of comparing the waveforms generated by various bit sequences in the transmitted signal. Although the method requires the manipulation of a rather large number of sequences, it makes no excessive demands on computer time as the required numerical operations can be made quite simple. Thus the curves in Figs. 3-5, representing data compressed from 48 complete simulations, required less than 20 minutes of UNIVAC 1110 computer time.

The computed examples appear to indicate that inherent channel degradation is not serious. This implies that waveforms associated with distinct messages remain almost as distinguishable as in the absence of intersymbol interference. There thus exists a substantial potential for receiver improvement relative to bit-by-bit detection. On the other hand, comparison of QPSK with O-QPSK and MSK reveals, at least for the type of rudimentary channel studied here, little advantage that would justify use of the latter, more complex formats.

ACKNOWLEDGEMENT

The author gratefully acknowledges the participation of Dr. Kai Ping Yiu in the creation of the satellite-channel simulation program.

REFERENCES

[1] G. D. Forney, Jr., "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference", *IEEE Trans. Inform. Theory,* Vol. IT-18, pp 363-378, May 1972.

[2] G. D. Forney, Jr., "Lower Bounds on Error Probability in the Presence of Large Intersymbol Interference", *IEEE Trans. Comm.,* Vol. COM-20, pp 76-77, 1972.

[3] H. Van Trees, *Detection, Estimation, and Modulation Theory,* Vol. III, Appendix, John Wiley & Sons, Inc., New York, 1971.

[4] G. Ungerboeck, "Adaptive Maximum-Likelihood Receiver for Carrier Modulated Data-Transmission Systems", *IEEE Trans. Comm.,* Vol. COM-22, pp 624-636, 1974.

[5] I. N. Anderson, "Sample Whitened Filters", *IEEE Trans. Inform. Theory,* Vol. IT-19, pp 653-660, 1973.

[6] M. F. Mesiya, P. J. McLane, and L. L. Campbell, "A Maximum Likelihood Receiver for Binary PSK Transmission over a Bandlimited Nonlinear Channel", *National Telecommunications Conf.,* 1975, pp 23-5 to 23-8.

[7] M. F. Mesiya, P. J. McLane, and L. L. Campbell, "Maximum Likelihood Sequence Estimation of Binary Sequences Transmitted Over Bandlimited Channels", *IEEE Trans. on Communications,* Vol. COM-25, pp 633-643, July 1977.

A Multi-User-Class, Blocked-Calls-Cleared, Demand Access Model

J. M. AEIN, SENIOR MEMBER, IEEE

Abstract—**Constraints on capacity allocation are investigated for circuit-switched demand access to a common transmission resource by user communities with differing traffic intensity and capacity requirements. A simple birth-death steady-state traffic model is used together with a geometrical representation of any given set of constraints placed upon user access. State probabilities have a simple product form which holds for a wide class of constraint sets. Performance characteristics are intrinsic to the upper surface of the constraint set.**

Paper approved by the Editor for Computer Communication of the IEEE Communications Society for publication without oral presentation. Manuscript received December 10, 1976; revised October 10, 1977. This work was supported in part by the Office of the Secretary of Defense, Director Telecommunications Command and Control, under Contract DAHC 1573C0200.

The author is with the Institute for Defense Analyses, Arlington, VA 22202.

I. INTRODUCTION

A simple birth-death traffic model is formulated to analyze access constraints and capacity allocated to users linearly sharing a common transmission resource. The problem derives from the use of Demand Assigned Multiple Access (DAMA) *circuit-switched* service in communication satellite systems. However, the model has wider application; for example, multispeed dial up data line multiplex structure at an access point to a wideband digital trunk. The general problem setting and physical interpretation is presented in Ref. 1. This paper supplements Ref. 1 with the theoretical development (those readers interested in store-and-forward systems should also see Refs. 2, 3).

A wideband transmission facility (e.g., satellite) is modeled in Fig. 1 as providing a pool of capacity which can be *linearly* subdivided and assigned to groups or classes of user terminals which then share their allotted capacity on a call-by-call basis. A call consumes an amount of capacity depending on the physical characteristics of its originating terminal e.g., satcom, receive sensitivity, transmit power, and bandwidth. For data access to a multiplexed wideband trunk, the capacity per call is simply the access line speed, normally $2^n \times 75$ bit/s. Strategies for sharing the transmission capacity are implemented by imposing access constraints on the user classes. Call traffic is modeled as a stationary independent Poisson arrival process with exponential holding times. Traffic parameters and capacity used per call can vary between classes but are identical within a class. Demand for transmission service is provided on a first-come, first-served basis within the access constraints imposed. For example, if there are K-user classes one can divide the available capacity into K-separate "channels", one for each user class (dedicated access). Alternatively, one could allow unrestricted access to all the capacity by all users (fully shared access).

