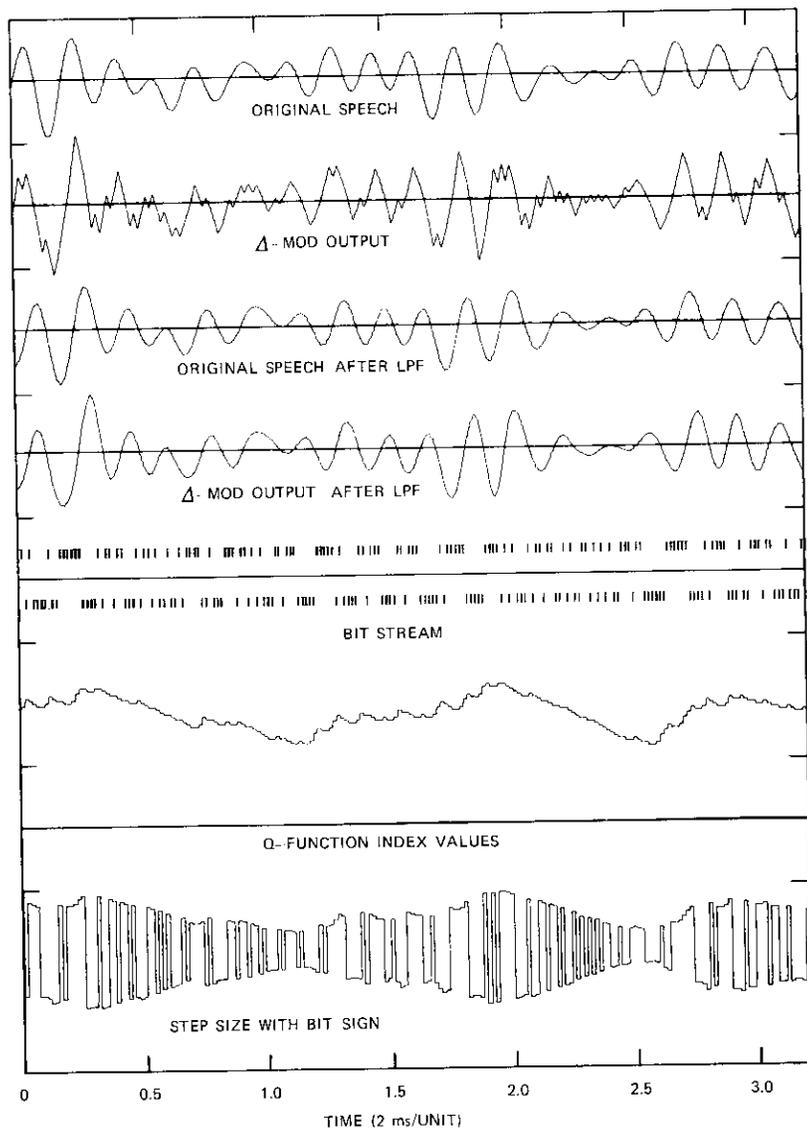
Figure 4. *Signal Waveforms in the ADM System*

In addition to this objective assessment, a subjective test was performed with about 100 listeners. Samples of analog telephone speech were compared with samples processed by the ADM system and by companded PCM. In the latter two cases, the speech samples were produced by programming the algorithms on the computer. The results are given in a subsequent section entitled "Simulation Results."

Speech sources

To minimize dependence on talker characteristics, long speech segments involving four different talkers, two women (JW and AS) and two men (FC and SC), were produced. Each talker spoke into a telephone handset connected to an artificial line and carrier facility. At the output of the carrier system, the signal was recorded to simulate a typical interface point with an earth station before transmission via satellite. Each talker read the same five phonetically balanced sentences in the following order:

- *a.* She caught Sue's fuchsia scarf.
- *b.* See the old pig's azure foot.
- *c.* Joe took father's shoe bench out.
- *d.* Ali Baba knew how men buy lime.
- *e.* He called time and time again.

The speech signals were sampled and digitized. The sampling frequency was 32 kHz to allow for preparation of the 32-kbit transmission rate of the delta modulator. The total length of the speech source was approximately 48 seconds, yielding more than 1.5 million samples. Some pertinent statistics of these speech samples are listed in Table 1; the composite power spectrum distribution of all samples is shown in Figure 5.

TABLE 1. STATISTICS OF SPEECH SOURCES

Talker	rms Value	Peak/rms (dB)	No. of Samples	Remarks
JW	209.74	13.86	430,080	Female
AS	170.61	16.17	360,448	Female
FC	180.45	15.74	344,064	Male
SC	193.71	15.47	409,600	Male
Overall	190.41	15.27*	1.54 Million	

* The largest value for the four talkers, 1,024, was used as the peak value.

averaging over the speech samples produced by two men and two women are shown in Figure 6 for levels ranging from -50 to 0 dBm0. Also shown for comparison are the performances of the 5-bit-per-sample A-law and μ-law PCM [11], [12].
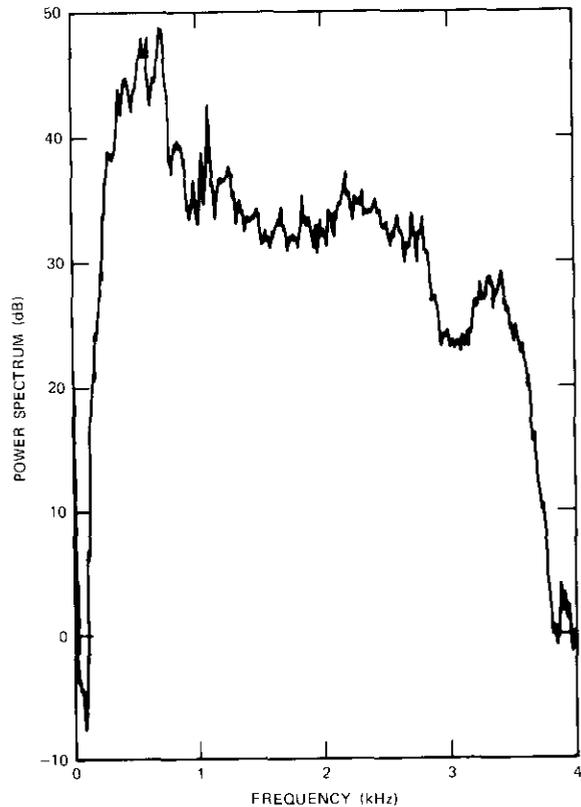
Figure 5. *Composite Spectrum of Speech Sources*

Simulation results

Objective evaluation of performance

As indicated previously, the distortion is estimated from the rms difference between the low-passed output of the delta modulator and the original speech processed through the same type of low-pass filter. It should be noted that the performance characteristic curves presented in this paper are the S/N_e values obtained by averaging the speech segments of all four talkers.

The rms speech power averaged over the speech samples is divided by the distortion determined above to yield S/N_e. The results obtained by

Figure 6. *Performance Characteristic Curves for ADM and PCM Systems*

The performance of the ADM system is similar to that of a companded PCM system [11], [12]. The optimum values are achieved at an input level in the neighborhood of -12.5 dBm0. The dynamic range of the overall speech input using 3-dB S/N_e degradation points from optimum is approximately 35 dB. The average computed S/N_e value for active speech is about 18 dB. It should be noted that an 18-dB S/N_e ratio is equivalent to the performance obtained with a 5-bit-per-sample A-law PCM system.

Subjective evaluation (ADM)

Subjective listening tests were conducted to assess the performance of the ADM system described in the preceding subsection. The ADM system was ranked relative to an analog channel having a continuously active

speech-to-Gaussian-noise ratio of 38 dB and relative to 5- and 7-bit error-free A-law companded PCM. The results indicated that the ADM system was equivalent in performance to both the 5-bit PCM processed speech and the 38-dB S/N analog speech, thus confirming a strong correlation between the objective evaluation and the subjective assessment.

Impact of bit errors

Bit-error rates in the range of 10^{-6} to 10^{-3} produced by a Gaussian noise generator were introduced in the transmission channel. For each error rate, three different error patterns were simulated. Quality evaluations were made by a jury of four communications engineers. The impact of an error rate smaller than 10^{-5} was not perceptible through either a high-fidelity loudspeaker or a telephone set. As the bit-error rate increased to 0.5×10^{-4}, the effect became noticeable when the samples were played back through a loudspeaker; amplitude modulation due to bit errors was being heard. However, the quality of the speech was still good. Moreover, this modulation phenomenon was hardly perceptible for bit-error rates up to 10^{-4} when a telephone handset was used. As the bit-error rate increased to 0.5×10^{-3}, the quality of speech played back through both a loudspeaker and a telephone handset deteriorated, but the intelligibility was still good. For a bit-error rate greater than 10^{-3}, the speech became somewhat amplitude modulated and noisy, and intelligibility was judged as fair.

Conclusions

A new ADM strategy, incorporating a nonlinear step size Q function, has been presented for the reconstruction of high-quality speech. An extensive computer simulation of ADM transmitted at a 32-kbps rate was made with the speech of two women and two men as input. Each talker read the same phonetically balanced sentences. The total length of the input source was approximately 48 seconds.

As an objective measure of the performance, the signal-to-distortion error was computed for the ADM system over a 50-dB dynamic range of signal level. The ADM performance characteristic curve exhibited a behavior similar to that of 5-bit companded PCM.

A subjective rank order comparison test was also conducted with approximately 100 listeners. The subjects indicated that the quality of the error-free reconstructed speech was equivalent to that of 5-bit companded PCM and to that of an analog channel having a continuously active speech-signal-to-Gaussian-noise ratio of 38 dB.

An investigation of the effect of bit-error rate in the range of 10^{-3} to 10^{-6} indicated that no effect is perceptible until the bit-error rate increases to 10^{-4}. As the bit-error rate reaches a value of 0.5×10^{-3}, the speech quality starts to deteriorate in the form of volume modulation, although good intelligibility is maintained. For a bit-error rate larger than 10^{-3}, the speech becomes noisy and increasingly modulated in volume, but still very intelligible.

The ADM incorporating the new slope control logic demonstrates that transmission at 32 kbps can provide good-quality telephone communications and that good performance can be expected at bit-error rates as high as 10^{-3}.

Acknowledgment

References

[1] E. M. Deloraine, S. Van Miero, and B. Derjavitch, "Methode et Système de Transmission par Impulsions," French Patent 932140, August 1946.

[2] L. J. Libois, "Un Nouveau Procédé de Modulation Codée la Modulation en Delta," *L'Onde Electrique,* Vol. 32, June 1952, pp. 26–31.

[3] F. deJager, "Delta Modulation, A Method of PCM Transmission Using the 1-Unit Code," Phillips Research Report 7, 1952, pp. 442–466.

[4] C. C. Cutler, "Differential Quantization of Communications Signals," U.S. Patent 2605361, July 29, 1952.

[5] J. E. Abate, "Linear and Adaptive Delta Modulation," *Proc. IEEE,* Vol. 55, No. 3, March 1967, pp. 298–308.

[6] J. A. Greefkes and F. deJager, "Continuous Delta Modulation," Phillips Research Report, Vol. 32, No. 2, 1968, pp. 233–246.

[7] J. A. Greefkes and K. Riemeus, "Code Modulation with Digitally Controlled Companding for Speech Transmission," Phillips Technical Report, Vol. 31, No. 11–12, 1970, pp. 335–353.

[8] J. A. Greefkes, "A Digitally Companded Delta Modulation Modem for Speech Transmission," International Conference on Communications, June 1970, *Proc.,* pp. 7-33–7-48.

[9] N. S. Jayant, "Adaptive Delta Modulation with a One-bit Memory," *Bell System Technical Journal,* Vol. 49, No. 3, March 1970, pp. 321–342.

[10] J. E. Gibson, *Nonlinear Automatic Control*, New York: McGraw-Hill Book Company, 1963, pp. 342–438.

[11] Members of the Technical Staff, Bell Laboratories, Inc., *Transmission Systems for Communications*, Fourth Edition, Winston-Salem: Western Electric Company, 1970.

[12] J. C. Su and S. J. Campanella, "Quantization Noise and Distortion Power Spectra of PCM Systems," *COMSAT Technical Review*, Vol. 4, No. 2, Fall 1974, pp. 393–410.

James C. Su received his B.S.E.E. from Taiwan University and his Ph.D. in electrical engineering from the University of Pennsylvania. From 1963 to 1967, while at the Moore School of Electrical Engineering, he worked on functional series representation of time-varying nonlinear systems with arbitrary input.

As a member of the technical staff at COMSAT *Laboratories since 1967, Dr. Su has been engaged in digital signal processing techniques for speech transmission. He has developed a system for regenerating the pointing data of synchronous satellites and a nonlinear phase-locked loop for fast acquisition of the signal.*

H. G. Suyderhoud was born in Borneo, Indonesia. He received the Ingenieur degree from Delft Technological University in the Netherlands in 1955, and the M.S. degree in applied mathematical statistics from Rutgers, New Brunswick, N.J., in 1966. Joining COMSAT *Laboratories in 1968, he worked extensively on research and development of echo cancellation techniques. Presently Manager of the Signal Processing Department of the Communications Processing Laboratory, he is responsible for signal processing research and development for satellite application using both digital and analog techniques.*
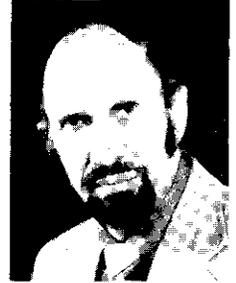
Prior to joining COMSAT, *Mr. Suyderhoud worked for 11 years at Bell Telephone Laboratories on exchange area transmission engineering, toll carriers, and picturephone. He is a member of the IEEE and the American Statistical Association.*

S. J. Campanella received a B.S.E.E. from the Catholic University of America, an M.S.E.E. from the University of Maryland, and a Ph.D. in electrical engineering from the Catholic University of America. Before joining COMSAT, *he was Manager of the Electronics Research Center at Melpar, Inc.*

He is presently Director of the Communications Processing Laboratory of COMSAT *Laboratories where he is responsible for the technical orientation of the efforts pursued in the Signal Processing and Multiplexing, Image Processing, Digital Control, and Modulation Techniques Departments.*

Dr. Campanella is a member of AIAA, AAAS, IEEE, Sigma Xi, Phi Eta Sigma, and the Acoustical Society of America. He is also a member of the board of directors of EASCON.

Phase ambiguity resolution in a 4-phase PSK modulation system with forward-error-correcting convolutional codes

Y. Tsuji

(Manuscript received March 19, 1976)

Abstract

When 4-phase PSK modulation and coherent detection are used, the phase ambiguity of a reference carrier at a receive side must be resolved. The phase ambiguity can be eliminated by using a differential coding technique or a synchronization word. However, when forward-error-correcting (FEC) codes are used, the conventional differential coding method implemented inside an FEC encoder and decoder pair (codec) has a drawback. That is, one erroneously detected phase will cause two consecutive false symbols (a burst error) even if the next phase is received correctly. In this case, the random error-correcting capability of the error decoder must be doubled to maintain the same performance. On the other hand, the use of a synchronization word will sometimes result in the reduction of transmission efficiency.

This paper investigates two methods of resolving phase ambiguity by differential coding which are compatible with the application of an FEC code. The first uses differential coding inside an FEC codec in conjunction with an interleaving technique. The second utilizes a syndrome pattern to resolve the primary phase ambiguity followed by differential coding outside an FEC codec to resolve the remaining ambiguity.

Introduction

Application of forward-error-correcting (FEC) codes to digital satellite communications systems such as single-channel-per-carrier (SCPC) [1], digital television communications (DITEC) [2], time-division multiple-access (TDMA) [3], and frequency-division multiple-access/phase shift keyed/digital speech interpolation (FDMA/PSK/DSI) [4] systems has been considered. A 4-phase PSK modulation system is usually adopted in these digital systems. When 4-phase PSK modulation and coherent detection are used, the recovered carrier phase ambiguity must be resolved on the basis of the received PSK signal. In addition, the phase rotation direction ambiguity due to frequency conversion must be resolved. The recovered carrier phase ambiguity can be eliminated by differential coding, as in the 4-phase PSK modem in the TDMA system [5]. Use of a synchronization word, as in the 4-phase PSK modem in the SPADE system, makes it possible to resolve both recovered carrier phase and phase rotation direction ambiguity [6].

When FEC codes are used, resolution of phase ambiguity becomes more involved. For example, with the conventional differential coding method implemented inside an FEC codec, one erroneously detected phase will cause two consecutive false symbols (a burst error) even if the next phase is received correctly [3], [7]. In this case, the random error-correcting capability of an error decoder must be doubled to maintain the same performance. On the other hand, the use of a synchronization word will sometimes result in the reduction of transmission efficiency. One method of overcoming this problem, which is peculiar to conventional differential coding, is to develop a new FEC code which corrects the pairwise error pattern which is most likely to appear in a differentially encoded coherent 4-phase PSK system [8].

This paper describes two alternative methods of resolving phase ambiguity which are compatible with the application of a conventional FEC code. The first uses differential coding inside an FEC codec, but it provides a means of breaking up the burst errors peculiar to conventional differential coding. This method still utilizes differential coding inside an FEC codec, but burst errors after differential decoding are changed into random errors at the input of an FEC decoder.

The second method is phase ambiguity resolution by differential coding used outside an FEC codec. This method utilizes a syndrome pattern in the error decoder followed by a differential decoder to resolve the phase ambiguity. Through observation of the syndrome pattern, the various

kinds of phase ambiguities degenerate into some phase ambiguities which cannot be distinguished by the syndrome pattern. To resolve these remaining ambiguities, a differential decoder is used after error correction. Reference 7 shows the configuration for recovered carrier phase ambiguity using a transparent coder.

This paper includes techniques for resolving both recovered carrier phase ambiguity and combined recovered carrier phase and phase rotation direction ambiguity when either a transparent or a nontransparent FEC coder is used. These studies are based on the following assumptions:

a. a 4-phase PSK modulation system,
b. systematic FEC convolutional codes $(n, n-1)$ with n even,
c. a synchronizer using a trial and error method in the decoder.

Classification of phase ambiguity

Phase rotation direction ambiguity

It is assumed that a PSK signal of intermediate frequency f_c is up-converted by the transmit local frequency $f_m - f_c$ to yield the PSK microwave frequency f_m. At the receive side, f_m is down-converted by the receiver local frequency to obtain the PSK signal at f_c. In this case, two receive local frequencies can be considered, as shown in Figure 1. That is, the receive local frequency may be selected so that it is higher than the receive microwave frequency by f_c, or it may be selected so that it is lower than the receive microwave frequency by f_c.

The resultant phase rotation direction will be reversed relative to the original phase rotation direction depending on the receive local frequency which is selected. This phenomenon is called phase rotation direction ambiguity. This ambiguity can usually be avoided by specifying the relationship between the transmit local frequency and the receive local frequency. Hence, it will not occur in PCM/PSK/TDMA and FDMA/PSK/DSI systems, for example. However, it is difficult to avoid in SPADE and SCPC systems.

The SPADE system has about 800 carriers in one 36-MHz satellite transponder. These carriers, located over a 36-MHz bandwidth from 98 to 134 MHz, are classified as lower band carriers (LBCs) and upper band carriers (UPCs), as shown in Figure 2. When a certain channel uses one of the LBCs, its return channel must use a UBC, which will result in a difference of 18 MHz between the carriers. For example, if channel i of station 1 is using $F(i)$, then the return channel j of station 2 uses $F(j)$, where $F(X)$
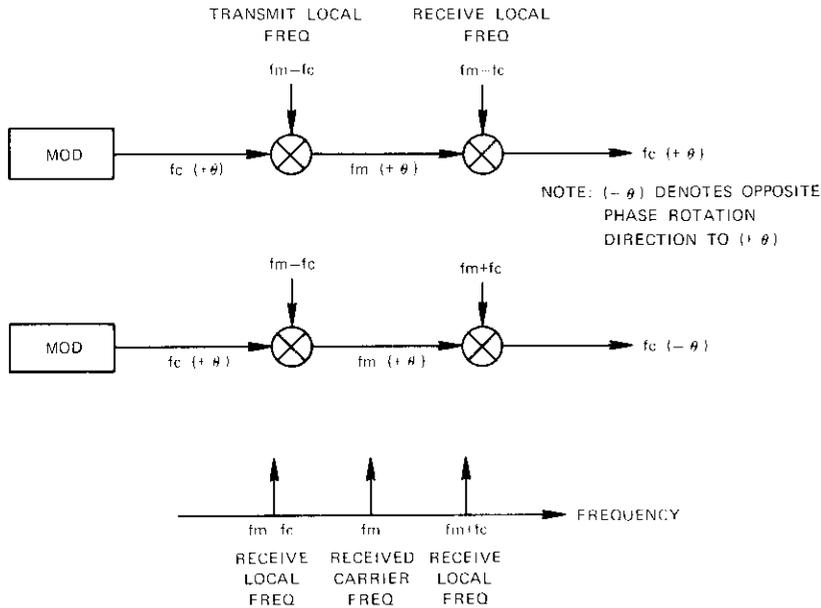
Figure 1. *Two Kinds of Receive Local Frequency*
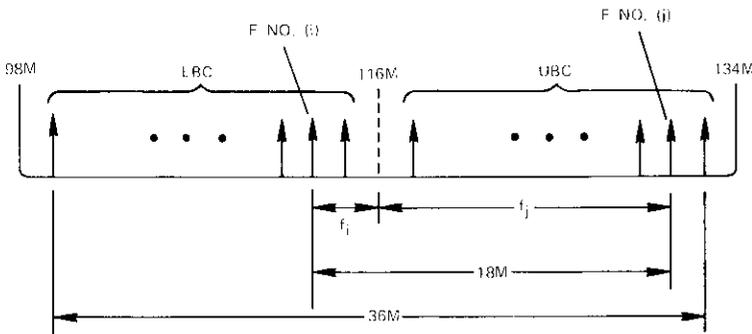
Figure 2. *Arrangement of Carrier Frequencies in the SPADE System*

denotes the frequency of channel X, and f_i and f_j are the frequency differences between 116 MHz and $F(i)$ and 116 MHz and $F(j)$, respectively. Even if the specified sideband is selected in the microwave frequency conversion and phase rotation direction ambiguity does not occur in this process, the resultant IF will exhibit phase rotation direction ambiguity,
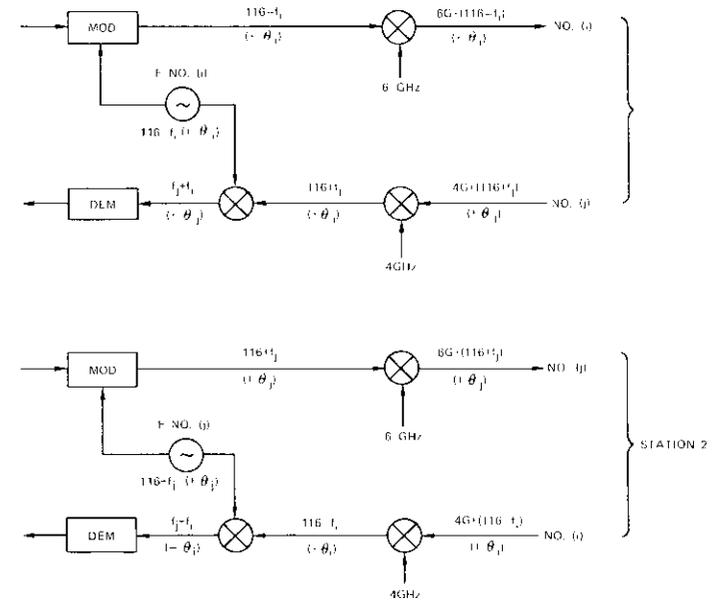
Figure 3. *Phase Rotation Direction Ambiguity in the SPADE System*

as shown in Figure 3, where $+\theta_i$ and $-\theta_i$ represent the normal and opposite phase rotations originating from channel i. In the SCPC system, which is a modification of the SPADE system, similar phase rotation direction ambiguity can be observed depending on the portion of the usable transponder bandwidth from which the received PSK signal comes.