The central objective here is the development of a geometrical approach in describing access constraints and their relation to the solution of the associated birth-death state equations. The analysis is limited to a blocked-calls-cleared mode of operation. For a multiuser class system which holds calls in queue when there is no capacity available, the analysis easily becomes intractable (Ref. 2) for all but the simplest access constraints and queueing disciplines. One must stipulate a consistent and comprehensive state-dependent "protocol" as to which of the multiclass calls in queue are to be served as a function of which call in progress terminates.

The methods used here were motivated by two recent papers (Refs. 4, 5). Reference 4 considers synchronous time-division multiplexing of different speed local loops onto a wideband, digital trunk. Consequently, the model considered only integer-valued call capacities. Further, call arrival and holding papameters, λ and μ, were unnecessarily restricted to be the same for all sources. Reference 5 models the satellite problem wherein

0090-6778/78/0300-0378$00.75 © 1978 IEEE

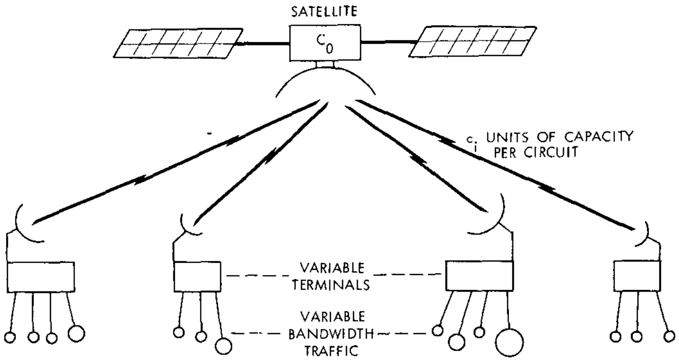

Fig. 1. Satellite System Configuration.

call capacity need not be integer multiples, nor need the traffic activity parameters all be the same. Reference 5 introduced for notational convenience the concept of representing the capacity access constraints as an allowable set of states, A, but did not develop the geometrical significance. Reference 2 provides a fuller discussion of current efforts in the context of conventional teletraffic theory and the rapidly developing theory of queueing networks.

II. MATHEMATICAL MODEL

A. System Definition

For each $i = 1, 2, \cdots, K$, define user class i with parameters $(N_i, \lambda_i, \mu_i, c_i)$ and let j_i equal the number of calls in progress from class i.

Traffic:

a. Population size $= N_i$, either finite (Engset model) or infinite (Erlang model).

b. Differential call arrival probability (Poisson arrivals)

$$= (N_i - j_i)\lambda_i dt \qquad \text{when } 0 \leqslant j_i \leqslant N_i < \infty; \text{ (Engset)}$$

$$= \tilde{\lambda}_i dt \qquad \text{when } N_i = \infty, \text{ i.e., } N_i\lambda_i \to \tilde{\lambda}_i \text{ (Erlang).}$$

c. Differential call holding probability $= \mu_i dt$ (exponential holding time).

State:

d. A vector $j \equiv (j_1, j_2, \cdots, j_K)$, where j_i is the number of calls in progress from user class i.

Linear Capacity Usage:

e. The capacity in use $C(j) \equiv \Sigma_{i=1}^K c_i j_i \leqslant C_0$, where c_i is capacity used by a class i call in progress, and C_0 is the total available capacity.

Access Constraint:

f. A set, A, of allowed states j.

g. A call is blocked from service if on its arrival it would move the system state to be outside A. Thus the set of blocked states for class i must be determined from A as $B_i = \{j \mid j + \delta_i \notin A\}$, where δ_i is a unit vector with value one at the i^{th} location and zeros elsewhere.

From (a) and (e) above,

$$A \subseteq \Omega \equiv \{j \mid C(j) \leqslant C_0; j_i \leqslant N_i\}.$$

The fundamental description of the "system" in the model used here is the specification of the user-class traffic parameters, the set A, and a solution for the probability distribution, $P(j), j \in A$. One then computes the $P(B_i)$ which measures the grade of service. Calls are provided "demand access" but need not have access to all of the capacity C_0. The A-set is a means of apportioning capacity availability to the different user classes.