Recovered carrier phase ambiguity

Because the locked-in phase may be any one of four possible phases, ambiguity of the fourth degree results. Figure 4 shows the variation of the demodulated signal with the locked-in phase of the carrier recovery circuit.
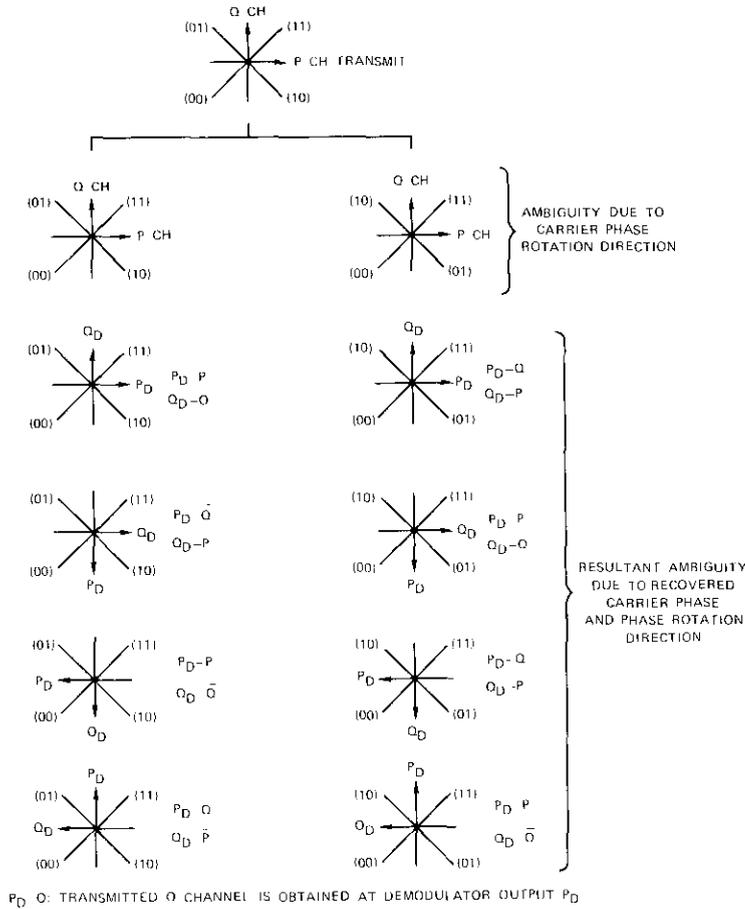
Figure 4. *Recovered Carrier Phase and Phase Rotation Direction Ambiguity*

Phase ambiguity resolution by differential coding inside an FEC codec

Recovered carrier phase ambiguity is resolved by differential coding inside an FEC codec in systems such as TDMA. The direction of phase rotation should be given to the differential decoder in transmission systems in which phase rotation direction ambiguity results from the frequency conversion system, as in the SCPC and SPADE systems.

PSK modulation with differential coding transmits information by using the amount of phase shift, and the information is received by comparing the difference between two consecutive phases. Hence, in the differential coding scheme, one erroneously detected phase will cause two consecutive false symbols (a burst error) even if the next phase is received correctly. Some means of breaking up these burst errors is necessary to permit the error decoder to work properly when phase ambiguity resolution by differential coding is used inside an FEC codec.

One advantage of this method is that it does not require synchronization time to eliminate the phase ambiguity for the error decoder. Thus, it can be used in either a burst mode signal or a continuous mode signal, although the bit-error rate equivalent to the decoder input is twice that without differential coding. Methods of breaking up the consecutive double errors which are associated with differential coding are described in the following subsections.

Columnwise writing and rowwise reading method

As shown in Figure 5, on the transmit side a train of input signals is applied to the FEC convolutional encoder. The encoder output is written columnwise in the buffer, and the signal stored in the buffer is read rowwise. These steps are reversed on the receive side. If the numbers of buffer rows and columns are selected to accommodate the constraint length of the convolutional codes, the burst error caused by differential coding will be converted into a random error in the receive buffer output.
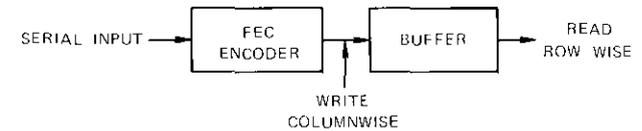
Figure 5. *Columnwise Writing and Rowwise Reading Method*

Symbol interleaving method

The symbol interleaving method is intended to eliminate the adverse effect of differential coding by encoding even and odd symbols independently so that a burst error is changed to a random error, hence justifying the use of the random error-correcting convolutional codes. Figure 6
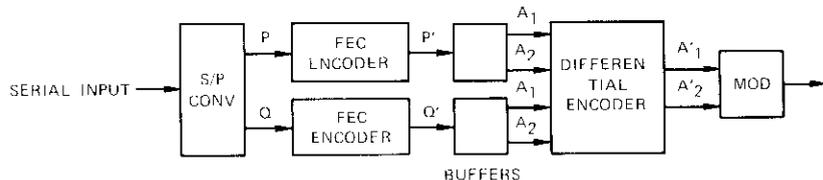
Figure 6. *Method of Applying Symbol Interleaving*

is a transmit side block diagram showing the application of this method to 4-phase PSK modulation with differential coding. Figure 7 shows
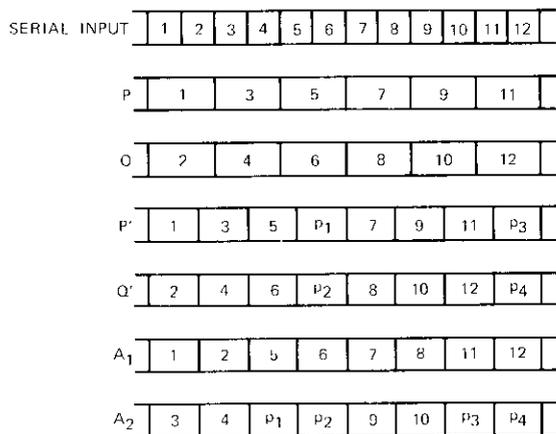
Figure 7. *Example of Bit Correspondence for a (4, 3) Code with Symbol Interleaving*

individual bit correspondence when the (4, 3) code is used in 4-phase PSK modulation with differential coding.

As shown in Figure 7, the symbol interleaving method comprises three steps:

a. STEP 1: An input signal train is changed into parallel signals P and Q.

b. STEP 2: Parallel signals P and Q are separately applied to the two FEC encoders.

c. STEP 3: Encoder outputs P' and Q' are applied to the buffer memories, and the bits corresponding to a symbol are alternately read out from two buffer memories as independent symbols.

These steps are reversed at the receive side. This method does not degrade the capability of FEC codes with simplified hardware.

M-symbol differential coding method

Differential coding in phase modulation refers to the transmission of the difference between the present phase and the previous phase, which corresponds to the information to be transmitted. Hence, the receiving side can regenerate the received information from the phase shift between two consecutive phases. In this technique, if the nth symbol on the transmission path is erroneous, then the $(n + 1)$th symbol is also erroneous.

For M-symbol differential coding in phase modulation, a natural extension of conventional differential coding, the difference between the present phase and the Mth previous phase corresponds to the information to be transmitted. Hence, the receiving side can recover the received information from the difference between one phase which has been received and that received M symbols previously. If the nth symbol on the transmission path is erroneous, then the $(n + M)$th symbol will also be erroneous. If M is determined to be sufficiently longer than the constraint length, a pairwise error in one constraint length can be avoided.

Phase ambiguity resolution by differential coding outside an FEC codec

It should be noted that phase ambiguity resolution by differential coding outside an FEC codec is to be used together with the synchronizer circuit of the FEC decoder shown in Figure 8. The synchronizer continually ob-
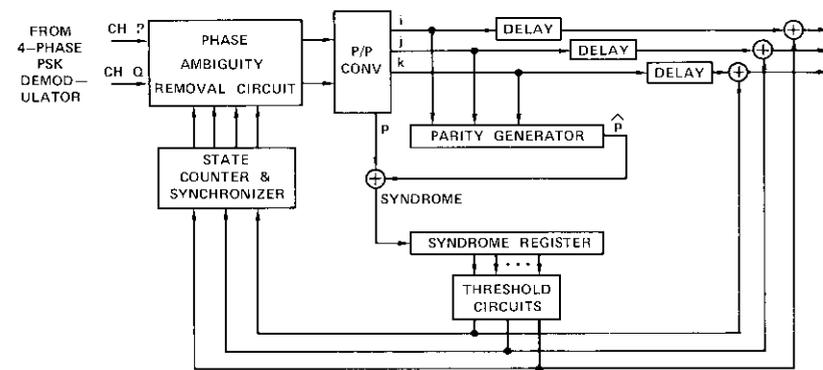
Figure 8. *(4, 3) Convolutional Code Decoder with Phase Ambiguity Resolving Circuit*

serves the syndrome pattern and compares the number of corrections with a predetermined threshold value. Once the threshold is exceeded, the state of the phase ambiguity resolving circuit is changed. If the new state is not correct, another state is selected until the proper state for ambiguity resolution is found. Differential coding is used to resolve the particular phase ambiguity which the synchronizer of the decoder cannot resolve because of the nature of convolutional codes.

Since the differential coding is used outside an FEC codec, the error-correcting decoder does not encounter the double-error phenomenon and only the bit-error rate of the error-correcting decoder output (improved bit-error rate) is doubled. Doubling the decoder output error rate results in a smaller E_b/N_o loss than doubling the input error rate because the curve of error rate versus E_b/N_o is steeper for the output. Thus, in terms of bit-error rate improvement characteristics, this method is superior to phase ambiguity resolution by differential coding inside an FEC codec.

The primary disadvantage of this method is that a relatively long time is required to resolve the phase ambiguity. Therefore, it may not be suitable for application to burst mode operation, as in a TDMA system. Since this method exploits the inherent characteristics of FEC convolutional codes, two kinds of subgenerators, $g(j)$, which define $(n, n-1)$ FEC convolutional codes, are assumed (see Figure 9). A transparent subgenerator, $g(j)$ $= [g_0(j), g_1(j), \ldots, g_{N-1}(j)]$, is defined as

$$\sum_{i=0}^{N-1} g_i(j) = 2l + 1$$

whereas a nontransparent subgenerator, $g(j) = [g_0(j), g_1(j), \ldots, g_{N-1}(j)]$, is defined as

$$\sum_{i=0}^{N-1} g_i(j) = 2l$$

where
l = integer
N = constraint length in blocks

$j = 1, 2, \ldots, n-1$.

If $g_i(j) = 0$, there is no connection to the ith register from the jth information stream, while if $g_i(j) = 1$, there is a connection to the ith register from the jth information stream. Codes which are generated are designated
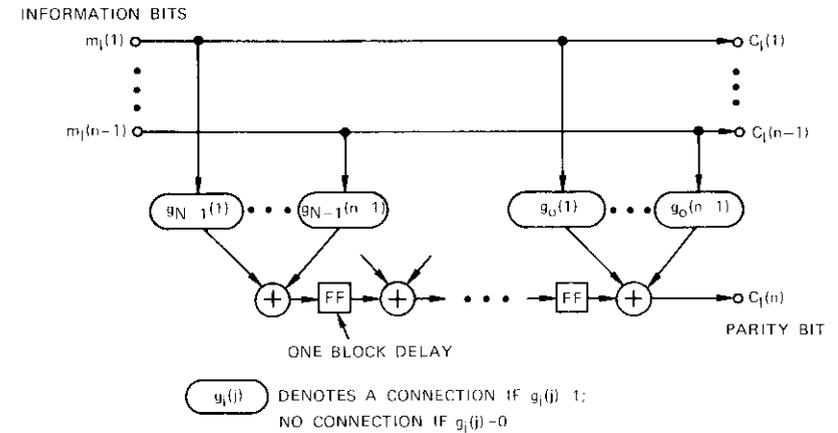
Figure 9. $(n, n-1)$ Convolutional Encoder and Transmit Format

as transparent or nontransparent depending on the generator which is used.

As described previously, two kinds of ambiguity states can be considered. One consists of recovered carrier phase ambiguity only, and the other consists of both covered carrier phase ambiguity and phase rotation direction ambiguity.

Recovered carrier phase ambiguity

PCM/PSK/TDMA and FDMA/PSK/DSI satellite communications systems are examples of systems which experience only recovered carrier phase ambiguity because the same sequence of frequency conversion can always be expected. In this case, there are four kinds of phase ambiguity:

Transmit		Receive		
P	P	$\bar{P}$	Q	$\bar{Q}$
Q	Q	$\bar{Q}$	$\bar{P}$	P

Therefore, Figure 10 can be assumed to be the carrier phase ambiguity resolution circuit. In this logic diagram, it may be assumed that the $(P, \overline{Q})$ state may remain unresolved due to the nature of the decoder, since $(Q, \overline{P})$, $(\overline{Q}, P)$ will be changed to (P, Q) or $(\overline{P}, Q)$ by the carrier phase ambiguity resolution circuit. It is also assumed that the decoder will not work properly in the $(Q, \overline{P})$ or $(\overline{Q}, P)$ state.
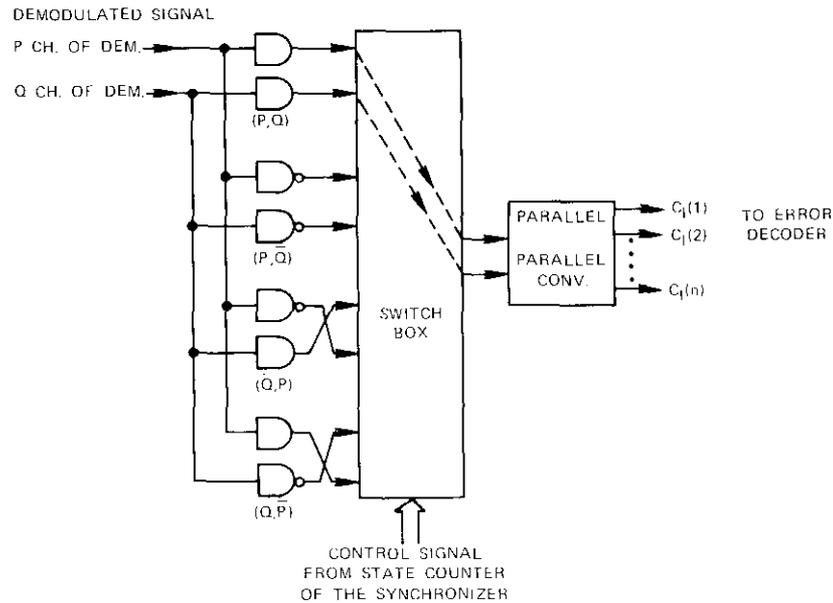
Figure 10. *Logic Diagram of Phase Ambiguity Resolution Circuit for Recovered Carrier Phase Ambiguity*

APPLICATION OF TRANSPARENT CODES

As shown in Figure 9, the n transmitted digits of the lth code block are expressed as follows:

$$C_l(j) = m_l(j), \qquad j = 1, 2, \ldots, n - 1$$

$$C_l(n) = \sum_{i=0}^{N-1} \sum_{j=1}^{n-1} m_{l-i}(j) \, g_i(j) \tag{1}$$

where $m_l(j) = n - 1$ message digits of the lth block

$$(j = 1, 2, \ldots, n - 1)$$

$C_l(n)$ = parity bit of the lth block.

When the receive ambiguity state is $(\overline{P}, \overline{Q})$, $m_{l-i}(j)$ becomes $\overline{m_{l-i}(j)}$. Therefore, estimated parity bit $\hat{C}_l(n)$, which is generated in the decoder by the same algorithm used in the encoder, is expressed as follows:

$$\hat{C}_l(n) = \sum_{i=0}^{N-1} \sum_{j=1}^{n-1} \overline{m_{l-i}(j)} \, g_i(j)$$

$$= \overline{C}_l(n) \tag{2}$$

since $\sum_{i=0}^{N-1} g_i(j) = 2l + 1$ for $j = 1, 2, \ldots, n - 1$ and n even.

Since estimated parity bit $\hat{C}_l(n)$ coincides with received parity bit $\overline{C}_l(n)$ in the case of $(\overline{P}, \overline{Q})$, the syndrome pattern will be all "0's" and the synchronizer will assume that the phase ambiguity has been resolved. In this case, if there is no error in the transmission line, $\overline{m_l}$ (the inverse of the transmitted message) will be obtained from the decoder output so that the binary differential encoder and decoder shown in Figure 11 will be necessary [7].
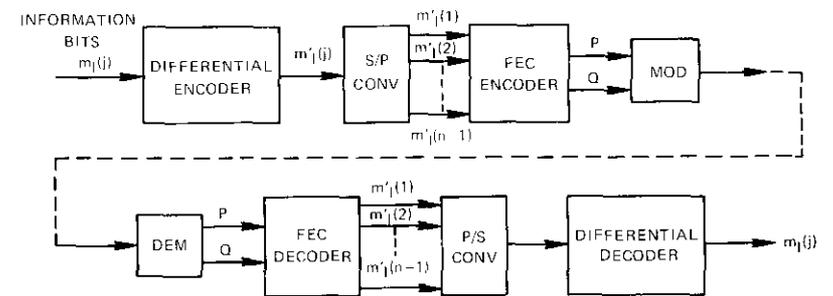
NOTE: $m'_l(j)$ DENOTES DIFFERENTIALLY ENCODED $m_l(j)$
AMBIGUITY: RECOVERED CARRIER PHASE ONLY
CODE: TRANSPARENT

Figure 11. *Phase Ambiguity Resolution by Differential Coding Outside an FEC Codec (case I)*

APPLICATION OF NONTRANSPARENT CODES

When the receive ambiguity state is $(\bar{P}, \bar{Q})$, the estimated parity bit $\hat{C}_l(n)$, which is generated in the decoder by the same algorithm used in the encoder, is expressed as follows:

$$\hat{C}_l(n) = \sum_{i=0}^{N-1} \sum_{j=1}^{n-1} \overline{m_{l-i}(j)} \, g_i(j)$$

$$= C_l(n) \tag{3}$$

since $\sum_{i=0}^{N-1} g_i(j) = 2l$ for $j = 1, 2, \ldots, n - 1$. Since estimated parity bit $\hat{C}_l(n)$ does not coincide with received parity bit $\bar{C}_l(n)$ in the case of $(\bar{P}, \bar{Q})$, the syndrome pattern becomes all "1's" and the synchronization process continues until the correct ambiguity resolution is found. Therefore, it is unnecessary to adopt differential encoding, since the synchronizer itself will automatically resolve the phase ambiguity.

Recovered carrier phase and phase rotation direction ambiguity

Because of the frequency conversions encountered in SPADE and SCPC satellite communications systems, these systems experience both recovered carrier phase and phase rotation direction ambiguity. In this case, there are eight kinds of phase ambiguities:

Transmit		Receive						
P	P	$\bar{P}$	Q	$\bar{Q}$	Q	$\bar{Q}$	$\bar{P}$	P
Q	Q	$\bar{Q}$	$\bar{P}$	P	P	$\bar{P}$	Q	$\bar{Q}$

Therefore, Figure 12 can be considered to be the carrier phase ambiguity resolution circuit. In this logic diagram, it may be assumed that states $(P, \bar{Q})$, $(\bar{P}, Q)$, and $(\bar{P}, \bar{Q})$ remain unresolved due to the nature of the decoder, since $(Q, \bar{P})$, $(\bar{Q}, P)$, (Q, P), and $(\bar{Q}, \bar{P})$ will be changed to state (P, Q), $(P, \bar{Q})$, $(\bar{P}, Q)$, or $(\bar{P}, \bar{Q})$ by the phase ambiguity resolution circuit. It is also assumed that the decoder will not work properly in the $(Q, \bar{P})$, $(\bar{Q}, P)$, (Q, P), and $(\bar{Q}, \bar{P})$ states.

APPLICATION OF TRANSPARENT CODES

For state $(\bar{P}, \bar{Q})$, $\overline{m_l}$ will be obtained from the decoder output, as shown in the case of transparent codes for recovered carrier phase ambiguity
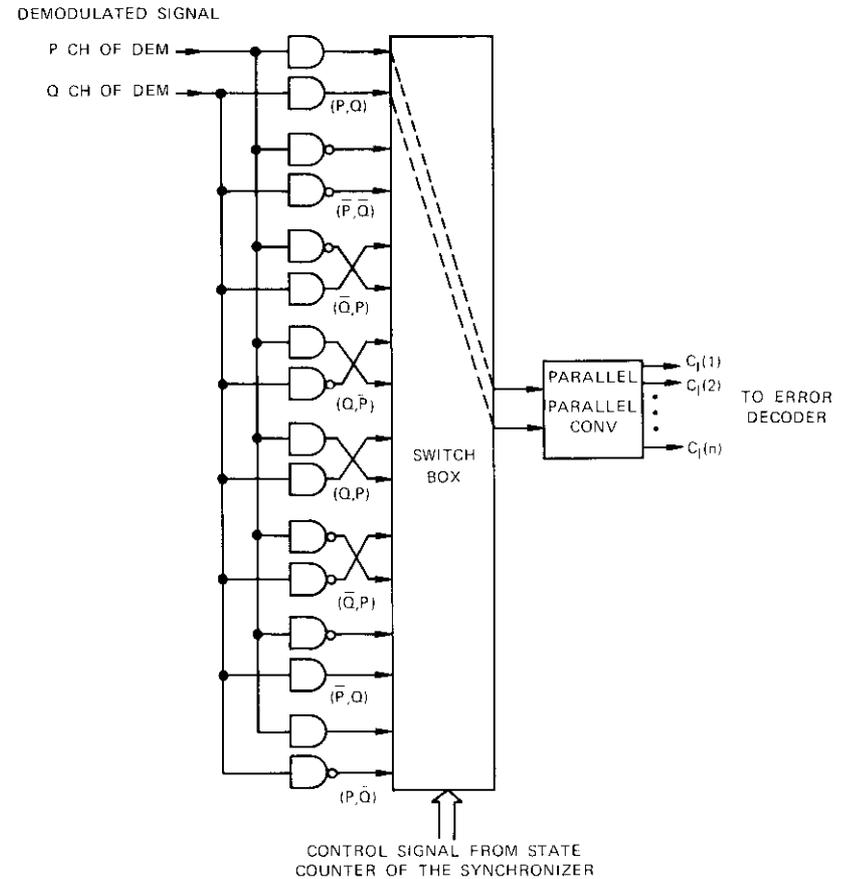
Figure 12. *Logic Diagram of the Phase Ambiguity Resolution Circuit for Both Recovered Carrier Phase and Phase Rotation Direction Ambiguity*

only. When the ambiguity state is $(\bar{P}, Q)$, $m_{l-i}(j)$ $(j = 1, 2, \ldots, n - 1)$ will be changed as follows at the receive side:

$$\overline{m_{l-i}(j)}: \quad j \leq \frac{n}{2}$$

$$m_{l-i}(j): \quad \frac{n}{2} < j \leq n - 1 \quad .$$

Therefore, estimated parity bit $\hat{C}_i(n)$, which is generated in the decoder, is expressed as

$$\hat{C}_i(n) = \sum_{i=0}^{N-1} \sum_{j=1}^{n/2} \overline{m_{L-i}}(j)\, g_i(j) + \sum_{i=0}^{N-1} \sum_{j=(n/2)+1}^{n-1} m_{L-i}(j)\, g_i(j)$$

$$= C_i(n), \quad \frac{n}{2} \text{ even}$$

$$= \overline{C}_i(n), \quad \frac{n}{2} \text{ odd} \quad . \tag{4}$$

Since estimated parity bit $\hat{C}_i(n)$ coincides with received parity bit $C_i(n)$ in state $(\overline{P}, Q)$ for $n/2$ even, $\overline{m_i}(j)$ $(j = 1, 2, \ldots, n/2)$, $m_i(j)$ $[j = (n/2) + 1, \ldots, n - 1]$ will be obtained from the decoder output. Estimated parity bit $\hat{C}_i(n)$ does not coincide with received parity bit $C_i(n)$ for $n/2$ odd. Therefore, the synchronizer itself can resolve phase ambiguity state $(\overline{P}, Q)$ for $n/2$ odd.

When the ambiguity state is $(P, \overline{Q})$, $m_{L-i}(j)$ will be changed as follows at the receive side:

$$m_{L-i}(j): \quad j \leq \frac{n}{2}$$

$$\overline{m_{L-i}}(j): \quad \frac{n}{2} < j \leq n - 1 \quad .$$

Hence, estimated parity bit $\hat{C}_i(n)$, which is generated in the decoder, is expressed as

$$\hat{C}_i(n) = \sum_{i=0}^{N-1} \sum_{j=1}^{n/2} m_{L-i}(j)\, g_i(j) + \sum_{i=0}^{N-1} \sum_{j=(n/2)+1}^{n-1} \overline{m_{L-i}}(j)\, g_i(j)$$

$$= \overline{C}_i(n), \quad \frac{n}{2} \text{ even}$$

$$= C_i(n), \quad \frac{n}{2} \text{ odd} \quad . \tag{5}$$

Since estimated parity bit $\hat{C}_i(n)$ coincides with received parity bit $\overline{C}_i(n)$ in state $(P, \overline{Q})$ for $n/2$ even, $m_i(j)$ $(j = 1, 2, \ldots, n/2)$, $\overline{m_i}(j)$ $[j = (n/2)$

$+ 1, \ldots, n - 1]$ will be obtained from the decoder output. For $n/2$ odd, the synchronizer itself can resolve phase ambiguity state $(P, \overline{Q})$. Therefore, to resolve the kinds of ambiguity described above for $n/2$ even, sequences $m_i(j)$ $(j = 1, 2, \ldots, n/2)$ and $m_i(j)$ $[j = (n/2) + 1, \ldots, n - 1]$ must be differentially encoded and decoded independently, as shown in Figure 13.

$m'_I(j)$: DIFFERENTIALLY ENCODED $m_I(j)$

AMBIGUITY: RECOVERED CARRIER PHASE AND PHASE ROTATION DIRECTION

CODE: TRANSPARENT
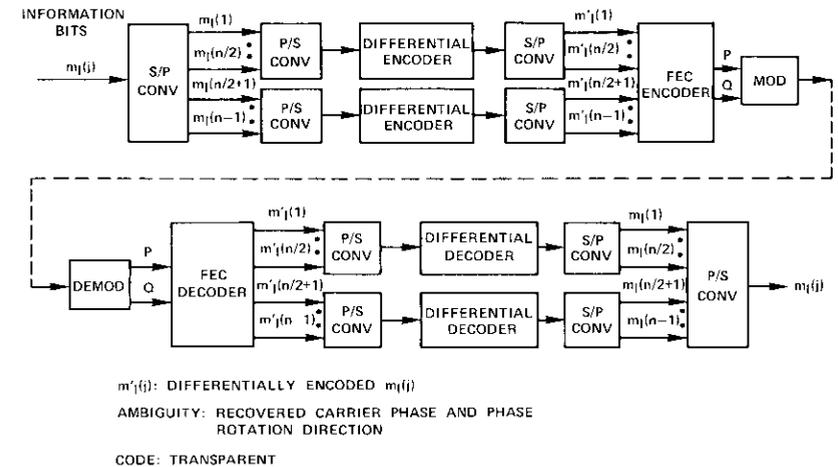
Figure 13. *Phase Ambiguity Resolution by Differential Coding Outside an FEC Codec (case 2)*

For $n/2$ odd, only the $(\overline{P}, \overline{Q})$ state remains unresolved at the error decoder. Therefore, it is sufficient for all information digit sequences $m_i(j)$ $(j = 1, 2, \ldots, n - 1)$ to be differentially encoded together, as shown in Figure 11.

APPLICATION OF NONTRANSPARENT CODES

For state $(\overline{P}, \overline{Q})$, as shown in equation (3), estimated parity bit $\hat{C}_i(n)$ does not coincide with received parity bit $\overline{C}_i(n)$. For state $(\overline{P}, Q)$, estimated parity bit $\hat{C}_i(n)$ can be calculated as in equation (4):

$$\hat{C}_i(n) = \sum_{i=0}^{N-1} \sum_{j=1}^{n/2} \overline{m_{L-i}}(j)\, g_i(j) + \sum_{i=0}^{N-1} \sum_{j=(n/2)+1}^{n-1} m_{L-i}(j)\, g_i(j)$$

$$= C_i(n) \quad . \tag{6}$$

Since estimated parity bit $\hat{C}_i(n)$ coincides with received parity bit $C_i(n)$ in state $(\bar{P}, Q)$, sequences $\overline{m_i}(j)$ $(j = 1, 2, \ldots, n/2)$, $m_i(j)$ $[j = (n/2) + 1, \ldots, n - 1]$ will be obtained from the decoder output.

In state $(P, \bar{Q})$ estimated parity bit $\hat{C}_i(n)$ can be calculated as in equation (5):

$$\hat{C}_i(n) = \sum_{i=0}^{N-1} \sum_{j=1}^{n/2} m_{L-i}(j) \, g_i(j) + \sum_{i=0}^{N-1} \sum_{J=(n/2)+1}^{n-1} \overline{m}_{L-i}(j) \, g_i(j)$$

$$= C_i(n) \quad . \tag{7}$$

Thus, estimated parity bit $\hat{C}_i(n)$ does not coincide with received parity bit $\overline{C}_i(n)$ in the case of $(P, \bar{Q})$. Therefore, to resolve the kinds of ambiguity described above, it is sufficient for sequence $m_i(j)$ $(j = 1, 2, \ldots, n/2)$ to be differentially encoded and decoded, as shown in Figure 14.

m'ᵢ(j): DIFFERENTIALLY ENCODED mᵢ(j)

AMBIGUITY: RECOVERED CARRIER PHASE
            AND PHASE ROTATION DIRECTION

CODE: NON–TRANSPARENT
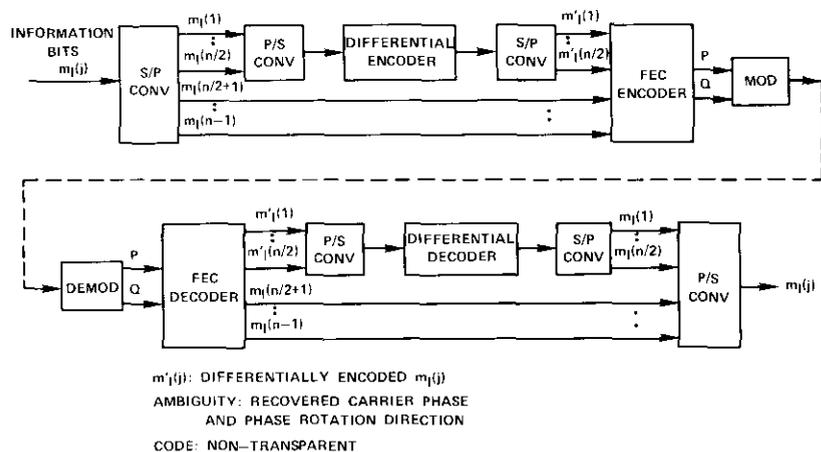
Figure 14. *Phase Ambiguity Resolution by Differential Coding Outside an FEC Codec (case 3)*

Conclusion

Two alternative methods of resolving phase ambiguity in a 4-phase PSK modulation system which uses FEC convolutional codes have been investigated. The first method uses recovered carrier phase ambiguity

resolution by differential coding inside an FEC codec. Some means of breaking up the occurrence of pairs of errors which are associated with this method is necessary to permit the error decoder to work properly. The following methods have been proposed: columnwise writing and rowwise reading, symbol interleaving, and M-symbol differential coding.

An advantage of these methods is that they do not require synchronization time to resolve the phase ambiguity, although the bit-error rate equivalent to the decoder input is twice that achieved without differential coding. Thus, this method is suitable for burst mode signals.

The second method of phase ambiguity resolution uses differential coding outside an FEC codec. In this case, a syndrome pattern is used in conjunction with the synchronizer to resolve both recovered carrier phase and phase rotation direction ambiguity. However, differential coding is still required to resolve some residual phase ambiguities which the synchronizer cannot resolve by itself. This is due to the transparencies of the codes, as described in the following. For recovered carrier phase ambiguity only, transparent or nontransparent FEC codes may be applied. For transparent codes the $(\bar{P}, \bar{Q})$ state remains unresolved at the error decoder so that all the information digits $m_i(j)$ $(j = 1, 2, \ldots, n - 1)$ must be differentially encoded. In the case of nontransparent codes, there is no need for differential coding, since the synchronizer itself can resolve the phase ambiguity.

For both recovered carrier phase and phase rotation direction ambiguity, the same two codes are applicable. In the case of transparent codes and $n/2$ even, states $(\bar{P}, \bar{Q})$, $(\bar{P}, Q)$, and $(P, \bar{Q})$ remain unresolved at the error decoder. Therefore, information digits $m_i(j)$ $(j = 1, 2, \ldots, n/2)$ and $m_i(j)$ $[j = (n/2) + 1, (n/2) + 2, \ldots, n - 1]$ are differentially encoded independently. For $n/2$ odd, only state $(\bar{P}, \bar{Q})$ remains unresolved at the error decoder output. Therefore, it is sufficient for all the information digits $m_i(j)$ $(j = 1, 2, \ldots, n - 1)$ to be differentially encoded together. For nontransparent codes, state $(\bar{P}, Q)$ remains unresolved at the error decoder. Therefore, it is sufficient for information digits $m_i(j)$ $(j = 1, 2, \ldots, n/2)$ to be differentially coded.

In the second method of phase ambiguity resolution, since the differential coding is used outside an FEC codec, the error-correcting decoder does not encounter the double-error phenomenon and only the bit-error rate of the error-correcting decoder output (improved bit-error rate) is doubled. Doubling the decoder output error rate results in a smaller E_b/N_o loss than doubling the input error rate because the curve of error rate versus E_b/N_o is steeper for the output. Thus, in terms of

bit-error rate improvement characteristics, the second method is superior to the first. The disadvantage of the second method is that a relatively long time is required to resolve the phase ambiguity. Therefore, it may not be suitable for application to burst mode operation, as in a TDMA or SPADE system.

Acknowledgments

References

[1] R. B. McClure, "Status and Progress of the INTELSAT SCPC System," IEEE 1975 International Conference on Communications, San Francisco, California, June 1975, *Proc.*, pp. 12-6–12-11.

[2] L. S. Golding, "DITEC—A Digital Television Communications System for Satellite Links," Second International Conference on Digital Satellite Communications, Paris, France, November 1972, *Proc.*, pp. 384–397.

[3] I. Matsushita, Y. Tsuji, and Y. Sohma, "Applications of Error-Correcting Codes to TDMA Satellite Communications," IEEE 1973 International Conference on Communications, Seattle, Washington, June 1973.

[4] R. C. Davis and R. J. F. Fang, "Channel Capacity Extension via FDMA/PSK/DSI," Third International Conference on Digital Satellite Communications, Kyoto, Japan, November 1975, *Proc.*, pp. 170–179.

[5] D. J. Withers and A. K. Jefferis, "A Time Division Multiple Access System for INTELSAT," AIAA 4th Communications Satellite Systems Conference, Washington, D.C., April 1972, Paper no. 72–538.

[6] E. R. Cacciamani, Jr. and C. J. Wolejsza, Jr., "Phase-Ambiguity Resolution in a Four-Phase PSK Communications System," *IEEE Transactions on Communication Technology*, COM-19, No. 6, Pt. II, December 1971, pp. 1200–1210.

[7] G. D. Forney, Jr. and E. K. Bower, "A High-Speed Sequential Decoder: Prototype Design and Test," *IEEE Transactions on Communication Technology*, COM-19, No. 5, Pt. II, October 1971, pp. 821–834.

[8] P. L. Chen and R. A. Rutledge, "Error Correcting Codes for Satellite Communications Channels," *IBM Journal of Research and Development*, Vol. 20, No. 2, March 1976, pp. 168–175.

Yoshikazu Tsuji received B.S. and M.S. degrees in electrical engineering from Kyoto University in 1965 and 1967, respectively. Immediately thereafter, he joined Fujitsu Ltd., where he was engaged in the development of digital satellite communications systems. In March 1975 he joined COMSAT Laboratories as an INTELSAT nominee. Since that time, he has been engaged in the development of SS-TDMA satellite communications systems. He is a member of IECE of Japan.

A model for microwave propagation along an earth-satellite path

D. J. FANG AND J. JIH

(Manuscript received March 15, 1976)

Abstract

Recent studies of microwave transmission through a precipitation medium have contributed to clarify the mechanism of propagation and have also provided a framework for the interpretation of measurements at centimeter and millimeter wavelengths. However, the problem of utilizing the results of theoretical studies to establish engineering models in order to make reliable predictions of propagation characteristics on a specific satellite-earth microwave link remains to be resolved. The primary difficulties are related to the effect of elevation angles on propagation parameters, which is not known in detail; the precise evaluation of the raindrop canting effect, for which equations are not available; and the scaling of the propagation parameters, for which proper procedures are not established.

The present paper attempts to provide engineering solutions to these problems. A slant path model is developed for electromagnetic wave propagation at centimeter and millimeter wavelengths through rain. Considerations which take into account the random canting effect of raindrops are provided. The scaling procedures for transferring propagation information from one polarization to

another, one frequency to another, and one elevation angle to another are demonstrated step by step.

Introduction

It is known that rain is a principal cause of signal attenuation and cross polarization in a terrestrial or satellite transmission link at centimeter and millimeter wavelengths. It can be shown that most of the current theoretical studies on microwave propagation through rain follow three steps:

   a. analysis of forward scattering for a plane electromagnetic wave incident upon a single raindrop,

   b. calculation of the diffraction of the wave penetrating perpendicularly across a thin layer of precipitation,

   c. multiple-layer iterative integration to derive the wave field which passes through a precipitation medium.

Papers indicating the scientific and technical findings obtained from all or part of these steps are numerous (see, for example, References 1-9). Some of these papers provide experimental data to substantiate the theoretical findings. Thus, theoretical studies are not only important in the elucidation of the physics of electromagnetic waves through precipitation, but also for establishing valid frameworks within which experimenters can apply and interpret the data collected from the various propagation experiments. On the other hand, for an engineer involved in estimating propagation parameters for the design of a specific microwave link, the results of existing theoretical studies are inadequate and/or inapplicable.

A review of the literature reveals many limitations in the theoretical studies. For instance, in step *a*, the numerical methods of matching the boundary conditions for a spheroidal raindrop are so tedious and expensive that only a very limited number of computations of forward scattering fields for selected frequencies, drop sizes, and drop temperatures have been performed [1], [3], [4], [9]. Also, in most computations, the propagation direction is assumed to be perpendicular to the raindrops' axes of symmetry, which renders the results inapplicable to slant path propagation. For step *b*, the Laws and Parsons drop size distribution [1] [4], [7] has been used consistently for simplicity, although it is now known that other distributions, such as the Marshall-Palmer [10] and Joss distributions [11], may be more realistic for heavy storms. In step *c*, the analysis is generally based on the electric field differentials from one layer to the next [5], [6], [8], [12] without including dispersion effects. These short-

comings, together with the meteorological uncertainties, such as the spatial rain rate distribution along a propagation path, provide inadequate building blocks for engineering design applications.

On the basis of available theoretical results, this paper proposes a format for compiling and editing the relevant data, and for making engineering interferences to supplement relevant yet scanty data, as required for practical applications on a slant path satellite-earth link. The following section introduces a format which will permit detailed description of the dependence of propagation parameters on elevation angles, a vital piece of information for slant path applications. The next section analyzes the effect of raindrop canting. This effect is much more complicated in slant path propagation than in terrestrial links, simply because the relative angles between the axes of the canted raindrops and wave vectors are not trivial. Precise formulas for the evaluation of the random canting effect on propagation parameters will also be presented.

Another difficulty which engineers must face when applying the theoretical results involves the scaling of propagation parameters from one frequency to another, one polarization to another, and one elevation angle to another. To demonstrate the procedure, a special example of slant path transmission will be discussed.

A new formulation of field equations for microwave propagation through a precipitation medium

The geometry for analyzing wave scattering through a rain cloud is shown in Figure 1. In a simplified example, the raindrops are presumed to fall vertically downward. The axis of symmetry of the spheroidal raindrops, denoted as $\vec{i}_c$, is aligned along the vertical axis. The x axis is chosen so that the incident wave normal lies in the x-z plane, making an angle ψ with the z axis. The unit vector for wave propagation, $\vec{i}_k$, is given by

$$\vec{i}_k = -\vec{i}_z \cos \psi - \vec{i}_x \sin \psi \ . \tag{1}$$

At the penetrating point, the incident wave field, $\vec{E}_o$, is decomposed into horizontally and vertically polarized components, E_{Ho} and E_{Vo}, respectively, for separate scattering analysis. The standard wave factor $\exp[i\omega t - ik(\vec{i}_k \cdot \vec{r})]$ is omitted for simplicity. The sense of polarization is such that the unit vectors $\vec{i}_H$ and $\vec{i}_V$ are specified as

$$\vec{i}_H = \vec{i}_y \tag{2a}$$
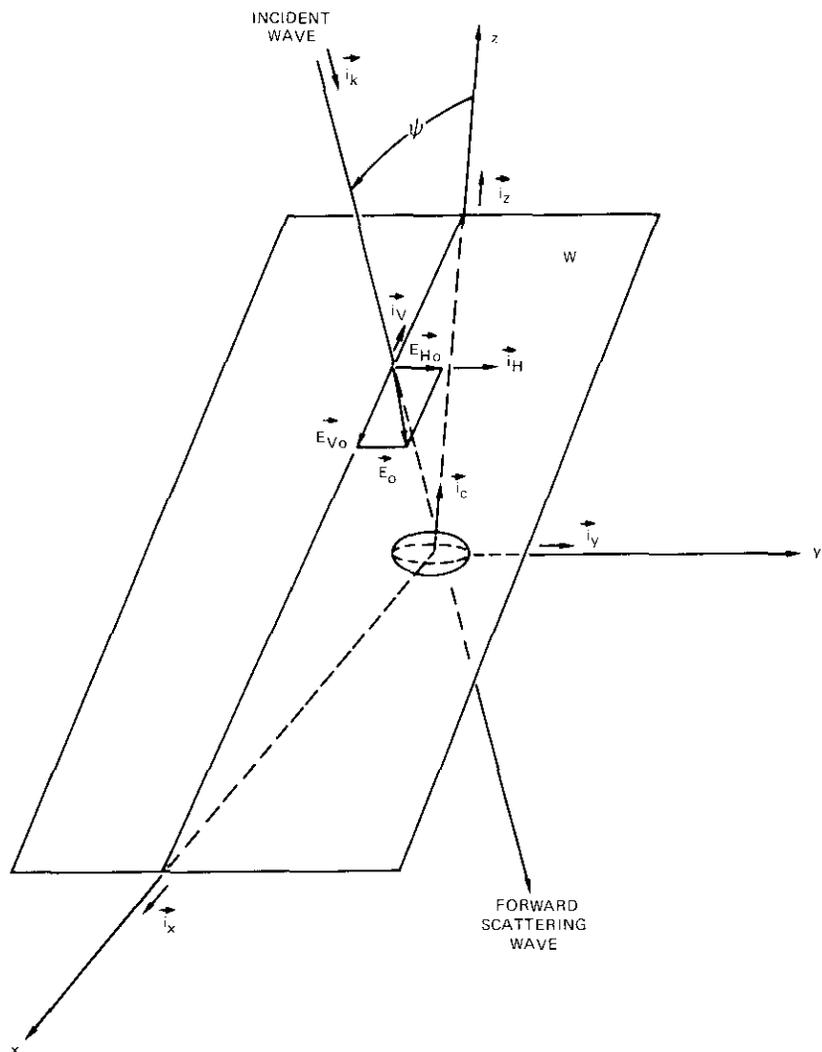
Figure 1. *Geometry for a Plane Wave Penetrating a Raindrop: Uncanted*
Drop with Its Symmetric Axis $(\vec{i_c})$ *Aligned Along the Vertical*
Axis (A general incident field $\vec{E_o}$ *can be decomposed into*
horizontally and vertically polarized fields $\vec{E}_{Ho}$ *and* $\vec{E}_{Vo}$ *for*
separate forward scattering analysis.)

$$\vec{i_V} = -\vec{i_x}\cos\psi + \vec{i_z}\sin\psi \quad . \tag{2b}$$

The established procedure for calculating the far-field intensity, E_H and E_V, for an incident plane wave penetrating a rain cell layer L km thick is to derive the equivalent complex wave factors, K_H and K_V, for a given frequency, f, rain rate, R, and propagation angle, ψ, such that [3], [7]

$$\begin{bmatrix} E_H(f,\,R,\,\psi,\,L) \\ E_V(f,\,R,\,\psi,\,L) \end{bmatrix}$$

$$= \begin{bmatrix} \exp\{-iK_H(f,\,R,\,\psi)\,L\} & 0 \\ 0 & \exp\{-iK_V(f,\,R,\,\psi)\,L\} \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix} \quad . \tag{3}$$

The propagation parameters, *i.e.*, attenuation, A, in dB/km and phase shift, Φ, in deg/km, are related to $K_{H,V}$ as follows:

$$A_{H,V}(\psi) = 8.686\ Im[K_{H,V}(\psi)] \tag{4a}$$

$$\Phi_{H,V}(\psi) = \frac{180}{\pi}\ Re[K_{H,V}(\psi)] \quad . \tag{4b}$$

For a terrestrial propagation link, $\psi = 90°$, values of $A_{H,V}$ and $\Phi_{H,V}$ at selected frequencies are tabulated in the literature [3], [7], [13]. Values for slant path propagation are very limited.

Slant path microwave propagation for a satellite-earth link can be properly modeled only by first determining the functional dependence of equation (3) on ψ. This can be done by introducing the following parameters:

$$R_p(f,\,R,\,\psi,\,L) = \exp\{-[a(f)\ R^{m(f)} + ib(f)\ R^{n(f)}]\ L\cos^2\psi\} \tag{5a}$$

$$R_m(f,\,R,\,\psi,\,L) = \sin^2\psi\ R_p(f,\,R,\,\psi,\,L) \tag{5b}$$

$$F_p(f,\,R,1) = \frac{1}{2}\{\exp[-iK_H(f,\,R,\,90)] + \exp[-iK_V(f,\,R,\,90)]\} \tag{6a}$$

$$F_m(f,\,R,\,1) = \frac{1}{2}\{\exp[-iK_H(f,\,R,\,90)] - \exp[-iK_V(f,\,R,\,90)]\} \tag{6b}$$

$$F_p(f, R, L) = [F_p(f, R, 1)]^L \tag{7a}$$

$$F_m(f, R, L) = L[F_p(f, R, 1)]^{L-1} F_m(f, R, 1) \quad . \tag{7b}$$

The formalization of parameters $F_p(f, R, L)$ and $F_m(f, R, L)$ is presented in Appendix A. It suffices herein to note that equation (6) gives the per-kilometer values ($L = 1$) of F_p and F_m based on terrestrial propagation conditions ($\psi = 90$) and that equation (7) can be used to evaluate F_p and F_m at any distance L once the per-kilometer values are known. Equation (3) can then be written as

$$
\begin{bmatrix} E_H(f, R, \psi, L) \\ E_V(f, R, \psi, L) \end{bmatrix} = F_p(f, R, L) \, R_p(f, R, \psi, L) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix}
$$
$$
+ F_m(f, R, L) \, R_m(f, R, \psi, L) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix} \quad . \tag{8}
$$

This formulation of field equations for microwave propagation through a precipitation medium, which is believed to be novel, provides a description of the electric fields as functions of rain rate, frequency, propagation angle, and path length. Values of $a(f)$, $b(f)$, $m(f)$, and $n(f)$ in equation (5a) can be assessed on the basis of available theoretical data of Oguchi [3], [13] and Chu [7]. The results are plotted in Figure 2, which also gives specific numerical magnitudes at $f = 4, 6, 11, 30$, and 34.8 GHz. Based on the same data source, the complex quantities $F_p(f, R, 1)$ and $F_m(f, R, 1)$ can be evaluated for specific rain rates and frequencies. The results are plotted in Figures 3a and 3b. Details of the formulation are given in Appendix A.