Examples of A-Sets:

1. $A = \Omega$ (Fig. 2a).

Fully Shared Capacity: All calls have full access to any available capacity.

2. $A = \prod_{i=1}^{K} \{j_i \mid c_i j_i \leqslant C_i\}, \qquad \text{where } \Sigma C_i = C_0$ (Fig. 2b).

Capacity Dedicated to Each User Class: Capacity is apportioned to each user class, with access by calls only from its class.

3. $A = \{j \mid j \in \Omega \text{ and } c_i j_i \leqslant C_i \leqslant C_0 \text{ for } i = M+1, \cdots, K\}$ (Fig. 2c).

Capacity-limited Access: Fully shared access to first M classes; last $K - M$ user classes have a capacity limit.

Choose values $\hat{C}_i$ such that $\Delta \equiv C_0 - \Sigma_{i=1}^K \hat{C}_i > 0$, then limit all the user classes at $C_i = \hat{C}_i + (C_0 - \Sigma_{i=1}^K \hat{C}_i)$ to produce

4. $A = \{j \mid c_i j_i \leqslant \hat{C}_i + \Delta, i = 1, 2, \cdots, K\}$ (Fig. 2d)

Dedicated with Shared Overflow Access: Provides $\hat{C}_0 = \Sigma C_i < C_0$ for use as a dedicated allocation and the reserve capacity, Δ, is used on a fully shared basis. When $j_i \leqslant \hat{j}_i = \hat{C}_i/c_i$, the i^{th} class is not affected by any other class. When $j_i > \hat{j}_i$, then the excess, $j_i - \hat{j}_i$, must compete for the reserve capacity, Δ, on a fully shared basis.

There are limitations in specifying A-sets; not only must $A \subseteq \Omega$ but, because of the birth-death model, A must be connected.

In every instance, an A-set is a geometrical solid in K dimensions resting on the positive coordinate axes. The choice of A-set affects the blocking states B_i and thus couples blocking between user classes.

All the A-sets in Fig. 2 have an extremely important property: the orthogonal projections from any point in the set A to the lower bounding planes formed by the coordinate axis are wholly contained in the set A. To coin a phrase, any figure possessing this last property is said to have coordinate-convexity. Note that A can possess coordinate-convexity without being convex. A property equivalent to coordinate-convexity is that the exterior surfaces of A can be decomposed into the lower coordinate planes plus an upper surface made up entirely of planar facets parallel to some coordinate plane (see Fig. 2).

Coordinate-convexity in A will provide a *sufficient* condition (necessity is an open question) for a separating solution, $P(j)$, to the K-dimensional equations of state. In this case $P(j)$ will have a product form even though the *coefficients* of the K-dimensional difference operator depend on the boundary of A:

$$P(j) = P_A(0) \cdot \prod_{i=1}^{K} \binom{N_i}{j_i} a_i^{j_i}$$

$$P_A(0) = \left(\sum_{j \in A} \prod_{i=1}^{K} \binom{N_i}{j_i} a_i^{j_i} \right)^{-1}. \qquad (1)$$

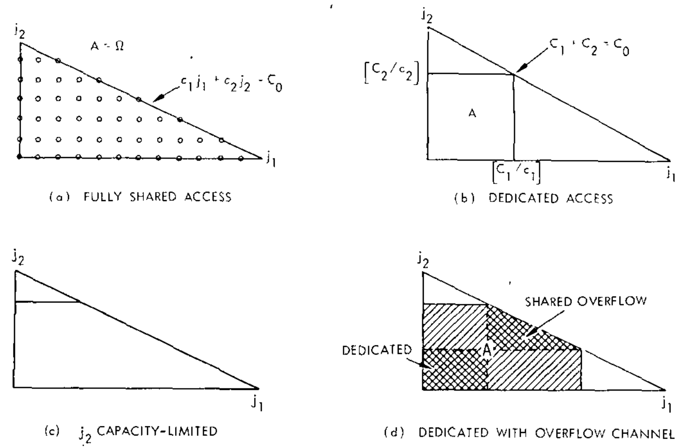

Fig. 2. Access Strategies and A-Sets (note that due to the discrete values of the j_i the upper boundaries of the A-Set are actually staircased).