For practical applications with f, R, ψ, and L as variables, equation (8) is more straightforward than other forms of equations which operate indirectly through parameters $A_{H,V}$ and $\Phi_{H,V}$. In actual calculations the inputs $F_p(f, R, 1)$ and $F_m(f, R, 1)$ can be generated from reliable terrestrial experiments if they yield more realistic values than those given in Figures 3a and 3b. It is also of interest to note that, in earlier studies [6], [7], [14] in which the ψ dependence could be described only through the differences of $A_H - A_V$ and $\Phi_H - \Phi_V$, the differences were modeled as proportional to $\sin^2 \psi$. Equation (8) provides a much more refined description of the ψ dependence; it indicates that the $\sin^2 \psi$ proportionality is good only to the first order for frequencies lower than 20 GHz.
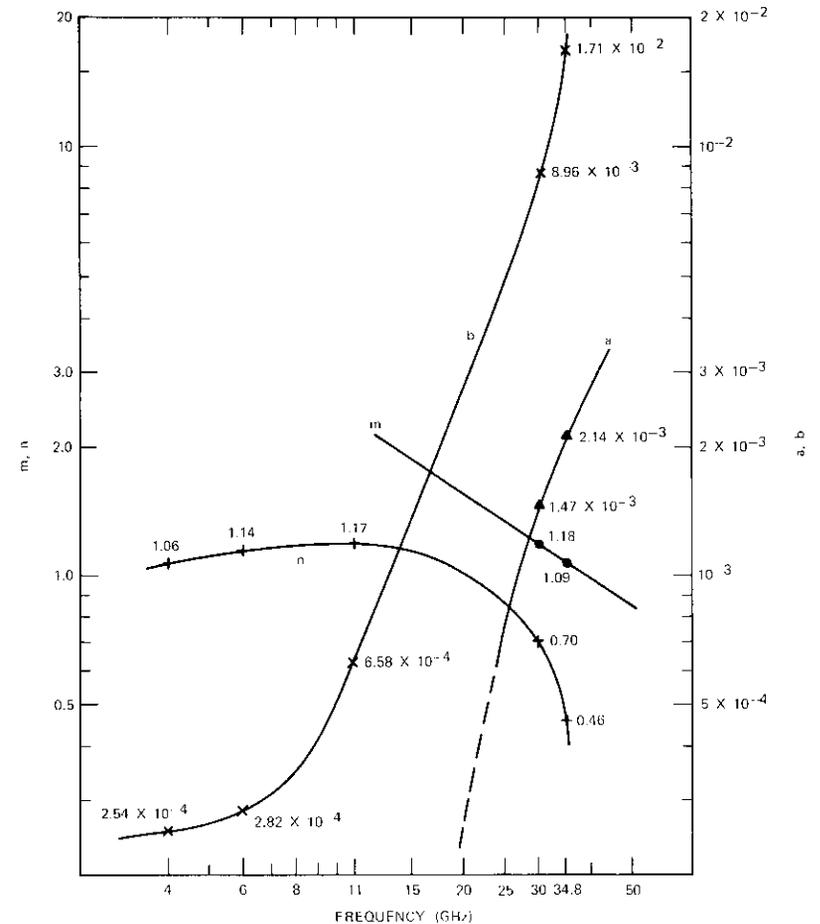
Figure 2. *Values of a, b, m, and n of the Modeling Equation (5) (The derivation is given in Appendix A; marked values are for 4-, 6-, 11-, 30-, and 34.8-GHz frequencies, respectively.)*
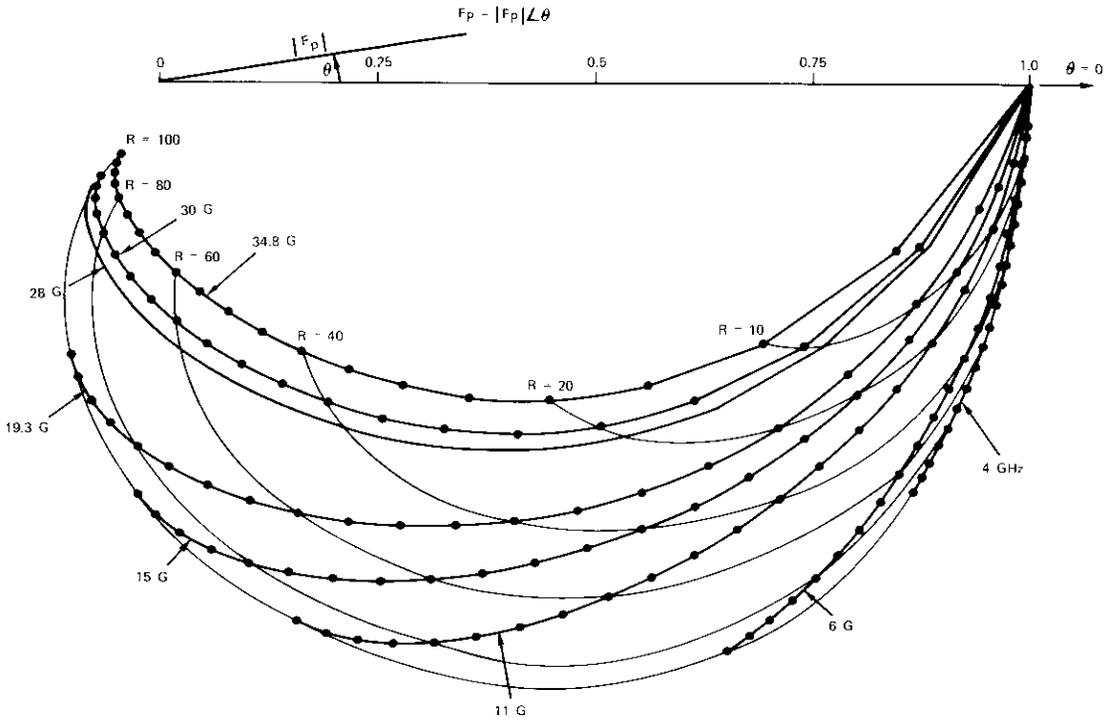
Figure 3a. *Plots of F_p in Polar Coordinates (precise values of F_p are available in Table A-1 in Appendix A)*
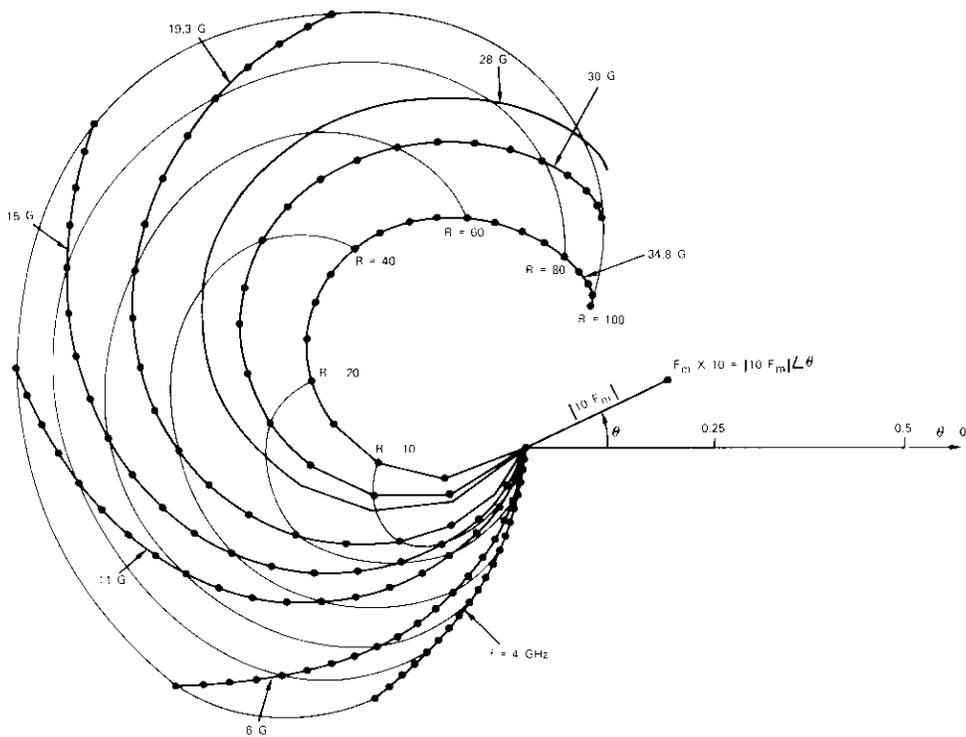
Figure 3b. *Plots of F_m in Polar Coordinates (precise values of F_m are available in Table A-1 in Appendix A)*

Random canting effect on microwave propagation

In applications involving cross-polarization isolation, the field equations should take into account the random canting effect of the raindrops. Since for each single raindrop the axis of symmetry, $\vec{i'_c}$, makes the polar angles θ' and Φ' with respect to the original x–y–z coordinates (see Figure 4), equation (8) will be valid in the new x'–y'–z' coordinate system only for those raindrops having a pointing direction which is the same as that of $\vec{i'_c}$. Furthermore, the field vectors are decomposed along the $\vec{i'_H}$ and $\vec{i'_V}$ directions, and a matrix transformation is required to bring the field vectors back into the original x–y–z system.

In terms of the procedures which can be found in Reference 14, equation (8) can be modified as follows:

$$\begin{bmatrix} E_H \\ E_V \end{bmatrix} = F_p T_1(\psi) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix} + F_m T_2(\psi) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix}$$

$$- F_m T_3(\psi) \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix} \tag{9}$$

where

$$T_1(\psi) = \int \exp\{-[aR^m + ibR^n] L \cos^2 \psi'\} \sin \theta'$$

$$\cdot p(\theta', \Phi') \, d\theta' \, d\Phi' \tag{10a}$$

$$T_2(\psi) = \int \sin^2 \psi' \exp\{-[aR^m + ibR^n] L \cos^2 \psi'\}$$

$$\cdot \left[1 - \frac{2 \sin^2 \theta' \sin^2 \Phi'}{\sin^2 \psi'} \right] \sin \theta'$$

$$\cdot p(\theta', \Phi') \, d\theta' \, d\Phi' \tag{10b}$$

$$T_3(\psi) = \int \sin^2 \psi' \exp\{-[aR^m + ibR^n] L \cos^2 \psi'\}$$

$$\cdot \frac{2 \sin \theta' \sin \Phi'}{\sin^2 \psi'} (\sin \psi \cos \theta'$$

$$- \cos \psi \sin \theta' \cos \Phi') \sin \theta'$$

$$\cdot p(\theta', \Phi') \, d\theta' \, d\Phi' \quad . \tag{10c}$$
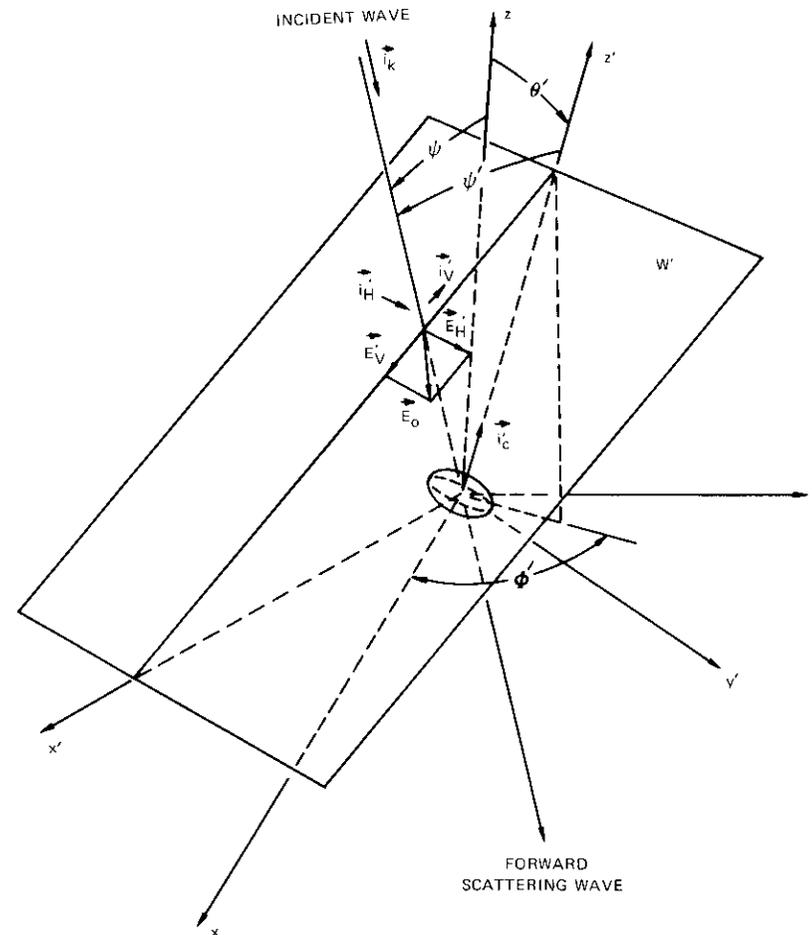
Figure 4. *Geometry for a Plane Wave Penetrating a Raindrop: Canted Drop with $\vec{i'_c}$ Making θ' and Φ' Polar Angles with Respect to the Original Coordinates (Forward scattering analysis can be made via $\vec{E}_H$ and $\vec{E}_V$.)*

Hence, a formal relationship between the transmitted and received fields is obtained for randomly canted rainfall events. The canting distribution function $p(\theta', \Phi')$ specifies the portion of raindrops with symmetric axes $\vec{i'_c}$ pointing at a direction within a solid angle $d\Omega'$ such that

$$d\Omega' = \sin\theta'\, d\theta'\, d\Phi' \quad , \qquad 0 \le \theta' < \frac{\pi}{2} \quad , \qquad 0 \le \Phi' < 2\pi$$

and

$$\int_{\theta'=0}^{\pi/2}\int_{\Phi'=0}^{2\pi} p(\Omega')\, d\Omega' = 1 \quad . \tag{11}$$

The angle between $\vec{i_c}$ and the z axis is ψ'; i.e.,

$$\cos\psi' = \sin\psi \sin\theta' \cos\Phi' + \cos\psi \cos\theta' \quad . \tag{12}$$

For equation (9), the common assumption [14] that canting distribution is independent of rain rate has been adopted; otherwise, the factors F_p and F_m would be included in the integrations of equation (10). Once the formal relationship shown in equation (9) has been established, three cases significant for engineering applications can be analyzed.

Case 1: Unique direction canting

For the case in which all raindrops are canted in a unique direction, e.g., θ' and Φ', the distribution function approximates a delta function, and equation (9) becomes

$$\begin{bmatrix} E_H \\ E_V \end{bmatrix} = F_p R_p(f,\, R,\, \psi',\, L) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix}$$

$$+ F_m R_m(f,\, R,\, \psi',\, L) \left[1 - 2\frac{\sin^2\theta'\sin^2\Phi'}{\sin^2\psi'} \right] \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix}$$

$$- F_m R_m(f,\, R,\, \psi',\, L)\, \frac{2\sin\theta'\sin\Phi'}{\sin^2\psi'}$$

$$\cdot (\sin\psi\cos\theta' - \cos\psi\sin\theta'\cos\Phi') \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix} \quad . \tag{13}$$

To provide worst-case estimates of cross-polarization levels [6], [14], [15], this equation is usually evaluated after assigning Φ' as 90° and ψ' as 45°.

Case 2: Turbulence

Knowledge of turbulence in a precipitation medium is usually rudimentary. However, it is known that [16], in the case of turbulence, the random pointing of $\vec{i_c}$ is azimuthally symmetric with reference to the local vertical z axis. The canting distribution function p can be described as

$$p(\theta',\, \Phi') = \frac{1}{2\pi}\, p(\theta') \quad . \tag{14}$$

It can be shown that, due to symmetry,

$$T_3(\psi) = 0 \tag{15}$$

and that

$$\begin{bmatrix} E_H \\ E_V \end{bmatrix} = F_p T_1(\psi) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix} + F_m T_2(\psi) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix} \quad . \tag{16}$$

The off-diagonal terms in the matrices of these equations are zero. This confirms the generally accepted belief that, although air is almost always turbulent during rainfall, turbulence alone will not produce any significant amount of depolarization for horizontally or vertically polarized waves. On the other hand, it is important to note that the turbulence cannot be disregarded in the propagation analysis because the random canting will degrade the transmission parameters $T_1(\psi)$ and $T_2(\psi)$, as indicated in equation (16).

Case 3: Prevailing wind

Another factor that causes raindrops to be canted is the presence of a prevailing wind system having a vertical nonuniform profile [16]. In this case, the mean canting direction is no longer vertical as in the previous case, but is specified by polar angles θ_s and Φ_s. This case can be analyzed by introducing a new coordinate system $x_s\, y_s\, z_s$ such that z_s is aligned along the mean direction of $\vec{i_c}$. It follows that the random canting will be azimuthally symmetric in the new system, and that equation (16) will still be applicable if all the formulations are done in reference to the s coordinate system. After a coordinate transformation,

$$\begin{bmatrix} E_H \\ E_V \end{bmatrix} = F_p T_1(\psi_s) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix}$$

$$+ F_m T_2(\psi_s) \left[1 - \frac{2\sin^2\theta_s\sin^2\Phi_s}{\sin^2\psi_s} \right] \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix}$$

$$- F_m T_2(\psi_s)\, \frac{2\sin\theta_s\sin\Phi_s}{\sin^2\psi_s}$$

$$\cdot(\sin \psi \cos \psi_s - \cos \psi \sin \psi_s \cos \Phi_s) \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} E_{H_o} \\ E_{V_o} \end{bmatrix} \qquad (17)$$

where $\qquad \cos \psi_s = \sin \psi \sin \theta_s \cos \Phi_s + \cos \psi \cos \theta_s$. $\qquad (18)$

The presence of off-diagonal terms in equation (17) indicates that the cross polarization exists.

Simulation of the canting distribution function

The actual $p(\theta', \Phi')$ function is dictated by meteorological factors including turbulence, the prevailing wind system and its spatial gradient, individual raindrop vibrations, and drop size distributions. Only preliminary studies relating to $p(\theta', \Phi')$ are currently available [16], [17]. To examine the random canting effect on microwave transmission, it is often necessary to perform theoretical calculations based on an approximately postulated profile of $p(\theta', \Phi')$ which is sufficiently versatile to simulate any realistic distribution of the canting.

The profile postulated here is

$$p(\theta', \Phi') = \frac{u + 1}{2\pi} \cos^u \theta \qquad (19a)$$

where $\qquad u = -\left\{ 1 + \left[\frac{\log 2}{\log (\cos \sigma_\theta)} \right] \right\} \qquad (19b)$

with a peak centered at $\theta' = 0$ whose width is determined by an angle σ_θ such that

$$\int_{\Phi'=0}^{2\pi} \int_{\theta'=0}^{\sigma_\theta} p(\theta', \Phi') \sin \theta' \, d\theta' \, d\Phi' = \frac{1}{2} \qquad (20)$$

i.e., 50 percent of the random pointing of $\vec{i_c}$ is within the solid angle centered at $\theta' = 0$ and bounded by $\theta' = \sigma_\theta$.

Values of T_1, T_2, and T_3 can be evaluated from equation (10) by first expanding the exponentials in Taylor series terms. Since the absolute value of the exponential term is much less than unity, only the first three terms up to the order of W^2, where

$$W = (aR^m + ibR^n) L \qquad (21)$$

are needed in the expansion. In terms of equations (12) and (19), term-by-term integrations of T_1, T_2, and T_3 are possible. After several passages,

$$T_1(\psi) = 1 - (1 + u \cos^2 \psi) \frac{W}{u + 3} + \frac{W^2}{2(u + 3)(u + 5)}$$
$$\cdot [3 + 6u \cos^2 \psi - (3 - u) u \cos^4 \psi] \qquad (22a)$$

$$T_2(\psi) = \frac{1}{(u + 3)(u + 5)(u + 7)} \left\{ u \left[u^2 + (12 - W) u + 35 - 7W \right. \right.$$
$$\left. + \frac{3}{2} W^2 \right] - u \left[(W + 1) u^2 + (12 + 4W - 3W^2) u + 35 \right.$$
$$\left. - 21W + \frac{15}{2} W^2 \right] \cos^2 \psi + u \left[W \left(1 + \frac{W}{2} \right) u^2 + W \right.$$
$$\left. \cdot (5 - 6W) u - 2W(7 - 5W) \right] \cos^4 \psi + \left[-\frac{1}{2} u^3 + 3u^2 \right.$$
$$\left. \left. - 4u \right] W^2 \cos^6 \psi \right\} \qquad (22b)$$

$$T_3(\psi) = 0 \quad . \qquad (22c)$$

An important application of these equations relates a physical distribution of random canting to the so-called "effective mean canting" used extensively by wave propagationists. "Effective mean canting" means that, although the pointing of $\vec{i_c}$ for the raindrops is random, its effect on microwave propagation can always be evaluated by assuming that all the raindrops point in a unique direction after properly adjusting the evaluated microwave field to take into account the randomness. Comparison of equations (9) and (13) with specific inputs from equations (19)–(22) makes it possible to find the effective mean canting direction as well as the magnitude required to adjust for modeling an arbitrarily canted rainfall event as a unique direction event.

Slant path propagation studies

The relationships between the transmitted field and received field given in equations (8), (9), (13), (16), and (17) can be reduced to the form

$$\begin{bmatrix} E_H \\ E_V \end{bmatrix} = \begin{bmatrix} a_1(f, \psi, L, R, \theta_s, \Phi_s, \sigma_\theta) & a_2(f, \psi, L, R, \theta_s, \Phi_s, \sigma_\theta) \\ a_3(f, \psi, L, R, \theta_s, \Phi_s, \sigma_\theta) & a_4(f, \psi, L, R, \theta_s, \Phi_s, \sigma_\theta) \end{bmatrix} \begin{bmatrix} E_{Ho} \\ E_{Vo} \end{bmatrix} . \quad (23)$$

Hence, a series of analytic equations is obtained which makes it possible to calculate numerically the values of a_1, a_2, a_3, and a_4 once the frequency (f), propagation angle (ψ), path length (L), rain rate (R), effective mean canting direction (θ_s and Φ_s), and random canting spread (σ_θ) are specified. The two basic theoretical inputs F_p and F_m can be read from Table A-1 or Figure 3, or derived from experimental results. If the frequency and rain rate are such that the values of F_p and F_m are not readily available, F_p and F_m can be estimated from the known values by proper scaling frequency and/or rain rate through the quantities $A_{H,V}$ and $\Phi_{H,V}$ introduced in equation (4).

Once values of a_1, a_2, a_3, and a_4 are determined, important propagation parameters such as attenuation (ATN) and cross-polarization discrimination ratio (XPD) can be evaluated. Definitions and formulas used to evaluate the parameters often differ somewhat in the various references due to different mathematical approaches and experimental configurations. Standard definitions and formulas are given in Table 1. (It should be noted that, in the present paper, a_2 is always equal to a_3.) The formulas for evaluating ATN and XPD in circular polarization are derived after using the following conversion equations:

$$\begin{bmatrix} \vec{i}_R & E_{Lo} & E_L \\ \vec{i}_L & E_{Ro} & E_R \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix} \begin{bmatrix} \vec{i}_H & E_{Ho} & E_H \\ \vec{i}_V & E_{Vo} & E_V \end{bmatrix} . \quad (24)$$

It is important to note that, if the meteorological conditions, such as rain rate, canting direction, and random spread, are not spatially uniform along the propagation path, the received field should be evaluated by iterating over the consecutive segments of path lengths, L_i, where the meteorological conditions are locally uniform. That is, the a-matrix in equation (23) should be written as

$$\begin{bmatrix} a_1(f, \psi, L) & a_2(f, \psi, L) \\ a_3(f, \psi, L) & a_4(f, \psi, L) \end{bmatrix}$$

$$= \prod_i \begin{bmatrix} a_1(f, \psi, L_i, R_i, \theta_{si}, \Phi_{si}, \sigma_{\theta i}) & a_2(f, \psi, L_i, R_i, \theta_{si}, \Phi_{si}, \sigma_{\theta i}) \\ a_3(f, \psi, L_i, R_i, \theta_{si}, \Phi_{si}, \sigma_{\theta i}) & a_4(f, \psi, L_i, R_i, \theta_{si}, \Phi_{si}, \sigma_{\theta i}) \end{bmatrix} \quad (25a)$$

where
$$L = \sum_i L_i . \quad (25b)$$

TABLE 1. DEFINITIONS AND FORMULAS FOR ATTENUATION AND CROSS POLARIZATION

Parameters	Transmitted Field, $E_{Ho}\vec{i}_H + E_{Vo}\vec{i}_V$ or $E_{Ro}\vec{i}_R + E_{Lo}\vec{i}_L$	Received Field, $E_H\vec{i}_H + E_V\vec{i}_V$ or $E_R\vec{i}_R + E_L\vec{i}_L$	Definitions and Formulas
Attenuation of horizontally polarized wave	E_{Ho}	E_H	$ATN_H = 10 \log \left\| \dfrac{E_H}{E_{Ho}} \right\|^2 = 20 \log |a_1|$
Attenuation of vertically polarized wave	E_{Vo}	E_V	$ATN_L = 10 \log \left\| \dfrac{E_V}{E_{Vo}} \right\|^2 = 20 \log |a_4|$
Cross polarization from horizontal to vertical waves	E_{Ho}	E_H and E_V	$XPD_H = 10 \log \left\| \dfrac{E_V}{E_H} \right\|^2 = 20 \log \left| \dfrac{a_3}{a_1} \right|$
Cross polarization from vertical to horizontal waves	E_{Vo}	E_H and E_V	$XPD_V = 10 \log \left\| \dfrac{E_H}{E_V} \right\|^2 = 20 \log \left| \dfrac{a_2}{a_4} \right|$
Attenuation of right-hand polarized wave	E_{Ro}	E_R	$ATN_R = 10 \log \left\| \dfrac{E_R}{E_{Ro}} \right\|^2 = 20 \log \left| \dfrac{a_1 + a_4}{2} \right|$
Attenuation of left-hand polarized wave	E_{Lo}	E_L	$ATN_L = 10 \log \left\| \dfrac{E_L}{E_{Lo}} \right\|^2 = 20 \log \left| \dfrac{a_1 + a_4}{2} \right|$
Cross polarization from right-hand to left-hand waves	E_{Ro}	E_R and E_L	$XPD_R = 10 \log \left\| \dfrac{E_L}{E_R} \right\|^2 = 20 \log \left| \dfrac{a_1 - a_4 - i2a_2}{a_1 + a_4} \right|$
Cross polarization from left-hand to right-hand waves	E_{Lo}	E_R and E_L	$XPD_L = 10 \log \left\| \dfrac{E_R}{E_L} \right\|^2 = 20 \log \left| \dfrac{a_1 - a_4 + i2a_2}{a_1 + a_4} \right|$

The validity of equation (25) can be verified by deliberately splitting a known path length, L, into two segments of identical meteorological conditions, e.g., σL and $(1 - \sigma) L$, where $\sigma < 1$, and ensuring that within the tolerance of equation (7) multiplication of the two matrices corresponding to the two segments will yield values of a_1, a_2, a_3, and a_4 which are the same as those obtained by evaluating a single path length.

A particularly interesting case is that of a local thunderstorm cell, where the total length of the slant path may be smaller than 1 to 2 km. Then the tedious matrix multiplication of equation (25) can be avoided by using the following approximation method. First, the following limits hold as $L_i = \Delta L_i$ becomes small:

$$F_p(f, R_i, \Delta L_i) \simeq 1 \tag{26a}$$

$$T_1(\psi_s) \simeq 1 - \frac{1 + u \cos^2 \psi_s}{u + 3} (aR_i^m + ibR_i^n) \Delta L_i \tag{26b}$$

$$F_m(f, R_i, \Delta L_i) \simeq \frac{F_m(f, R_i, 1)}{F_p(f, R_i, 1)} \Delta L_i \tag{26c}$$

$$T_2(\psi_s) \simeq \frac{u}{u + 3} \sin^2 \psi_s \quad . \tag{26d}$$

The element matrices involved in the multiplication can then be simplified by deleting all the $(\Delta L_i)^2$ terms. The final result is

$$\begin{bmatrix} a_1(f, \psi, L) & a_2(f, \psi, L) \\ a_3(f, \psi, L) & a_4(f, \psi, L) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \left\{ \int_0^L D_i(L_i)\, dL_i \right\} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$+ \left\{ \int_0^L E_i(L_i)\, dL_i \right\} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$
$$+ \left\{ \int_0^L F_i(L_i)\, dL_i \right\} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{27}$$

where, according to equation (17),

$$D_i = - \frac{1 + u \cos^2 \psi_s}{u + 3} (aR_i^m + ibR_i^n) \tag{28a}$$

$$E_i = \frac{F_m(f, R_i, 1)}{F_p(f, R_i, 1)} \frac{u}{u + 3} [\sin^2 \psi_s - 2 \sin^2 \theta_s \sin^2 \Phi_s] \tag{28b}$$

$$F_i = - \frac{F_m(f, R_i, 1)}{F_p(f, R_i, 1)} \frac{u}{u + 3} 2 \sin \theta_s \sin \Phi_s$$

$$\cdot (\sin \psi \cos \psi_s - \cos \psi \sin \psi_s \cos \Phi_s) \quad . \tag{28c}$$

Further calculations of ATN and XPD can then proceed in accordance with the formulas in Table 1.

To demonstrate the procedures to be used in a communications application, the following example will be considered. An 11-GHz terrestrial propagation experiment was conducted along a 10-km east-west path. The western site transmitted a left-hand circularly polarized wave, while the eastern site received in both left- and right-hand circularly polarized channels. Records of the experiment included the following items: ATN_L, XPD_L, rough rain gauge data, and prevailing wind information. The assignment is to reconstruct the records for establishing the levels of ATN_H and XPD_H at 28 GHz along a propagation path from the eastern station westward toward a satellite at a 30° elevation angle.

It is assumed that, at a specific time t_o during a precipitation event, the records indicated that ATN_L and XPD_L were -3.5 and -25.0 dB, respectively; the prevailing wind was southward; and the readings of the rain gauge at the time ± 3 minutes to t_o were fluctuating within the range of 35 to 75 mm/hr. It is then necessary to determine the estimated values of ATN_H and XPD_H at 28 GHz along the 30° slant path at t_o.

The available data concerning the terrestrial propagation path at instant t_o are as follows:

frequency, f:	11 GHz
propagation angle, ψ:	90°
rain rate, R_g:	35–75 mm/hr
path length, L_g:	≤ 10 km
azimuthal angle of raindrops, Φ_s:	90°
zenith angle of raindrops, θ_s:	unknown
random canting spread angle, σ_θ:	unknown

The first step is to determine all the available combinations of the unknowns, i.e., R_g, L_g, θ_s, and σ_θ, that will produce the correct attenuation and cross polarization. This can be achieved by evaluating ATN_L and XPD_L based on equations (17)–(22) using various combinations of numerical inputs of R_g, L_g, θ_s, and σ_θ and by drawing contour plots of Δ, where

$$\Delta = |ATN_L - (-3.5)| + |XPD_L - (-25.0)| \quad . \qquad (29)$$

The acceptable combinations are those which yield small values of Δ, e.g., less than 0.1 dB. There are infinite combinations, and for illustrative purposes a selection of the combinations is shown in Figure 5.
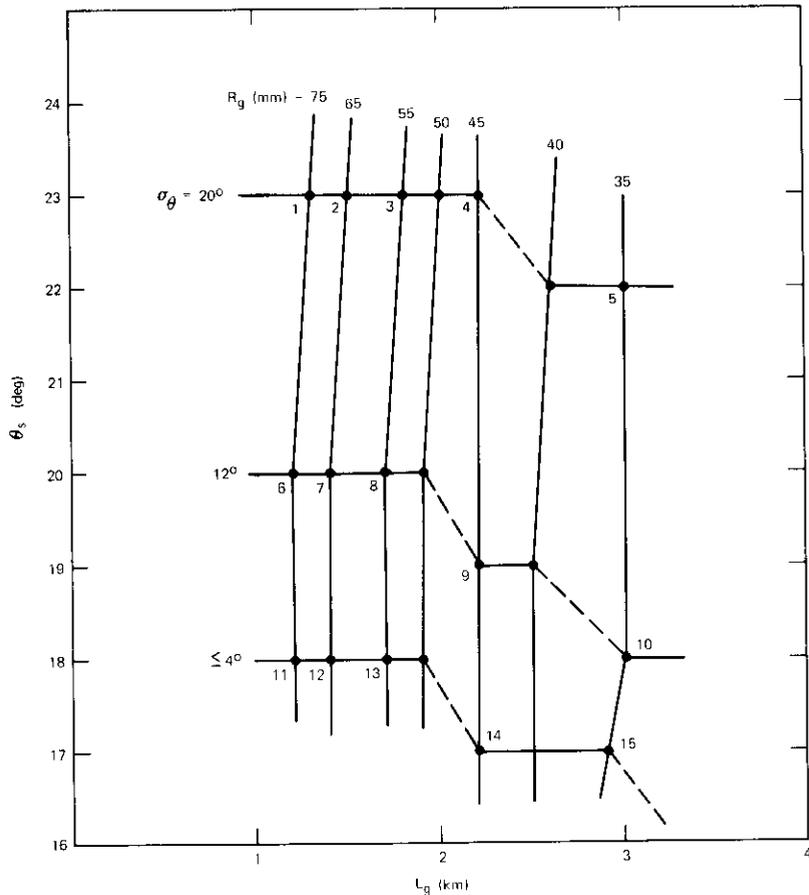
Figure 5. *A Selection of Combinations of L, θ_s, R, and σ_θ Which, Together with $\psi = 90°$ and $\Phi_s = 90°$, Will Yield ATN_L and XPD_L of -3.50 and -25.0 dB, Respectively, within an Accuracy of $\Delta \leq 0.1$ dB*

The next step is to perform the calculation at 28 GHz. The values of F_R and F_m at 28 GHz, which are not available in Tables A-1 and A-2 (Appendix A), are derived through scaling as shown in Figure 3. It is assumed that the height-elevated rain rate, R_h, equals the ground rain rate, R_g.* From the first step of the calculation,

$$
\begin{aligned}
f &= 28 \text{ GHz} \\
\psi &= 60° \\
R_h &= R_g \\
L &= L_g \cos \psi \\
\theta_s &= \theta_s \\
\sigma_\theta &= \sigma_\theta \\
\Phi_s &= 90°
\end{aligned}
\left.\begin{array}{l}
\\
\\
\text{input from the first} \\
\text{calculation step as} \\
\text{shown in Figure 5}
\end{array}\right.
$$

Final results of ATN_{II} and XPD_{II} are presented in Table 2, which shows that, although precise prediction is not possible, ATN_{II} and XPD_{II} at

TABLE 2. PREDICTED VALUES OF ATN_{II}, XPD_{II}, ATN_L, AND XPD_L AT 28 GHz BASED ON AVAILABLE INFORMATION AT 11 GHz

Point in Figure 5	f = 28 GHz			
	ATN_H	XPD_H	ATN_L	XPD_L
1	21.9	15.7	20.0	14.6
2	21.0	15.6	20.1	14.5
3	21.5	15.4	20.6	14.3
4	21.6	15.3	20.6	14.2
5	22.9	15.1	21.9	13.7
6	19.6	15.8	18.4	14.0
7	20.0	15.3	18.8	13.8
8	20.6	15.2	19.4	13.5
9	21.9	14.9	20.6	13.0
10	23.3	14.7	21.9	12.5
11	19.8	15.7	18.4	13.4
12	20.1	15.4	18.7	13.2
13	20.8	15.1	19.4	12.9
14	22.2	14.9	20.6	12.4
15	22.6	14.7	21.1	12.2
Average	21.3	15.3	20.0	13.5
Range	19.6–23.3	14.7–15.8	18.4–21.9	12.2–14.6

* Reference 18 deals with a method applicable whenever the above equality does not hold.

28 GHz can be estimated within an accuracy of ±1.5 dB on the basis of the 11-GHz terrestrial experiment records. For comparison, ATN_L and XPD_L at 28 GHz are also shown in the table.

Conclusions

The principal thrust of this paper is to provide an analytic description of the functional dependence of the electric field on frequency, propagation angle, rain rate, path length, and random canting conditions. The description provided herein allows propagation specialists to perform in-depth theoretical calculations from which empirical curves suitable for various engineering applications can be generated. The applicability of these curves to many of the recognized experimental results is currently being evaluated.

Acknowledgment

The authors wish to express their gratitude to Dr. T. Oguchi for supplying unpublished data.

References

[1] T. Oguchi and Y. Hosoya, "Scattering Properties of Oblate Raindrops and Cross Polarization of Radio Waves Due to Rain, (Part II): Calculation at Microwave and Millimeter Wave Region," *Journal of the Radio Research Laboratories*, Vol. 21, No. 105, 1974, pp. 191–259.

[2] J. W. Ryde and D. Ryde, "Attenuation of Centimeter and Millimeter Wave by Rain, Hail, Fogs, and Clouds," Report 8670, General Electric Co. Research Labs, Wembley, England, May 1945.

[3] T. Oguchi, "Attenuation and Phase Rotation of Radio Waves Due to Rain: Calculations at 19.3 and 34.8 GHz," *Radio Science*, Vol. 8, No. 1, 1973, pp. 31–38.

[4] J. A. Morison and M. J. Cross, "Scattering of a Plane Electromagnetic Wave by Axisymmetric Raindrops," *Bell System Technical Journal*, Vol. 53, No. 6, 1974, pp. 955–1019.

[5] P. A. Watson and M. Arbabi, "Rainfall Crosspolarisation at Microwave Frequencies," *Proc. IEE*, Vol. 120, No. 4, 1973, pp. 413–418.

[6] B. G. Evans and J. Troughton, "Calculation of Cross-Polarization Due to Precipitation," *IEEE* Conference on Propagation of Radio Waves at Frequencies Above 10 GHz, April 1973, Conference Publication No. 98, pp. 162–171.

[7] T. S. Chu, "Rain-Induced Cross-Polarization at Centimeter and Millimeter Wavelengths," *Bell System Technical Journal*, Vol. 53, No. 8, 1974, pp. 1557–1579.

[8] D. T. Thomas, "Cross-Polarization Distortion in Microwave Radio Transmission Due to Rain," *Radio Science*, Vol. 6, No. 10, 1971, pp. 833–839.

[9] W. J. Vogel, B. M. Fannin, and A. W. Straiton, "Polarization Effects for Millimeter Wave Propagation in Rain," Technical Report 73-1, University of Texas, Austin, December 1973.

[10] J. S. Marshall and W. Palmer, "The Distribution of Raindrops with Size," *Journal of Meteorology*, Vol. 5, 1948, pp. 165–166.

[11] J. Joss and A. Waldvogel, "Raindrop Size Distribution and Doppler Velocities," 14th Radar Meteorology Conference, American Meteorological Society, Boston, Massachusetts, 1970, pp. 153–156.

[12] R. G. Medhurst, "Rainfall Attenuation of Centimeter Waves: Comparison of Theory and Measurement," *IEEE Transactions on Antennas and Propagation*, AP-13, No. 4, 1965, pp. 550–564.

[13] T. Oguchi, Private Communications.

[14] D. J. Fang, "Attenuation and Phase Shift of Microwaves Due to Canted Raindrops," *COMSAT Technical Review*, Vol. 5, No. 1, Spring 1975, pp. 135–156.

[15] R. R. Taur, "Rain Depolarization: Theory and Experiment," *COMSAT Technical Review*, Vol. 4, No. 1, Spring 1974, pp. 187–190.

[16] G. Brussaard, "Rain Induced Cross Polarization and Raindrop Canting," *Electronics Letters*, Vol. 10, No. 20, 1974, pp. 411–412.

[17] M. J. Saunders, "Cross Polarization at 18 and 30 GHz Due to Rain," *IEEE Transactions on Antennas and Propagation*, AP-19, 1971, pp. 273–277.

[18] D. J. Fang and J. M. Harris, "A New Method of Estimating Microwave Attenuation Over a Slant Propagation Path Based on Rain Gauge Data," *IEEE Transactions on Antennas and Propagation*, AP-24, 1976, pp. 381–384.

Appendix A. Derivation of Equation (8)

Equation (8) will be identical to equation (3) if

$$F_p(f, R, L)\, R_p(f, R, \psi, L) = \frac{1}{2}\{\exp[-iK_H(f, R, \psi)\, L]$$

$$+ \exp[-iK_V(f, R, \psi)\, L]\} \qquad \text{(A-1)}$$

$$F_m(f, R, L)\, R_m(f, R, \psi, L) = \frac{1}{2}\{\exp[-iK_H(f, R, \psi)\, L]$$

$$- \exp[-iK_V(f, R, \psi)\, L]\} \quad . \qquad \text{(A-2)}$$

Based on these equations and equations (4a) and (4b), $F_p R_p$ and $F_m R_m$ are calculated as shown in Tables A-1 through A-4 using theoretical data supplied by Oguchi [A1], [A2] and Chu [A3].

It can be seen the last columns in the tables that, for rain rates less than 50 mm/hr,

$$\left| \frac{F_m(f, R, 1) \ R_m(f, R, \psi, 1)}{F_p(f, R, 1) \ R_p(f, R, \psi, 1)} \right| \ll 1 \quad . \tag{A-3}$$

Taylor series approximation then implies that

$$F_p(f, R, L) = \frac{1}{R_p(f, R, \psi, L)} \, [F_p(f, R, 1) \ R_p(f, R, \psi, 1)]^L$$

$$F_m(f, R, L) = \frac{L}{R_m(f, R, \psi, L)} \, [F_p(f, R, 1) \ R_p(f, R, \psi, 1)]^{L-1}$$

$$\cdot F_m(f, R, 1) \ R_m(f, R, \psi, 1) \quad .$$

These equations can be readily reduced to yield equation (7). Equation (7) is valid even for rain rates greater than 50 mm/hr, for which inequality (A-3) is, in a few cases, no longer appropriate. This is because the effective path length, L, for storms of $R \geq 50$ mm/hr is generally short, e.g., less than or very close to unity [A4]–[A6].

The values of $a(f)$, $b(f)$, $m(f)$, and $n(f)$ shown in Figure 2 are generated by trial and error. Their validity can best be shown by tabulating the ratio of either $F_p R_p$ or $F_m R_m$ generated by using equations (A-1) and (A-2) to that generated by using equations (5) and (6); i.e.,

$$C(f, R, \psi)$$

$$= \frac{(1/2)\{\exp[-iK_H(\psi)] + \exp[-iK_V(\psi)]\} \exp\{[aR^m + ibR^n]\cos^2\psi\}}{(1/2)\{\exp[-iK_H(90)] + \exp[-iK_V(90)]\}} \tag{A-4}$$

$$D(f, R, \psi)$$

$$= \frac{(1/2)\{\exp[-iK_H(\psi)] - \exp[-iK_V(\psi)]\} \exp\{[aR^m + ibR^n]\cos^2\psi\}}{(1/2)\{\exp[-iK_H(90)] - \exp[-iK_V(90)]\} \sin^2\psi} \tag{A-5}$$

in Tables A-5 and A-6 for $\psi = 70°$, $50°$, and $30°$ based on the available data at $f = 4, 6, 11, 30,$ and 34.8 GHz. For practical purposes, both C and D are unity in magnitude and zero in phase. Therefore, F_p and F_m at $L = 1$ km are indeed independent of the propagation angle ψ as postulated in the paper, and $F_p R_p$ and $F_m R_m$ given by equations (5) and (6) are indeed consistent with those defined by equations (A-1) and (A-2). Equation (7) implies further that F_p and F_m are insensitive to ψ at longer path lengths.

TABLE A-1. VALUES OF $F_p R_p$ AND $F_m R_m$ AT $\psi = 90°$ AND $L = 1$ km[a]

f (GHz)	R (mm)	$F_p R_p = A_p \angle \theta_p^\circ$		$F_m R_m = A_m \angle \theta_m^\circ$		(A_m/A_p) $\times 100$
		A_p	θ_p	$A_m \times 100$	θ_m	
4.0[b]	0.25	0.99998	−0.145	0.00646	−91.084	0.00646
4.0	1.25	0.99990	−0.569	0.03491	−91.629	0.03491
4.0	2.50	0.99981	−1.035	0.07304	−92.175	0.07306
4.0	5.00	0.99962	−1.895	0.15531	−93.147	0.15537
4.0	12.50	0.99901	−4.236	0.42121	−95.687	0.42163
4.0	25.00	0.99783	−7.897	0.90605	−99.637	0.90801
4.0	50.00	0.99499	−14.964	1.97267	−107.110	1.98260
4.0	100.00	0.98782	−28.620	4.27463	−121.371	4.32736
4.0	150.00	0.97862	−42.274	6.75161	−135.624	6.89912
4.0[c]	0.25	0.99997	−0.144	0.00576	−91.245	0.00576
4.0	1.25	0.99989	−0.562	0.03133	−91.744	0.03134
4.0	2.50	0.99979	−1.021	0.06563	−92.293	0.06564
4.0	12.50	0.99892	−4.178	0.37673	−95.787	0.37714
4.0	25.00	0.99763	−7.792	0.81100	−99.716	0.81293
4.0	50.00	0.99458	−14.714	1.76358	−107.070	1.77319
4.0	100.00	0.98688	−28.290	3.83960	−121.297	3.89065
4.0	150.00	0.97799	−41.652	5.98334	−135.020	6.11799
5.0[c]	0.25	0.99996	−0.180	0.00724	−91.579	0.00725
5.0	1.25	0.99981	−0.706	0.03945	−92.342	0.03946
5.0	2.50	0.99962	−1.285	0.08291	−93.128	0.08295
5.0	12.50	0.99776	−5.299	0.48384	−98.117	0.48492
5.0	25.00	0.99463	−9.931	1.04237	−103.903	1.04801
5.0	50.00	0.98654	−18.867	2.26617	−114.300	2.29708
5.0	100.00	0.96448	−36.486	4.85359	−134.196	5.03232
5.0	150.00	0.93737	−53.903	7.42621	−151.561	7.92238
6.0[c]	0.25	0.99994	−0.217	0.00864	−92.016	0.00864
6.0	1.25	0.99969	−0.853	0.04793	−93.087	0.04795
6.0	2.50	0.99935	−1.556	0.10128	−94.243	0.10134
6.0	12.50	0.99554	−6.463	0.59128	−101.457	0.59393
6.0	25.00	0.98870	−12.139	1.26822	−108.763	1.28272
6.0	50.00	0.97079	−23.118	2.70076	−121.885	2.78203
6.0	100.00	0.92386	−44.665	5.58868	−145.987	6.04925
6.0	150.00	0.87097	−65.928	8.28603	−168.854	9.51360
11.0[b]	0.25	0.99971	−0.409	0.01837	−94.745	0.01838
11.0	1.25	0.99804	−1.638	0.10360	−98.901	0.10381
11.0	2.50	0.99530	−3.003	0.21738	−102.286	0.21841
11.0	5.00	0.98860	−5.524	0.45541	−107.175	0.46066
11.0	12.50	0.96389	−12.373	1.19403	−117.541	1.23875
11.0	25.00	0.91601	−22.739	2.40206	−130.582	2.62230
11.0	50.00	0.81422	−41.876	4.54823	−152.442	5.58603
11.0	100.00	0.62879	−76.128	7.39110	160.355	11.75452
11.0	150.00	0.48101	−108.055	8.84560	136.984	18.38966