Note that the K-fold product on the j_i depends only on the traffic parameters of the user classes and is wholly independent of A. The impact of various access strategies (choices of A-sets and available capacity) are manifested only in $P_A(0)$, a crucial number. $P_A(0)$ is determined by the normalizing condition $P(A) = 1$. It is the probability that there are no calls in progress from any user class. This implies for system simulations that one need only monitor the percent of time the system is empty (rather than all $\boldsymbol{j} \in A$) and the traffic intensities a_i which will then determine $P(\boldsymbol{j})$.

Equation 1 has a deceptively simple form which becomes computationally difficult for moderate K and C_0 values. Remember that the $\boldsymbol{j} \in B_i$ must be determined (or tested for) and $\Sigma_{\boldsymbol{j}\in B_i} P(\boldsymbol{j})$ accumulated.

For an A-set which is not coordinate-convex, there exists at least one $\boldsymbol{j}$ state, $\boldsymbol{j}'$, in A with at least one coordinate i' such that $j'_{i'} \geqslant 1$ and the state $\boldsymbol{j}' - \delta_{i'}$ is no longer in A. This implies that the termination of a call in progress from an i' user class cannot take place until some other event (e.g., call arrival or departure from other than an i' user) transpires. Physically, the capacity relinquished by a departing call would not be reassignable to a new call arrival until the system transitions to a suitable $\boldsymbol{j}$ state. Coordinate-convex A-sets guarantee that call completion from all $\boldsymbol{j}$ states returns the newly freed capacity for use by new call arrivals.

A-sets lacking coordinate-convexity are certainly undesirable in communications systems. For application to the sharing of communications capacity, it is reasonable to strive for A-sets that possess coordinate-convexity. System implementation must ensure prompt call take down upon call termination.

B. Equations of State

From the properties of the birth-death process $P(\boldsymbol{j})$ must satisfy a K-dimensional second order difference equation with boundary conditions, and $P(A) \equiv 1$.

For brevity, the following vector notation is used. Let $\boldsymbol{j}$, μ, λ, and N be K-dimensional positive vectors of j_i, μ_i, λ_i, and N_i, respectively. Let I be a K-vector of all ones.

Define

$$(\mathrm{a}, \mathrm{b}) \triangleq \sum_{i=1}^{K} a_i b_i,$$

$P(\boldsymbol{j}) \triangleq$ scalar probability function

$\boldsymbol{j} \pm \delta_i \triangleq (j_1, j_2, \cdots, j_i \pm 1, j_{i+1}, \cdots, j_K)$

$\lambda\Delta^{-1}P(\boldsymbol{j}) \triangleq$ backward difference vector whose i^{th} component is $\lambda_i P(\boldsymbol{j} - \delta_i)$

$\mu\Delta^{+1}P(\boldsymbol{j}) \triangleq$ forward difference vector whose i^{th} component is $\mu_i P(\boldsymbol{j} + \delta_i)$.

For any point $\boldsymbol{j}$ in the strict interior of A, $\mathring{A}$, the homogeneous difference equation that $P(\boldsymbol{j})$ must satisfy is given by

$$[(\mu, \boldsymbol{j}) + (\lambda, N - \boldsymbol{j})]\, P(\boldsymbol{j}) - (N + I - \boldsymbol{j}, \lambda\Delta^{-1}P(\boldsymbol{j})) - (\boldsymbol{j} + I, \mu\Delta^{+1}P(\boldsymbol{j})) = 0, \qquad \text{for all } \boldsymbol{j} \in \mathring{A}, \tag{2}$$

The equations of state must also be specified on the boundaries of A. With coordinate-convexity, the boundaries of A separate (see Fig. 2) into a) K lower coordinate planar boundaries $\underline{A}(i)$, formed by the A-set resting upon the positive coordinate axes, and b) an upper boundary $\bar{A}$. Each planar lower boundary $\underline{A}(i)$ is composed of states $\boldsymbol{j}$ with nonzero-valued coordinates only at j_i and j_{i+1}. (In the case of $\underline{A}(K)$, the nonzero coordinates would be j_K and j_1 so that $i + 1$ is modulo K.)

Since a call cannot occur from below the lower boundary $\underline{A}(i)$, there can be no backward difference, $\lambda\Delta^{-}$, in the $K - 2$ directions normal to $\underline{A}(i)$ (i.e., the j_l for $l \pm i, i+1 \pmod K$)). In addition, the corresponding terms in the sum $(\mu, \boldsymbol{j})$ must also vanish, as they correspond to a call completion to a state below $\underline{A}(i)$. This is accounted for automatically by letting $j_l \equiv 0$ $l = i, i \mp 1 \pmod K$.