TABLE A-1. VALUES OF $F_p R_p$ AND $F_m R_m$ AT $\psi = 90°$ AND $L = 1$ km[a]
(Continued)

f (GHz)	R (mm)	$F_p R_p = A_p \angle \theta_p°$		$F_m R_m = A_m \angle \theta_m°$		(A_m/A_p) × 100
		A_p	θ_p	$A_m \times 100$	θ_m	
11.0 [e]	0.25	0.99970	−0.406	0.01645	−94.895	0.01646
11.0	1.25	0.99804	−1.612	0.09134	−99.020	0.09152
11.0	2.50	0.99534	−2.954	0.19363	−102.234	0.19453
11.0	12.50	0.96488	−12.153	1.07188	−117.079	1.11089
11.0	25.00	0.91851	−22.366	2.16181	−130.000	2.35361
11.0	50.00	0.82103	−41.113	4.11145	−151.401	5.00765
11.0	100.00	0.63821	−75.306	6.79138	171.350	10.64131
11.0	150.00	0.49380	−106.965	8.13634	138.369	16.47713
15.0 [e]	0.25	0.99929	−0.560	0.02295	−97.971	0.02297
15.0	1.25	0.99521	−2.214	0.12691	−104.093	0.12752
15.0	2.50	0.98895	−4.026	0.26547	−108.391	0.26844
15.0	12.50	0.92775	−15.970	1.41701	−127.326	1.52736
15.0	25.00	0.84782	−28.777	2.75795	−143.706	3.25300
15.0	50.00	0.70027	−51.722	4.87814	−170.737	6.96607
15.0	100.00	0.47344	−92.789	7.07185	143.428	14.93720
15.0	150.00	0.32042	−131.100	7.55109	102.172	23.56615
18.1 [b]	0.25	0.99882	−0.685	0.03139	−100.311	0.03142
18.1	1.25	0.99211	−2.704	0.17311	−107.850	0.17449
18.1	2.50	0.98222	−4.879	0.35940	−113.218	0.36590
18.1	5.00	0.96100	−8.771	0.74062	−120.563	0.77067
18.1	12.50	0.89382	−18.830	1.85267	−135.630	2.07274
18.1	25.00	0.78729	−33.431	3.46476	−154.934	4.40088
18.1	50.00	0.60476	−59.317	5.69326	173.801	9.41415
18.1	100.00	0.35974	−104.617	7.15352	121.852	19.88509
18.1	150.00	0.21336	−146.474	6.57781	75.268	30.82904
19.3 [e]	0.25	0.99862	−0.724	0.03006	−101.412	0.03011
19.3	1.25	0.99107	−2.829	0.16515	−109.430	0.16664
19.3	2.50	0.98013	−5.092	0.34476	−114.994	0.35175
19.3	12.50	0.88430	−19.555	1.76234	−138.935	1.99292
19.3	25.00	0.77035	−34.667	3.25863	−159.281	4.23008
19.3	50.00	0.57996	−61.203	5.19945	168.244	8.96518
19.3	100.00	0.32830	−107.976	6.21271	114.330	18.92400
19.3	150.00	0.18672	−150.795	5.39867	66.444	28.91396
30.0 [b]	0.25	0.99597	−1.131	0.05351	−110.725	0.05373
30.0	1.25	0.97622	−4.281	0.29025	−124.634	0.29732
30.0	2.50	0.95017	−7.503	0.58618	−133.906	0.61692
30.0	5.00	0.89889	−13.035	1.13888	−146.144	1.26698
30.0	12.50	0.75812	−26.380	2.50140	−168.936	3.29949
30.0	25.00	0.57469	−44.317	3.80753	165.360	6.62534
30.0	50.00	0.33664	−73.600	4.43103	127.708	13.16235
30.0	100.00	0.12799	−121.100	3.18444	71.499	24.87996
30.0	150.00	0.05212	−163.384	1.83971	24.612	35.29517

TABLE A-1. VALUES OF $F_p R_p$ AND $F_m R_m$ AT $\psi = 90°$ AND $L = 1$ km[a]
(Continued)

f (GHz)	R (mm)	$F_p R_p = A_p \angle \theta_p$		$F_m R_m = A_m \angle \theta_m$		(A_m/A_p) × 100
		A_p	θ_p	$A_m \times 100$	θ_m	
34.8 [e]	0.25	0.99449	−1.291	0.05714	−115.605	0.05746
34.8	1.25	0.96852	−4.776	0.30277	−132.330	0.31261
34.8	2.50	0.93520	−8.261	0.59817	−142.848	0.63961
34.8	12.50	0.70617	−27.695	2.28977	179.510	3.24253
34.8	25.00	0.50697	−45.329	3.20918	153.757	6.33008
34.8	50.00	0.27447	−73.097	3.33174	117.729	12.13888
34.8	100.00	0.09269	−118.005	2.06967	64.763	22.32809
34.8	150.00	0.03454	−158.563	1.08545	20.864	31.42561
50.0 [e]	0.25	0.98766	−1.765	0.08238	−133.791	0.08341
50.0	1.25	0.93719	−5.958	0.38321	−156.421	0.40889
50.0	2.50	0.88041	−9.724	0.69928	−168.939	0.79426
50.0	12.50	0.57446	−27.463	1.99070	154.122	3.46533
50.0	25.00	0.36766	−41.719	2.34363	132.412	6.37453
50.0	50.00	0.17105	−63.344	1.97056	104.039	11.52043
50.0	100.00	0.04621	−101.845	0.95265	61.317	20.61481
50.0	150.00	0.01388	−137.956	0.40228	23.459	28.97725

[a] The values are equivalent to those of F_p and F_m plotted in Figure 3, since $R_p = R_m = 1$ at $\psi = 90°$.
[b] Original inputs of $A_{H,V}$ and $\Phi_{H,V}$ are from Chu [A3].
[e] Original inputs of $A_{H,V}$ and $\Phi_{H,V}$ are from Oguchi [A2]. Oguchi's inputs at 19.3 and 34.8 GHz are also available in Reference A1.

TABLE A-2. VALUES OF $F_p R_p$ AND $F_m R_m$ AT $\psi = 70°$ AND $L = 1$ km

f (GHz)	R (mm)	$F_p R_p = A_p \angle \theta_p°$		$F_m R_m = A_m \angle \theta_m°$		(A_m/A_p) × 100
		A_p	θ_p	$A_m \times 100$	θ_m	
4.0 [a]	0.25	0.99998	−0.144	0.00506	−91.214	0.00506
4.0	1.25	0.99989	−0.564	0.02758	−91.752	0.02758
4.0	2.50	0.99979	−1.025	0.05725	−92.312	0.05726
4.0	12.50	0.99891	−4.202	0.33312	−95.803	0.33349
4.0	25.00	0.99763	−7.843	0.71605	−99.764	0.71775
4.0	50.00	0.99462	−14.819	1.54642	−107.192	1.55480
4.0	100.00	0.98709	−28.516	3.38269	−121.533	3.42692
4.0	150.00	0.97851	−41.995	5.28279	−135.361	5.39881
6.0 [a]	0.25	0.99994	−0.218	0.00768	−92.001	0.00768
6.0	1.25	0.99969	−0.856	0.04226	−93.093	0.04227
6.0	2.50	0.99935	−1.562	0.08992	−94.233	0.08998
6.0	12.50	0.99556	−6.498	0.52152	−101.491	0.52385
6.0	25.00	0.98878	−12.214	1.12067	−108.853	1.13338
6.0	50.00	0.97113	−23.280	2.37600	−122.096	2.44662
6.0	100.00	0.92498	−45.029	4.94299	−146.334	5.34391
6.0	150.00	0.87311	−66.513	7.33048	−169.425	8.39579

TABLE A-2. VALUES OF F_pR_p AND F_mR_m AT $\psi = 70°$ AND $L = 1$ km

f (GHz)	R (mm)	$F_pR_p = A_p \angle \theta_p°$		$F_mR_m = A_m \angle \theta_m°$		(A_m/A_p) $\times 100$
		A_p	θ_p	$A_m \times 100$	θ_m	
11.0 [a]	0.25	0.99970	−0.407	0.01461	−94.867	0.01462
11.0	1.25	0.99804	−1.618	0.08081	−99.052	0.08097
11.0	2.50	0.99535	−2.966	0.17076	−102.272	0.17156
11.0	12.50	0.96498	−12.234	0.94976	−117.113	0.98423
11.0	25.00	0.91872	−22.548	1.90992	−130.188	2.07889
11.0	50.00	0.82144	−41.519	3.63170	−151.817	4.42112
11.0	100.00	0.63868	−76.277	6.00329	170.399	9.39954
11.0	150.00	0.49417	−108.544	7.22188	136.922	14.61402
30.0 [b]	0.25	0.99597	−1.134	0.04789	−110.434	0.04808
30.0	1.25	0.97612	−4.301	0.25892	−124.388	0.26525
30.0	2.50	0.94989	−7.547	0.52297	−133.609	0.55056
30.0	5.00	0.89815	−13.131	1.01424	−145.899	1.12926
30.0	12.50	0.75591	−26.616	2.21834	−168.859	2.93465
30.0	25.00	0.57047	−44.770	3.35163	164.995	5.87524
30.0	50.00	0.33065	−74.458	3.85204	126.961	11.65004
30.0	100.00	0.12254	−122.575	2.71044	70.039	22.11876
30.0	150.00	0.04857	−165.294	1.52898	22.627	31.47713
34.8 [a]	0.25	0.99448	−1.295	0.05047	−115.606	0.05075
34.8	1.25	0.96836	−4.798	0.26770	−132.395	0.27645
34.8	2.50	0.93479	−8.308	0.52847	−142.942	0.56534
34.8	12.50	0.70361	−27.906	2.01374	179.098	2.86203
34.8	25.00	0.50271	−45.695	2.81474	153.168	5.59917
34.8	50.00	0.26937	−73.710	2.89259	116.849	10.73841
34.8	100.00	0.08891	−118.901	1.76422	63.521	19.84322
34.8	150.00	0.03235	−159.648	0.90481	19.198	27.96649

[a] Original inputs of $A_{H,V}$ and $\Phi_{H,V}$ are from Oguchi [A2].
[b] Original inputs of $A_{H,V}$ and $\Phi_{H,V}$ are from Chu [A3].

TABLE A-3. VALUES OF F_pR_p AND F_mR_m AT $\psi = 50°$ AND $L = 1$ km

f (GHz)	R (mm)	$F_pR_p = A_p \angle \theta_p°$		$F_mR_m = A_m \angle \theta_m°$		(A_m/A_p) $\times 100$
		A_p	θ_p	$A_m \times 100$	θ_m	
4.0 [a]	0.25	0.99997	−0.145	0.00340	−91.254	0.00340
4.0	1.25	0.99989	−0.569	0.01833	−91.761	0.01833
4.0	2.50	0.99979	−1.036	0.03840	−92.309	0.03841
4.0	12.50	0.99891	−4.262	0.22150	−95.865	0.22174
4.0	25.00	0.99763	−7.969	0.47649	−99.897	0.47762
4.0	50.00	0.99469	−15.090	1.02520	−107.468	1.03067
4.0	100.00	0.98755	−29.083	2.25279	−122.094	2.28120
4.0	150.00	0.97959	−42.871	3.52096	−136.226	3.59433

TABLE A-3. VALUES OF F_pR_p AND F_mR_m AT $\psi = 50°$ AND $L = 1$ km

f (GHz)	R (mm)	$F_pR_p = A_p \angle \theta_p°$		$F_mR_m = A_m \angle \theta_m°$		(A_m/A_p) $\times 100$
		A_p	θ_p	$A_m \times 100$	θ_m	
6.0 [a]	0.25	0.99994	−0.219	0.00515	−91.989	0.00515
6.0	1.25	0.99969	−0.864	0.02811	−93.093	0.02812
6.0	2.50	0.99935	−1.578	0.05937	−94.264	0.05941
6.0	12.50	0.99562	−6.587	0.34711	−101.570	0.34864
6.0	25.00	0.98902	−12.410	0.74718	−108.986	0.75548
6.0	50.00	0.97195	−23.693	1.58786	−122.471	1.63369
6.0	100.00	0.92762	−45.941	3.29477	−147.226	3.55184
6.0	150.00	0.87815	−67.971	4.89347	−170.842	5.57248
11.0 [a]	0.25	0.99970	−0.409	0.00963	−94.884	0.00963
11.0	1.25	0.99804	−1.632	0.05358	−99.086	0.05369
11.0	2.50	0.99537	−2.997	0.11354	−102.289	0.11407
11.0	12.50	0.96524	−12.429	0.62782	−117.407	0.65042
11.0	25.00	0.91923	−22.992	1.27130	−130.659	1.38300
11.0	50.00	0.82221	−42.554	2.41768	−152.843	2.94047
11.0	100.00	0.63943	−78.709	4.00041	167.966	6.25624
11.0	150.00	0.49472	−112.548	4.79492	132.996	9.69222
30.0 [b]	0.25	0.99595	−1.144	0.03138	−110.790	0.03150
30.0	1.25	0.97586	−4.359	0.17232	−124.493	0.17658
30.0	2.50	0.94917	−7.670	0.34792	−133.861	0.36656
30.0	5.00	0.89625	−13.383	0.67388	−146.397	0.75188
30.0	12.50	0.75019	−27.239	1.46699	−169.881	1.95549
30.0	25.00	0.55982	−45.967	2.19431	163.473	3.91967
30.0	50.00	0.31572	−76.648	2.46939	124.090	7.82156
30.0	100.00	0.11021	−126.238	1.64309	65.668	14.90927
30.0	150.00	0.04086	−170.061	0.87386	17.105	21.38531
34.8 [a]	0.25	0.99444	−1.305	0.03338	−115.821	0.03357
34.8	1.25	0.96796	−4.856	0.17806	−132.460	0.18395
34.8	2.50	0.93375	−8.426	0.35116	−143.185	0.37607
34.8	12.50	0.69710	−28.432	1.32892	178.215	1.90634
34.8	25.00	0.49209	−46.611	1.83803	151.751	3.73517
34.8	50.00	0.25692	−75.200	1.84379	114.694	7.17641
34.8	100.00	0.08023	−121.017	1.06577	60.832	13.28328
34.8	150.00	0.02763	−162.076	0.52206	15.883	18.89503

[a] Original inputs of $A_{H,V}$ and $\Phi_{H,V}$ are from Oguchi [A2].
[b] Original inputs of $A_{H,V}$ and $\Phi_{H,V}$ are from Chu [A3].

TABLE A-4. VALUES OF $F_p R_p$ AND $F_m R_m$ AT $\psi = 30°$ AND $L = 1$ km

f (GHz)	R (mm)	$F_p R_p = A_p \angle \theta_p°$		$F_m R_m = A_m \angle \theta_m°$		(A_m/A_p) $\times 100$
		A_p	θ_p	$A_m \times 100$	θ_m	
4.0 *	0.25	0.99997	−0.146	0.00148	−91.151	0.00148
4.0	1.25	0.99989	−0.575	0.00785	−91.739	0.00786
4.0	2.50	0.99979	−1.048	0.01658	−92.307	0.01658
4.0	12.50	0.99890	−4.331	0.09418	−95.943	0.09428
4.0	25.00	0.99763	−8.114	0.20209	−100.052	0.20257
4.0	50.00	0.99475	−15.395	0.44310	−107.743	0.44544
4.0	100.00	0.98791	−29.725	0.95831	−122.752	0.97004
4.0	150.00	0.98044	−43.864	1.49997	−137.209	1.52989
6.0 *	0.25	0.99994	−0.221	0.00218	−92.035	0.00218
6.0	1.25	0.99969	−0.872	0.01196	−93.100	0.01196
6.0	2.50	0.99935	−1.596	0.02532	−94.287	0.02534
6.0	12.50	0.99568	−6.689	0.14827	−101.678	0.14892
6.0	25.00	0.98925	−12.625	0.32154	−109.205	0.32503
6.0	50.00	0.97280	−24.165	0.67853	−122.900	0.69750
6.0	100.00	0.93031	−46.977	1.40764	−148.301	1.51308
6.0	150.00	0.88316	−69.608	2.09559	−172.511	2.37283
11.0 *	0.25	0.99970	−0.412	0.00411	−94.905	0.00411
11.0	1.25	0.99805	−1.649	0.02283	−99.026	0.02288
11.0	2.50	0.99540	−3.032	0.04864	−102.366	0.04864
11.0	12.50	0.96552	−12.645	0.27028	−117.540	0.27993
11.0	25.00	0.91975	−23.489	0.53943	−131.262	0.58650
11.0	50.00	0.82295	−43.712	1.03425	−154.086	1.25676
11.0	100.00	0.63965	−81.435	1.70377	165.119	2.66358
11.0	150.00	0.49423	−117.013	2.03878	128.468	4.12512
34.8 *	0.25	0.99440	−1.317	0.01429	−115.684	0.01437
34.8	1.25	0.96750	−4.922	0.07570	−132.559	0.07824
34.8	2.50	0.93255	−8.559	0.14959	−143.548	0.16041
34.8	12.50	0.68974	−29.018	0.56406	177.607	0.81779
34.8	25.00	0.48025	−47.621	0.76269	149.966	1.58810
34.8	50.00	0.24347	−76.791	0.75206	112.303	3.08891
34.8	100.00	0.07169	−123.147	0.40815	57.728	5.69331
34.8	150.00	0.02329	−164.524	0.18902	11.801	8.11571

* Original inputs of $A_{H,V}$ and $\Phi_{H,V}$ are from Oguchi [A2].

TABLE A-5. VALUES OF $C(\psi)$ AT $\psi = 70°$, $50°$, AND $30°$

f (GHz)	R (mm)	$C(70) = A_C \angle \theta_C°$		$C(50) = A_C \angle \theta_C°$		$C(30) = A_C \angle \theta_C°$	
		A_C	θ_C	A_C	θ_C	A_C	θ_C
4.0 a	2.5	1.00000	0.001	1.00000	0.001	0.99999	0.002
4.0	12.5	1.00000	0.001	0.99999	0.003	0.99998	0.006
4.0	25.0	1.00000	0.001	1.00001	0.005	1.00000	0.009
4.0	50.0	1.00004	0.002	1.00012	0.005	1.00017	0.009
4.0	100.0	1.00022	−0.002	1.00068	−0.000	1.00105	0.004
4.0	150.0	1.00053	0.002	1.00163	−0.000	1.00251	−0.001
6.0 a	2.5	1.00000	−0.001	1.00000	−0.003	0.99999	−0.006
6.0	12.5	1.00002	−0.001	1.00008	−0.005	1.00014	−0.011
6.0	25.0	1.00009	−0.001	1.00032	−0.009	1.00056	−0.011
6.0	50.0	1.00035	0.002	1.00119	0.003	1.00207	0.001
6.0	100.0	1.00120	−0.004	1.00407	−0.005	1.00698	−0.005
6.0	150.0	1.00246	−0.013	1.00825	−0.025	1.01400	−0.016
11.0 a	2.5	1.00001	0.001	1.00003	0.002	1.00005	0.004
11.0	12.5	1.00010	0.004	1.00037	0.023	1.00066	0.052
11.0	25.0	1.00023	0.009	1.00079	0.048	1.00136	0.099
11.0	50.0	1.00050	0.024	1.00143	0.075	1.00233	0.152
11.0	100.0	1.00074	−0.006	1.00191	0.007	1.00226	0.060
11.0	150.0	1.00077	−0.027	1.00187	−0.103	1.00089	−0.101
30.0 b	2.5	1.00021	0.070	1.00074	0.235		
30.0	5.0	1.00 32	0.089	1.00112	0.305		
30.0	12.5	1.00047	0.115	1.00145	0.381		
30.0	25.0	1.00029	0.118	1.00089	0.365		
30.0	50.0	0.99941	0.069	0.99724	0.226		
30.0	100.0	0.99588	0.031	0.98960	0.182		
30.0	150.0	0.99308	0.090	0.98133	0.388		
34.8 a	2.5	1.00024	0.128	1.00085	0.451	1.00152	0.820
34.8	12.5	1.00030	0.155	1.00097	0.555	1.00167	1.021
34.8	25.0	0.99992	0.137	0.99976	0.494	0.99952	0.933
34.8	50.0	0.99906	0.079	0.99688	0.341	0.99441	0.741
34.8	100.0	0.99624	0.056	0.98967	0.349	0.98631	0.959
34.8	150.0	0.99362	0.062	0.98529	0.538	0.98435	1.391

a Original inputs of $A_{H,V}$ and $\Phi_{H,V}$ are from Oguchi [A2].
b Original inputs of $A_{H,V}$ and $\Phi_{H,V}$ are from Chu [A3].

TABLE A-6. VALUES OF $D(\psi)$ AT $\psi = 70°, 50°$, AND $30°$

f (GHz)	R (mm)	$D(70) = A_D \angle \theta_D°$		$D(50) = A_D \angle \theta_D°$		$D(30) = A_D \angle \theta_D°$	
		A_D	θ_D	A_D	θ_D	A_D	θ_D
4.0 [a]	2.5	0.98790	−0.015	0.99706	0.000	1.01063	0.014
4.0	12.5	1.00138	0.010	1.00191	0.010	0.99995	0.003
4.0	25.0	0.99988	0.003	1.00121	0.001	0.99676	−0.005
4.0	50.0	0.99303	−0.014	0.99062	−0.018	1.00500	0.018
4.0	100.0	0.99771	−0.011	0.99983	−0.004	0.99835	−0.016
4.0	150.0	0.99988	0.004	1.00279	0.012	1.00276	0.023
6.0 [a]	2.5	1.00555	0.015	0.99895	−0.002	1.00000	−0.009
6.0	12.5	0.99887	−0.000	1.00039	0.005	1.00306	−0.005
6.0	25.0	1.00072	−0.016	1.00397	0.038	1.01414	0.033
6.0	50.0	0.99630	−0.048	1.00189	−0.009	1.00495	0.032
6.0	100.0	1.00163	0.013	1.00463	0.032	1.00749	−0.006
6.0	150.0	1.00187	0.000	1.00638	0.031	1.01162	0.007
11.0 [a]	2.5	0.99873	−0.025	0.99928	−0.009	1.00018	−0.049
11.0	12.5	1.00346	0.051	0.99811	−0.029	1.00863	0.082
11.0	25.0	1.00052	0.002	1.00212	0.015	0.99812	−0.040
11.0	50.0	1.00033	0.013	1.00206	0.074	1.00621	0.066
11.0	100.0	1.00106	0.015	1.00378	0.026	1.00349	−0.041
11.0	150.0	1.00519	0.105	1.00426	0.108	1.00231	0.046
30.p [b]	2.5	1.01086	0.411	1.01326	0.447		
30.0	5.0	1.00970	0.430	1.01241	0.400		
30.0	12.5	1.00773	0.428	1.01142	0.296		
30.0	25.0	1.00456	0.205	1.00906	0.129		
30.0	50.0	1.00177	0.179	1.00984	−0.344		
30.0	100.0	1.00264	0.046	1.01055	−0.511		
30.0	150.0	1.00298	0.014	1.01323	−0.442		
34.4 [a]	2.5	1.00121	0.081	1.00279	0.279	1.00467	0.418
34.8	12.5	0.99988	−0.046	1.00284	−0.003	1.01051	0.442
34.8	25.0	1.00163	−0.086	1.00529	−0.229	1.00304	−0.566
34.8	50.0	1.00088	−0.189	1.00430	−0.592	1.01217	−0.990
34.8	100.0	1.00265	−0.290	1.00331	−0.569	1.00597	−0.933
34.8	150.0	1.00139	−0.518	1.00953	−0.931	1.01684	−1.710

[a] Original inputs of $A_{H,V}$ and $\Phi_{H,V}$ are from Oguchi [A2].
[b] Original inputs of $A_{H,V}$ and $\Phi_{H,V}$ are from Chu [A3].

References

[A1] T. Oguchi, "Attenuation and Phase Rotation of Radio Waves Due to Rain: Calculations at 19.3 and 34.8 GHz," *Radio Science*, Vol. 8, No. 1, 1973, pp. 31–38.

[A2] T. Oguchi, Private Communications.

[A3] T. S. Chu, "Rain-Induced Cross-Polarization at Centimeter and Millimeter Wavelengths," *Bell System Technical Journal*, Vol. 53, No. 8, 1974, pp. 1557–1579.

[A4] L. J. Ippolito, "Summary and Evaluation of Results from the ATS Millimeter Wave Experiment," *Earth-Satellite Propagation Above 10 GHz*, NASA Document X-751–72–208, May 1972.

[A5] H. W. Evans, "Attenuation on Earth-Space Paths at Frequencies up to 30 GHz," International Conference on Communications, June 1971, Montreal, Canada, *Proc.*, pp. 27–1 to 27–5.

[A6] International Radio Consultative Committee (C.C.I.R.), "Propagation Data Required for Space Telecommunications Systems," U.K. Contribution, Document S/154-E, October 1973.

Dickson J. Fang received his B.S.E.E. in 1962 from Taiwan University, Taipei, China, and his M.S.E.E. and Ph.D. from Stanford University in 1964 and 1967. From 1967 to 1674 he was a radio physicist at Stanford Research Institute, and he authored 17 papers on wave propagation, radio auroras, and ionospheric irregularities. While on leave from Stanford (1969–1970) he was a visiting professor at Taiwan University, where he published a graduate textbook on electromagnetic theory and five articles on the college education system. He joined COMSAT *in 1974 and is currently responsible for research on propagation over 10-GHz frequencies.*

Jiausen Jih received a B.S. in mechanical engineering from National Taiwan University in 1956, an M.S. in aeronautical engineering from the University of Maryland in 1962, and a D.Sc. in mechanical engineering from George Washington University in 1976. In January 1969 he joined COMSAT *as a member of the technical staff. He is presently Manager of the Engineering Applications Department of the Computer Center. He is a member of ACM.*
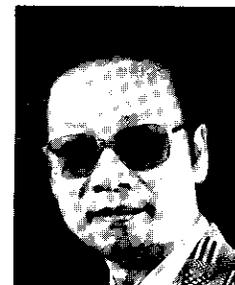
CTR Notes

TV cochannel interference on a PCM-PSK SCPC system

D. Kurjan and M. Wachs

(Manuscript received May 12, 1976)

Measurements have been made while investigating the effects of co-channel TV frequency modulating an interfering carrier on a multicarrier single-channel-per-carrier (SCPC) system with PCM-PSK modulation. This note presents results which deal with a single voice channel.

Experimental setup

The SCPC system under consideration would consist of several hundred 64-kbps 4-phase PSK channel slots in a single transponder. Since it would be impractical to generate that many carriers, multicarrier operation was simulated by using six 32-kbps 2-phase PSK carriers independently PSK modulated and clocked. The total power in these six carriers was adjusted so that the operating point of the traveling wave tube amplifier (TWTA) was at an input backoff of 10 dB. The SCPC carrier under test was located

at a frequency corresponding to a carrier-to-intermodulation (C/I) ratio of approximately 20 dB.

Figure 1 is a schematic of the experimental setup. An SCPC carrier is generated and up-converted to 4 GHz along with the six PSK cotransponder carriers. The up-converter is linear, providing a 2-carrier C/I greater than 40 dB. After amplification by the TWT, the composite is down-converted to 70 MHz. A common local oscillator is used for up- and down-conversion to eliminate translation oscillator instability. The TV interference and thermal noise are added at IF on the down-link. The TV interference is centered on the desired SCPC channel frequency and is represented as a frequency modulated carrier driven by a triangular dispersal wave with a peak-to-peak deviation of 1 MHz.

Test results

The performance of the SCPC modem in a back-to-back configuration without added interference is shown in the bit-error rate (BER) vs carrier-to-noise (C/N) ratio curve of Figure 2. It can be seen that an error probability (P_e) of 10^{-4} occurs at a C/N of 11.65 dB. The corresponding ratio of energy per bit to noise density (E_b/N_o) is computed as follows:

$$E_b/N_o = (C/N) BT_b$$

where B = channel bandwidth = 38 kHz

$$T_b = \text{information bit period} = \frac{1}{64 \text{ kbps}}$$

$$= 1.56 \times 10^{-5} \text{ seconds.}$$

Then,

$$(E_b/N_o)_{\text{meas}} = 11.65 - 2.26 = 9.4 \text{ dB}$$

while theoretical calculations yield

$$(E_b/N_o)_{\text{th}} = 8.4 \text{ dB} \quad .$$

The performance of the SCPC channel through the INTELSAT IV satellite simulator without external (i.e., TV) interference is shown in Figure 3. The BER vs C/N curve for a C/I of 30 dB is the same as that for the back-to-back case (Figure 2). For C/I ratios of 20 and 15 dB, a P_e of 10^{-4}
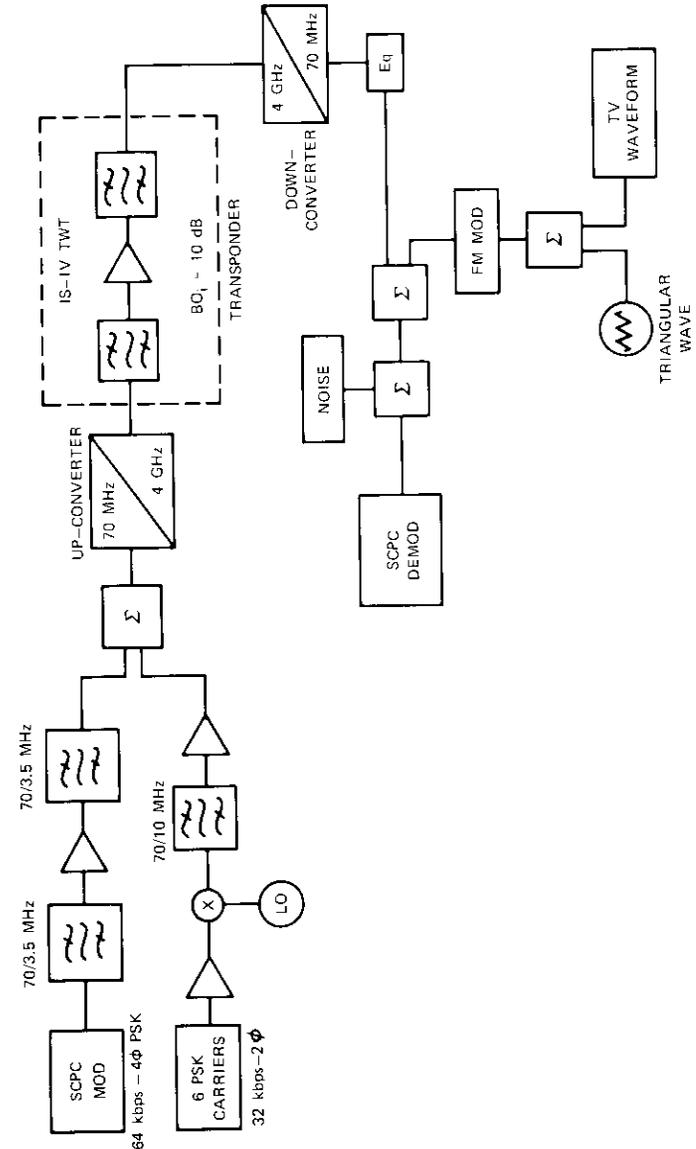
Figure 1. Block Diagram of Experimental Setup

requires C/N ratios of 12.45 and 16.6 dB, respectively. It should be noted that C/I refers to the modulated carrier power relative to the intermodulation power measured after the demodulator input filter without added thermal noise or interference.
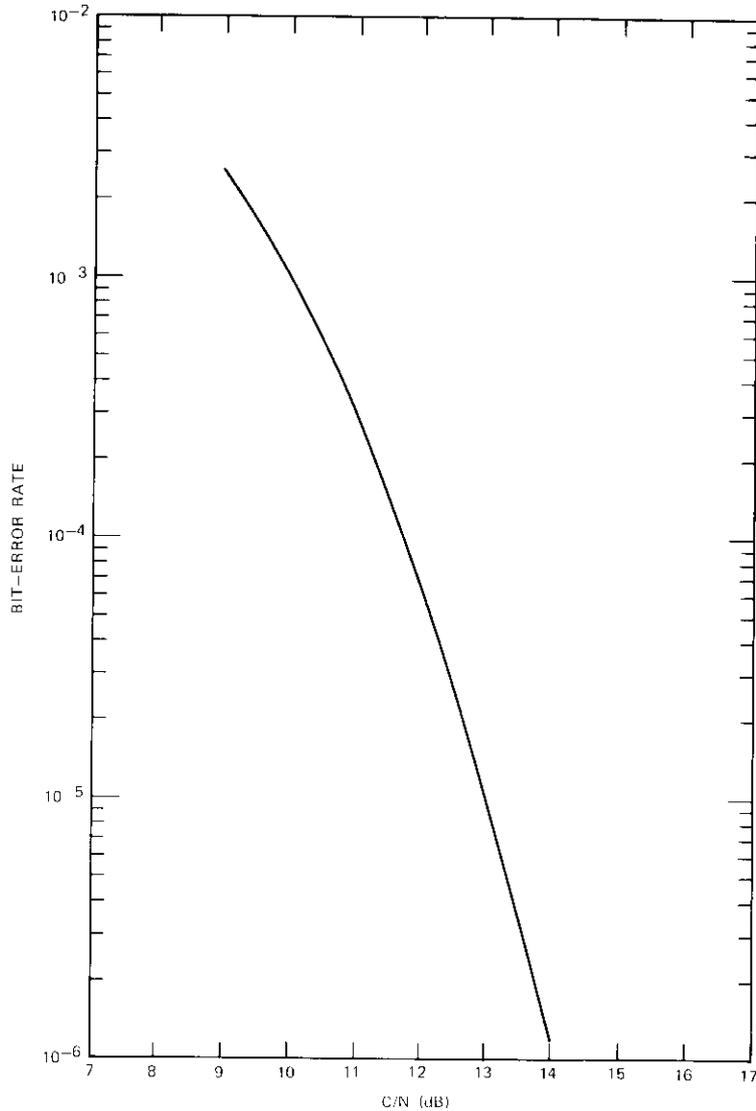
Figure 2. *SCPC Modem Back-to-Back Performance*
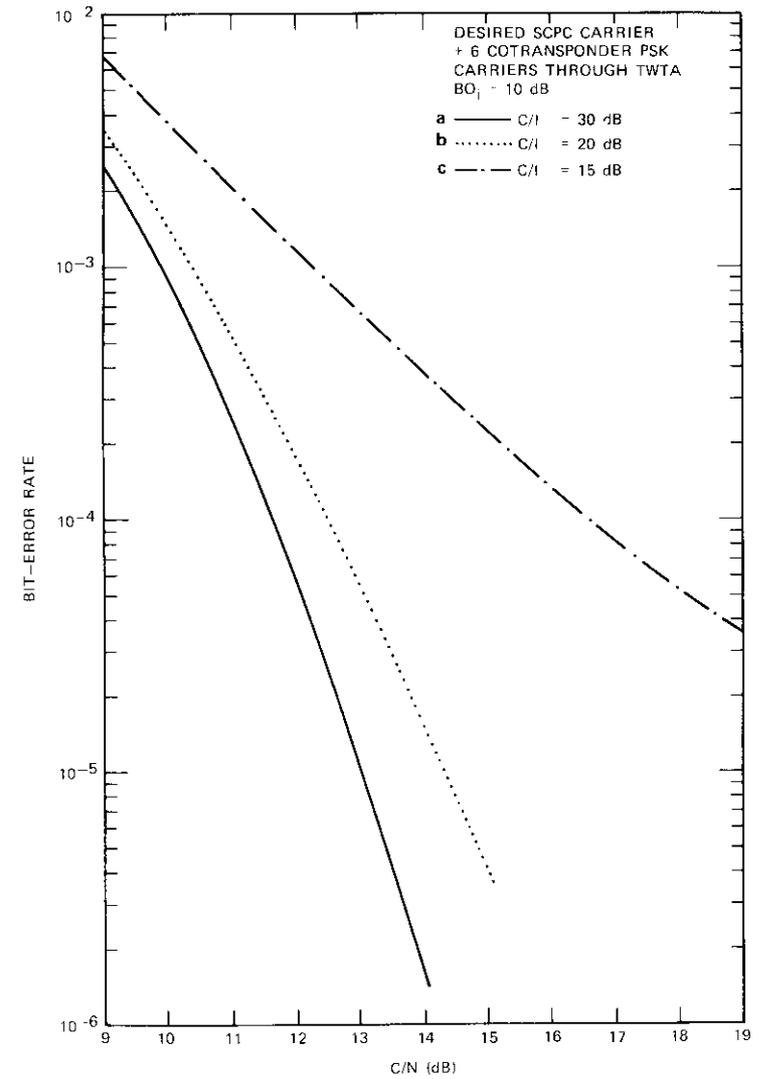
Figure 3. SCPC Performance Through a Transponder

The carrier-to-interference (C/INT) ratio is defined as the ratio of modulated carrier power to the power of an unmodulated external inter-

ference source at the demodulator input without added thermal noise (or intermodulation, in the case of transponder operation). The SCPC performance in a back-to-back mode with a frequency dispersal (TV) interference source added is shown in curves *a–c* of Figure 4. It can be seen that a C/INT of 10 dB yields an increase in the required C/N of less than 0.75 dB for a 10^{-4} bit-error probability. The average modulated (sweeping) interference power appearing after the demodulator input filter is 14.2 dB less than the power of the unmodulated interference source, since the swept interference source is in the channel bandwidth 3.8 percent of the time in these tests. Therefore, for a C/INT of 10 dB, the average power ratio of modulated carrier to modulated interference source at the channel filter output is 24.2 dB.

Curve *d* in Figure 4 shows the result of an unmodulated interference source at 70 MHz with the same C/INT (24 dB). In this case, performance is seen to be better than for the modulated interference source, even though the average power levels of the interference source at the channel filter output are the same. This performance difference is a result of transient effects in the demodulator caused by the spreading wave sweeping through the channel rather than being a stationary carrier within the band. Performance through the transponder with the spreading wave interference source added is shown in Figure 5, which considers the case of a 20-dB C/I for a range of interference levels. As can be seen from the figure, the presence of the swept interference source results in a further performance degradation of 0.5 dB for a C/INT of 15 dB, while a C/INT of 10 dB yields a 1.2-dB degradation for a P_e of 10^{-4}.

The effect of the modulating frequency (f_m) of the interference source is shown in Figure 6. It can be seen that the average BER does not vary significantly from 20 to 1,200 Hz. Even though the average BER is almost constant, the actual number of errors per interference sweep will vary directly with the time during which the interference source is in the channel bandwidth. No periodic effects were noticed in the quality of the channel over the entire range of sweep frequencies.

A "window" error rate, which corresponds to the error rate in the period during which the interference source sweeps through the channel bandwidth, can be computed from Figures 3 and 5 by eliminating the error rate due to noise and intermodulation alone. The resulting window error rate curves are shown in Figure 7. The fact that these error rates are not constant indicates that the additional errors in this period are not independent of the thermal noise.

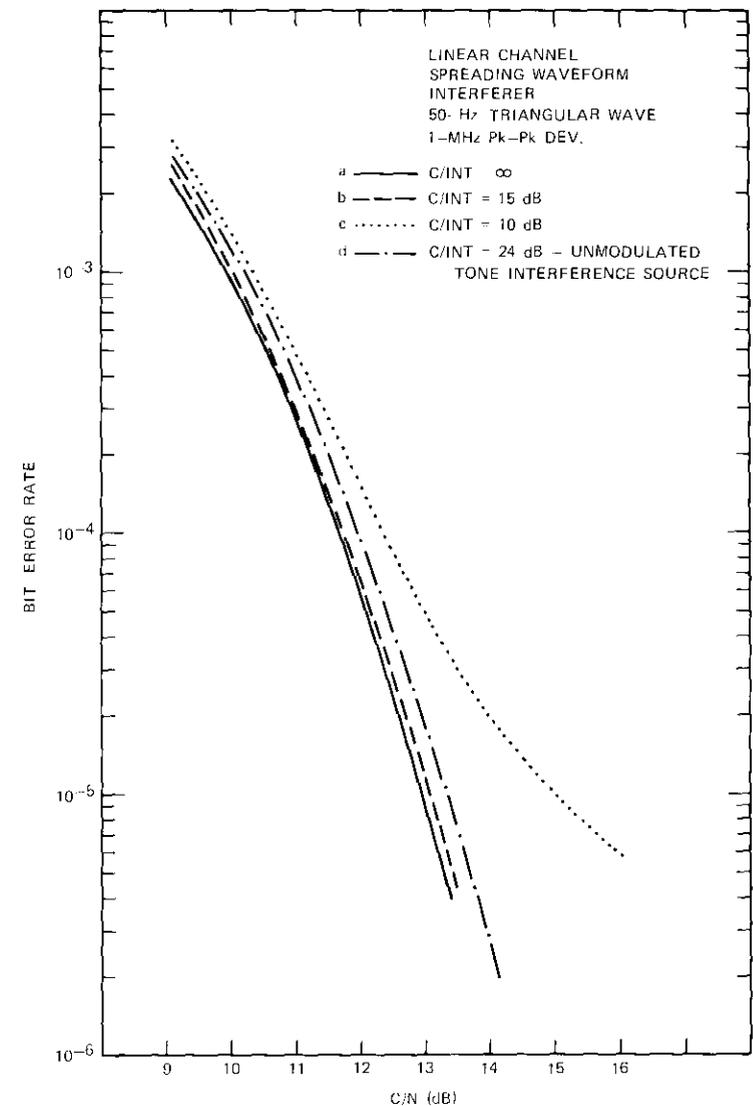For a typical operating condition of C/N = 15 dB and C/I ∼ 20 dB,

Figure 4. *SCPC Performance with Spreading Wave (back to back)*

C/INT was varied to determine the point at which the synchronization became unreliable. It was found that at C/INT = 9 dB occasional bursts

DESIRED SCPC CARRIER + 6 COTRANSPONDER
PSK CARRIERS THROUGH TWTA
BO_i – 10 dB
50 Hz; 1 MHz Pk Pk DEV.
TRIANGULAR SPREADING WAVE
C/I = 20 dB

a ——— C/INT – ∞
b — — — C/INT – 20 dB
c ········· C/INT – 15 dB
d —·—· C/INT – 10 dB
e —··—·· C/INT – 8 dB
f —···—··· C/INT = 6 dB
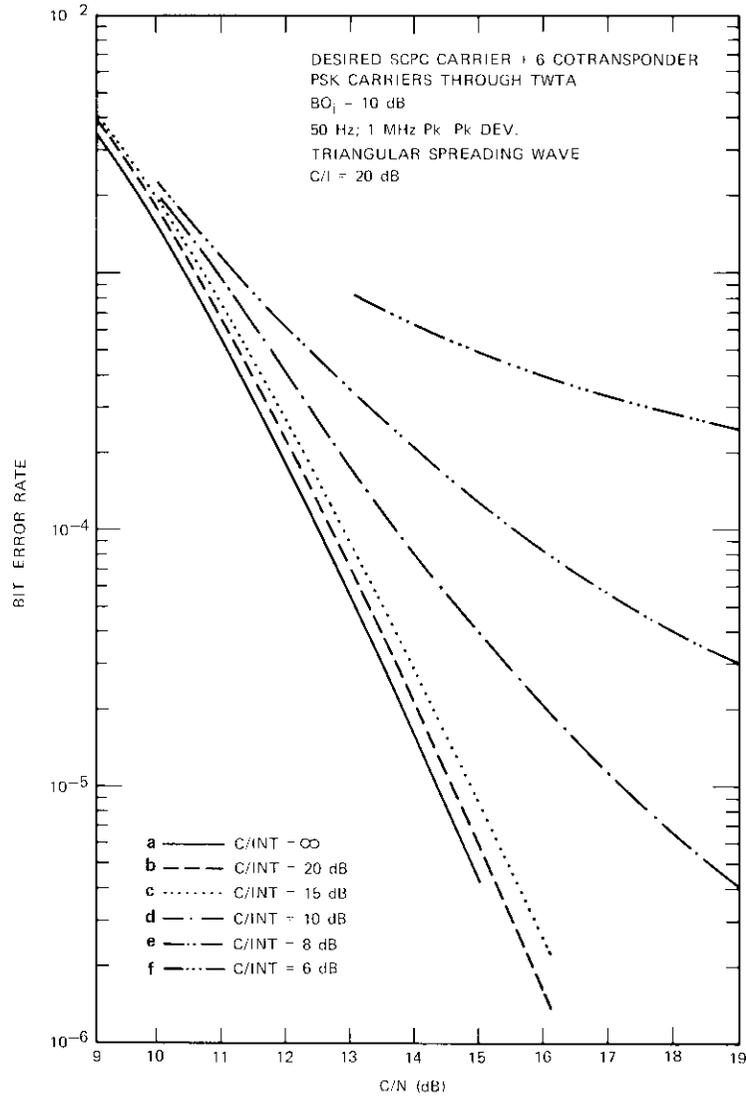
Figure 5. *SCPC Performance Through a Transponder Plus Spreading Wave Interference*

of approximately 100 errors occurred, corresponding to the loss of start-of-message synchronization.

DESIRED SCPC CARRIER + 6 COTRANSPONDER
PSK CARRIERS THROUGH TWTA
BO_i – 10 dB
C/I = 19 dB
C/N   15 dB
C/INT ≈ 10 dB
50 Hz; 1–MHz Pk–Pk DEV.
TRIANGULAR SPREADING WAVE
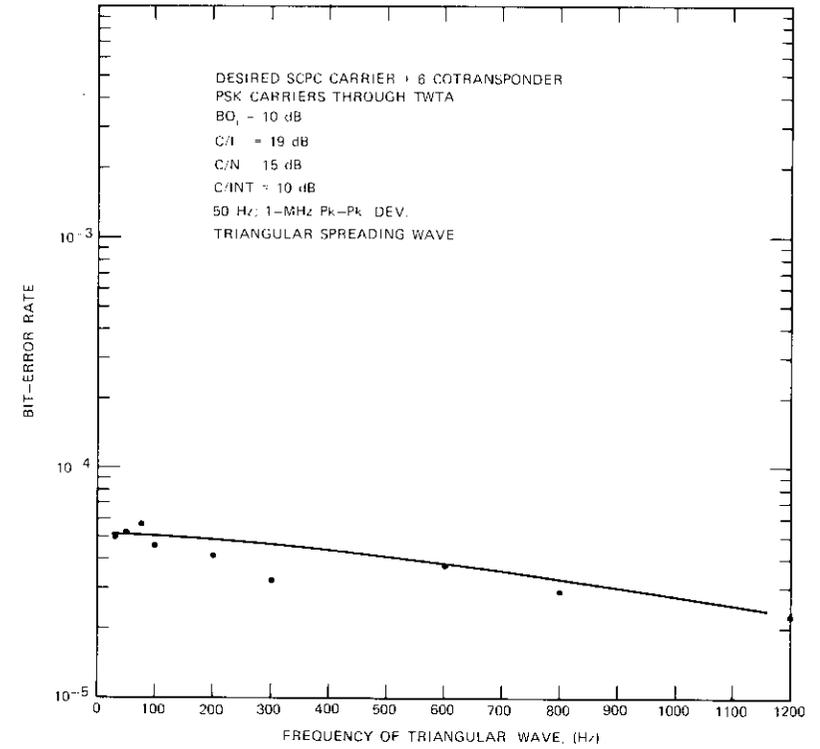
Figure 6. *Effect of Interference Source Modulating Frequency*

The SCPC channel performance through the simulator with both the spreading wave and TV signal (black field plus sync pulse and color burst) is shown in curves *a* and *b* of Figure 8. There is no significant difference between the corresponding curves of Figures 5 and 8 to within measurement accuracy. In this case, a spreading wave alone yields slightly higher error rates. Curve *c* in Figure 8 shows that error rates are lower for color bars than for a black field.

Conclusions

Based on the experimental results presented in this note, the following general conclusions can be drawn. In the PCM-PSK SCPC system, the error rate increase caused by a wideband sweeping inference source whose average power within a particular channel slot is at a certain level cannot
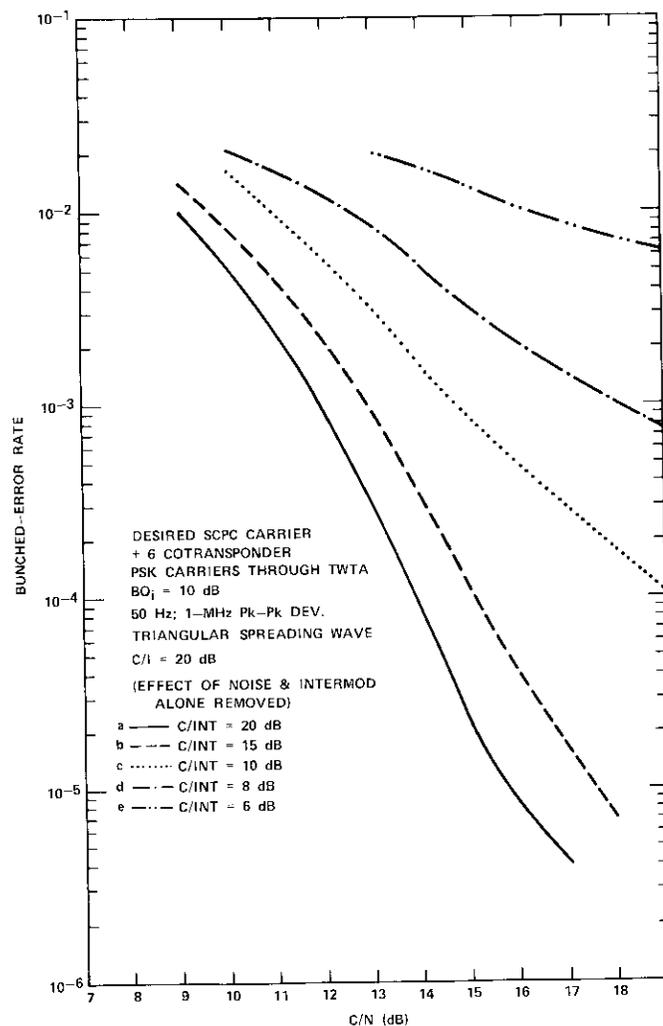
Figure 7. *Bunched Error Rate vs C/N*

Figure 8. *SCPC Performance with TV Signal and Spreading Wave*

be duplicated by a cochannel CW interference source with the same average in-band power.

The average BER for SCPC operation including intermodulation, AM/PM, and interference is given in Table 1. The increase in the average BER is

due to the occurrence of interference-induced errors; at times these errors bunch as the sweeping interference source impinges on the SCPC demodulator passband. The bunched or window error rate, which is the rate occurring when the interference energy is within the demodulator input filter bandwidth, is given in Table 2.

TABLE 1. SCPC AVERAGE BER

C/N (dB)	C/I (dB)	C/INT (dB)	Average BER
15	20	∞	4.2×10^{-6}
15	20	20	5.1×10^{-6}
15	20	15	8.5×10^{-6}
15	20	10	4.0×10^{-5}

TABLE 2. SCPC WINDOW ERROR RATE

C/N (dB)	C/I (dB)	C/INT (dB)	Bunched Rate	Average Errors per Pass*
15	20	∞	4.2×10^{-6}	0.000167
15	20	20	2.4×10^{-5}	0.0138
15	20	15	1.1×10^{-4}	0.045
15	20	10	9.4×10^{-4}	0.383

* Computed for 30-Hz sweep rate.

Based on a 30-Hz 1-MHz peak-to-peak interference sweep, the interference carrier is within the receiver's 38-kHz passband for 633 μs (or approximately 40 bits) per pass. The effect of the impulsive nature of the interference on telephone channel quality has been examined. Based on 3-minute testing periods, Table 3 summarizes the audible click occurrences vs operating conditions.

TABLE 3. AUDIBLE CLICK OCCURRENCES VS OPERATING CONDITIONS

C/N (dB)	C/I (dB)	C/INT (dB)	Average Audible Clicks in 3 min.
18	18.5	15	4
18	18.5	20	2.25
18	18.5	25	2
18	18.5	30	1

Acknowledgment

The authors wish to thank S. Fry for his assistance with the measurements.

Pitch axis adaptive compensation

JOHN HSING

(Manuscript received April 12, 1976)

Introduction

This note introduces a method which shows promise for the more accurate attitude control which may be required in future communications satellites. Satellite attitude control accuracy is improved by introducing self-adjusting parameters to compensate for disturbances of unknown magnitude at known frequencies. The application of a Liapunov-based parameter adaptive control technique to a satellite 3-axis attitude control system in the normal on-station mode will be described. The ability of the adaptive technique to compensate for some partially unknown external disturbances will be demonstrated analytically and substantiated by the results of a digital computer simulation.

The parameter adaptive control technique discussed in this note is based primarily on References 1 and 2. However, an article by Landau [3] reviews the state-of-the-art of adaptive control systems and lists 253 references.

A discussion of the control loop design for the pitch axis will be used to demonstrate the derivation of the adaptive control law. It will be shown that the results of a digital computer simulation of the system match the theoretical predictions very well. A comparison of pitch errors due to impulsive, step, and cyclic disturbance torques with and without parameter adaptive compensation shows that the parameter adaptive control technique substantially improves the satellite attitude performance.

Dr. Hsing is a member of the technical staff in the Stabilization and Structures Department of the Spacecraft Laboratory, COMSAT Laboratories.

System definition and pitch control law

Consider a 3-axis-stabilized satellite in a geostationary orbit with the reference coordinates and nominal angular momentum vector defined in Figure 1. With an earth sensor, this type of biased momentum wheel
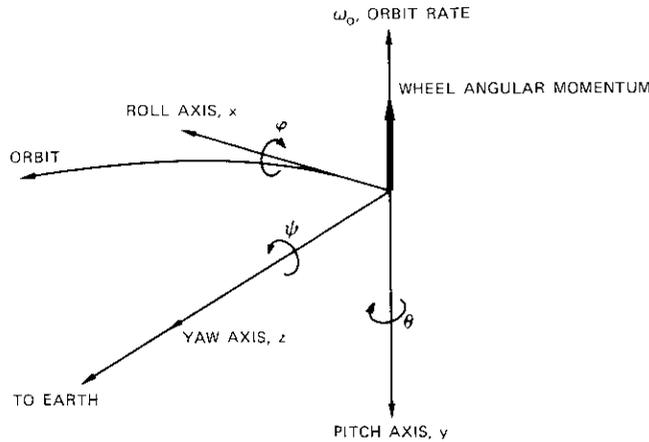

Figure 1. *Reference Frame and Wheel Angular Momentum*

offers pitch control torque through wheel speed modulation. Based on a classical formulation [4] which assumes a rigid body with internal momentum, the linearized pitch attitude differential equation is

$$I_y\ddot{\theta} + \dot{h}_{yc} = T_{dy} \tag{1}$$

where I_y is the pitch moment of inertia, θ is the pitch angle, h_{yc} represents the control torque, and T_{dy} represents disturbance torques with unknown magnitudes at known frequencies. For ease of analysis it will be assumed that

$$T_{dy} = a_0 + a_1 \cos \omega t + b_1 \sin \omega t \tag{2}$$

where a_0, a_1, and b_1 are unknown constants and ω is some known frequency.

The objective is to select the control function in equation (1) so that the pitch error due to these external disturbances asymptotically converges to zero. A satisfactory control law is

$$\dot{h}_{yc} = K_y(\tau_y\dot{\theta} + \theta) + k_1(t) + k_2(t) \cos \omega t + k_3(t) \sin \omega t \tag{3}$$

where K_y and τ_y are the pitch autopilot gain and time constant, respectively, and $k_1(t)$, $k_2(t)$, and $k_3(t)$ are the adaptive parameters which will self-adjust so that the controlled pitch loop is asymptotically stable.

Substituting equations (2) and (3) into equation (1) yields

$$I_y\ddot{\theta} + K_y\tau_y\dot{\theta} + K_y\theta = \delta_1(t) + \delta_2(t) \cos \omega t + \delta_3(t) \sin \omega t \tag{4}$$

where

$$\delta_1(t) \triangleq a_0 - k_1(t) \tag{5a}$$

$$\delta_2(t) \triangleq a_1 - k_2(t) \tag{5b}$$

$$\delta_3(t) \triangleq b_1 - k_3(t) \tag{5c}$$

are the parameter misalignments contributing to the driving torque of the controlled pitch differential equation [equation (4)].

The adaptive parameters $k_i(t)$ are adjusted on the basis of Liapunov's direct method.

Liapunov's direct method

Liapunov's direct method is a technique used to design an adaptive controller for adjusting feedback gains so that the controlled system is asymptotically stable. The most attractive feature of this method is that the stability of a system can be ascertained directly from the equations describing the system without recourse to explicit form solutions.

Equation (4) can be rewritten in the vector matrix form

$$\dot{\underline{y}}(t) = A\underline{y}(t) + \underline{b} \sum_{i=1}^{3} \delta_i(t)\, g_i(t) \tag{6}$$

where

$$\underline{y}(t) \triangleq [\theta, \dot{\theta}]^T \tag{7}$$

$$A \triangleq \begin{bmatrix} 0 & 1 \\ -\dfrac{K_y}{I_y} & -\dfrac{K_y\tau_y}{I_y} \end{bmatrix} \tag{8}$$

$$\underline{b} \triangleq \left[0, \dfrac{1}{I_y}\right]^T \tag{9}$$

and $g_1(t) = 1$, $g_2(t) = \cos \omega t$, and $g_3(t) = \sin \omega t$. The control function in equation (6) is directly related to the adaptive parameters of equation (5). These parameters will be defined so that $\underline{y} = \underline{0}$ is an asymptotically stable equilibrium point.

A potential Liapunov function of positive-definite quadratic form is chosen:

$$V = \underline{y}^T P \underline{y} + \sum_{i=1}^{3} \frac{1}{\alpha_i} (\delta_i + \beta_i \underline{y}^T P \underline{b} g_i)^2 \tag{10}$$

where the superscript T represents the transpose of a vector, P is a symmetric positive-definite matrix to be defined later, and α_i and β_i are arbitrary positive constants. Taking the time derivative of equation (10) and substituting equation (6) and its transpose for $\dot{\underline{y}}$ and $\dot{\underline{y}}^T$ yields

$$\dot{V} = \underline{y}^T(PA + A^TP)\,\underline{y} + 2\underline{y}^T P\underline{b} \sum_{i=1}^{3} \delta_i g_i$$

$$+ 2 \sum_{i=1}^{3} \frac{1}{\alpha_i} (\delta_i + \beta_i \underline{y}^T P \underline{b} g_i) \left[\dot{\delta}_i + \frac{d}{dt} (\beta_i \underline{y}^T P \underline{b} g_i) \right] \tag{11}$$

where the relationship $\underline{y}^T P \underline{b} = \underline{b}^T P \underline{y}$ has been used.

For a stable matrix A, there exists a unique positive-definite symmetric matrix P which is the solution of

$$PA + A^TP = -Q \tag{12}$$

where Q is a given positive-definite symmetric matrix [5]. Thus, the first term in equation (11) will be negative-definite if A is selected properly. For $\dot{V}$ to be negative-definite, which is a design requirement, the second and third terms in equation (11) should at least be negative-semidefinite. This will be true if the $\dot{\delta}_i$'s in equation (11) are defined as

$$\dot{\delta}_i = -\alpha_i \underline{y}^T P \underline{b} g_i - \beta_i \frac{d}{dt} (\underline{y}^T P \underline{b} g_i) \tag{13}$$

for $i = 1, 2, 3$.

Substituting equations (12) and (13) into equation (11) gives

$$\dot{V} = -\underline{y}^T Q \underline{y} - 2 \sum_{i=1}^{3} \beta_i (\underline{y}^T P \underline{b} g_i)^2 \tag{14}$$

which is negative-definite as expected. According to Liapunov's direct method [5], the controlled pitch loop is asymptotically stable with respect to the partially unknown disturbance torques.

The adjustment mechanism for parameters $k_i(t)$ can be obtained by substituting equation (5) into equation (13):

$$\dot{k}_i(t) = \alpha_i \underline{y}^T P \underline{b} g_i + \beta_i \frac{d}{dt} (\underline{y}^T P \underline{b} g_i) \tag{15}$$

for $i = 1, 2, 3$. If equation (15) is used to adjust the adaptive parameters which form the control function of the pitch loop, the attitude error will be driven to zero due to Liapunov's design of the control law.

Simulation results and comparison

It has been shown in the previous section that the control torque in equation (3) with the adaptive parameters defined in equation (15) guarantees that the controlled pitch loop is asymptotically stable. To provide high damping and rapid response to the driving function in equation (4), values of K_y and τ_y are nominally chosen to critically damp the system. During the simulation, $K_y = 0.7$ N-m/rad and $\tau_y = 50$ s were chosen, implying a critically damped pitch loop. The positive-definite symmetric matrix Q in equation (12) was selected as a unit matrix and the positive constants in equation (15) were arbitrarily selected as $\alpha_i = \beta_i = 10$ for $i = 1, 2, 3$. Equations (7), (9), and (12) can be used to rewrite the adaptive parameters defined in equation (15) as

$$k_1(t) = \frac{\alpha_1}{I_y} \int (p_{12}\theta + p_{22}\dot{\theta})\,dt + \frac{\beta_1}{I_y} (p_{12}\theta + p_{22}\dot{\theta}) \tag{16}$$

$$k_2(t) = \frac{\alpha_2}{I_y} \int (p_{12}\theta + p_{22}\dot{\theta}) \cos \omega t \, dt$$

$$+ \frac{\beta_2}{I_y} (p_{12}\theta + p_{22}\dot{\theta}) \cos \omega t \tag{17}$$

$$k_3(t) = \frac{\alpha_3}{I_y} \int (p_{12}\theta + p_{22}\dot{\theta}) \sin \omega t \, dt$$

$$+ \frac{\beta_3}{I_y} (p_{12}\theta + p_{22}\dot{\theta}) \sin \omega t \qquad (18)$$

where p_{12} and p_{22} are the components of the matrix P. The corresponding pitch control loop is shown in Figure 2.
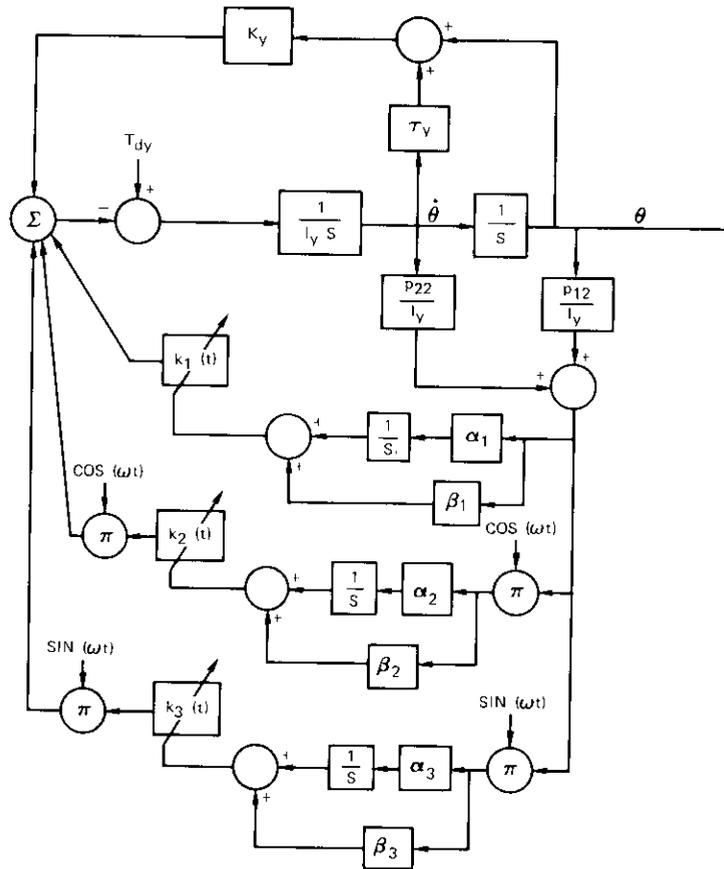
Figure 2. *Pitch Adaptive Control Loop*

This control system has been simulated on a digital computer with $I_y = 440$ kg-m^2. The pitch responses for impulsive, step, and cyclic (orbital rate) disturbance torques are plotted in Figure 3 and summarized in Table 1. The angles have been normalized as $\theta / |T_{dy}|$ with units of
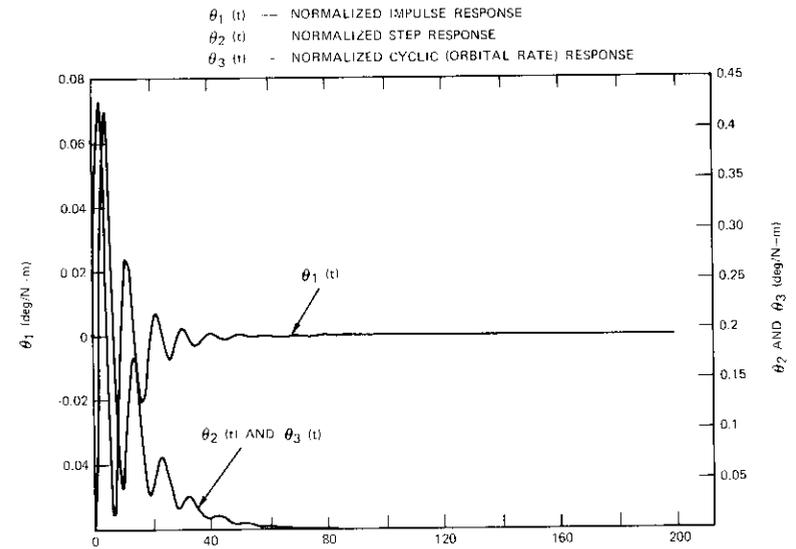
Figure 3. *Pitch Response with Adaptive Compensation*

deg/N-m, where $|\cdot|$ represents the maximum magnitude of a variable. These results substantiate the analysis discussed in previous sections.

TABLE 1. IMPROVEMENTS USING ADAPTIVE COMPENSATION

Disturbance	Without Adaptive Compensation		With Adaptive Compensation	
	Max. Error (deg/N-m)	Time Constant (s)	Max. Error (deg/N-m)	Time Constant (s)
Impulsive	0.6	80	0.07	20
Step	80	55	0.42	20
Cyclic (orbital rate)	80	∞	0.42	20

For ease of comparison, the normalized responses produced by a pitch autopilot without handling unknown disturbances, i.e., $k_i(t) = 0$ for $i = 1, 2, 3$, are plotted in Figure 4, where the steady-state errors due to step and cyclic disturbances are constant and cyclic, respectively. The comparison shows that both transient and steady-state attitude performance can be improved through adaptive compensation.

Figure 4. *Pitch Response without Adaptive Compensation*

Three cases were compared. In the first comparison an impulsive disturbance torque was used. Figure 4 shows the expected response of a critically damped control loop without adaptive compensation. When adaptive compensation is introduced, the rise to about 0.07 deg/N-m in the first few seconds is about the same. However, the pitch error is cut off and exhibits oscillations with rapidly diminishing amplitude. The maximum error is decreased by a factor of nine, and the time constant by a factor of four.
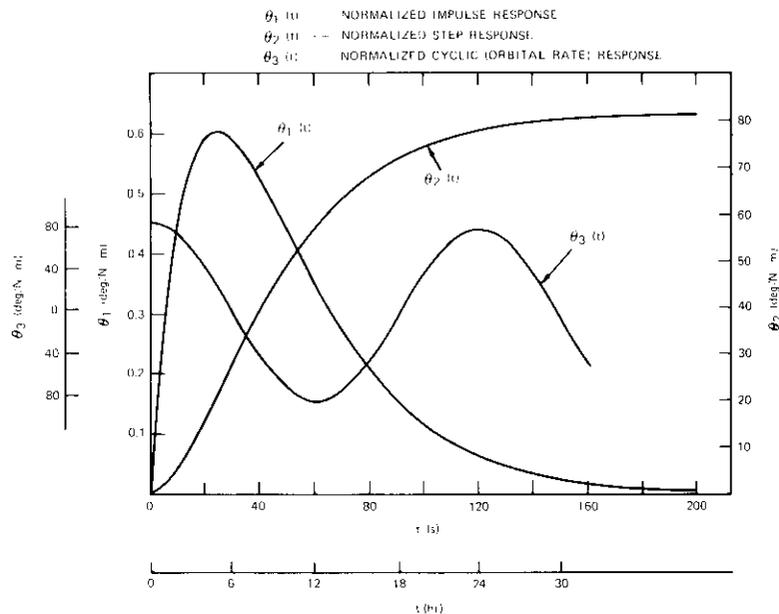
For the second comparison, the pitch response to a step disturbance without adaptive compensation is shown in Figure 4. Since the disturbance continues in time, a normalized error of 80 deg/N-m continues indefinitely. When adaptive compensation is added (see Figure 3), the maximum error is reduced to 0.42 deg/N-m and damps out rapidly to zero. The decrease of more than two orders of magnitude in the maximum error generated demonstrates the effectiveness of adaptive compensation.

The last comparison is for a cyclic solar pressure torque. Without adaptive compensation, Figure 4 shows the expected cyclic pitch error, with a normalized amplitude of 80 deg/N-m. When adaptive compen-

sation is added, Figure 3 shows that the maximum error has decreased by more than two orders of magnitude, and that, after the transient decays, the error is even smaller. In Figure 3, θ_2 and θ_3 share the same curve due to the low frequency of the solar pressure torque. It is also possible to adaptively compensate for cyclic disturbances at other frequencies, and the pitch error asymptotically converges to the null.

Conclusions

A parameter adaptive control system based on Liapunov's direct method has been applied to a pitch attitude control loop design in the normal on-station mode of a 3-axis-stabilized geostationary satellite. The ability of the adaptive technique to compensate for some unknown disturbance torques has been demonstrated for both transient and steady-state conditions. With the adaptive scheme, the pitch errors due to partially unknown disturbance torques will asymptotically converge to zero, whereas without adaptive compensation they will not. It can be concluded that transient and steady-state attitude performance can be substantially improved through adaptive compensation.

Acknowledgment

References

[1] P. C. Parks, "Liapunov Redesign of Model Reference Adaptive Control Systems," *IEEE Transactions on Automatic Control*, AC-11, 1966, pp. 262–267.

[2] J. W. Gilbart, R. V. Monopoli, and C. F. Price, "Improved Convergence and Increased Flexibility in the Design of Model Reference Adaptive Control Systems," Ninth Symposium on Adaptive Processes, the University of Texas at Austin, December 7–9, 1970.

[3] I. D. Landau, "A Survey of Model Reference Adaptive Techniques—Theory and Applications," *Automatica*, Vol. 10, Pergamon Press, 1974, pp. 353–379.

[4] M. G. Lyons, K. L. Lebsock, and E. D. Scott, "Double Gimbaled Reaction Wheel Attitude Control System for High Altitude Communications Satellites," AIAA Guidance, Control and Flight Mechanics Conference, Hofstra University, Hempstead, New York, August 16-18, 1971, AIAA Paper 71-949.

[5] J. C. Hsu and A. J. Meyer, *Modern Control Principles and Applications*, New York: McGraw-Hill Book Co., 1968.

Translations of Abstracts in this issue

Modèle d'assignation et de programmation des paquets AMRT

A. K. SINHA

Sommaire

Cet article propose une expression mathématique simple du problème de l'assignation et de la programmation des paquets d'information dans un système d'accès multiple par répartition dans le temps (AMRT) destiné à un système général de télécommunications par satellite dans lequel on choisit de façon arbitraire le réseau de stations terriennes, ainsi que le diagramme de couverture de l'antenne et la configuration du répéteur d'un satellite. On présente les concepts applicables aux problèmes de superposition de différents diagrammes de couverture, sur une même station terrienne, et du recouvrement dans le temps de différents paquets, ainsi que ceux dus aux contraintes des équipements des stations terriennes. On met également en évidence certains paramètres utiles à l'évaluation de l'efficacité de l'utilisation du système. Enfin, on propose un algorithme semi-analytique permettant de programmer les paquets AMRT de façon à réduire au minimum les équipements requis aux stations terriennes et à optimiser l'efficacité de la programmation pour des données de trafic spécifiées et une configuration de système donnée. On présente un exemple obtenu grâce à un programme d'ordinateur qui vient d'être élaboré suivant cette méthode.

Restauration automatique de l'orthogonalité de polarisation dans les systèmes de satellites à réutilisation de fréquence

D. F. DiFonzo, W. S. Trachtman et A. E. Williams

Sommaire

Cet article décrit des réseaux rétablissant automatiquement l'orthogonalité de polarisation. Ces réseaux peuvent être utilisés pour minimiser les brouillages réciproques qui se produisent dans une liaison de télécommunication par