Similarly, when $\boldsymbol{j}$ belongs to the upper boundary $\bar{A}$, of A, it is not possible for a call to arrive or complete from a disallowed state above the boundary. Thus, for $\boldsymbol{j} \in \bar{A}$, Eq. 2 must be modified by setting to zero the $K - 2$ vector coordinate components normal to $\bar{A}$ at $\boldsymbol{j}$ of the forward difference vector $\mu\Delta^{+1}$ and the corresponding coordinates in the $(\lambda, N - \boldsymbol{j})$ sum.

The homogeneous system of state difference equations in its briefest form is given by Eq. 2 with boundary conditions determined by

$$P(A) = \sum_{\boldsymbol{j}\in A} P(\boldsymbol{j}) = 1.$$

For $\boldsymbol{j} \in \underline{A}(i)$, then for $z = i, i + 1 \pmod K$

$$\{\lambda\Delta^{-1}P(\boldsymbol{j})\}_z = 0.$$

This condition must also be applied to those $\boldsymbol{j}$ states on a coordinate axis, $\boldsymbol{j} \equiv (0, 0, j_i, 0, 0)$ and $\boldsymbol{j} \equiv 0$. The coordinate axis i is the intersection of $\underline{A}(i-1)$ and $\underline{A}(i)$, while $\boldsymbol{j} \equiv 0$ is the intersection of all K lower boundary planes $\underline{A}(i)$. One must set the indicated terms to zero simultaneously for these states.

For $\boldsymbol{j} \in \bar{A}$, let $V(\boldsymbol{j})$ be the set of those indices i of the $\boldsymbol{j}$ vector for which $\boldsymbol{j} + \delta_i$ is outside A. Then, for $\boldsymbol{j} \in \bar{A}$ and $i \in V(\boldsymbol{j})$,

$$\{\mu\Delta^{+1}P(\boldsymbol{j})\}_i = 0,$$

$$(\lambda, N - \boldsymbol{j}) = \sum_{i \notin V(\boldsymbol{j})} \lambda_i (N_i - j_i).$$

For those classes i which have infinite source population, the λ inner product terms can be modified by passing to the limit $N_i\lambda_i \rightarrow \tilde{\lambda}_i$.

C. Separating Solution

For all $\boldsymbol{j} \in A$ assume that $P(\boldsymbol{j}) > 0$ can be factored into the product of K functions $P_i(j_i)$ and seek to generate K second-order homogeneous one-dimensional difference equations. Let D_i be a one-dimensional second-order difference operator on a function, $F(s); s = 0, 1, 2, \cdots$.

$$D_i[F(s)] \triangleq (\mu_i s + (N_i - s)\lambda_i)F(s) - (N_i + 1 - s)\lambda_i F(s-1) - (s+1)\mu_i F(s+1), \tag{3}$$

where μ_i, λ_i, N_i are the traffic source parameters for user class i (and hence the subscript on D).

Substituting the assumed product form of $P(\boldsymbol{j})$ into Eq. 2, using Eq. 3 and then dividing by $P(\boldsymbol{j}) > 0$, produces

$$\sum_{i=1}^{K} P_i^{-1}(j_i) \cdot D_i[P_i(j_i)] = 0; \qquad \text{for } \boldsymbol{j} \in A^{\circ}. \tag{4a}$$

For the lower boundaries of A

$$\sum_{l=i}^{i+1} P_l^{-1}(j_l)D_l[P_l(j_l)] = \sum_{\substack{l=1 \\ l \neq i, i+1}}^{K} P_l^{-1}(0)(N_l\lambda_l P_l(0) - \mu_l P_l(1)), \qquad \text{for } \boldsymbol{j} \in \underline{A}(i), j_i \geqslant 1, j_{i+1} \geqslant 1 \tag{4b}$$

$$P_i^{-1}(j_i)D_i[P_i(j_i)] = \sum_{\substack{l=1 \\ l \neq i}}^{K} P_l^{-1}(0)(N_l\lambda_l P_l(0) - \mu_l P_l(1)),$$

$$\text{for } j \in \underline{A}(i) \cap \underline{A}(i+1) \equiv \text{positive } i^{\text{th}} \text{ coordinate axis} \tag{4c}$$

$$\sum_{i=1}^{K} P_i^{-1}(0)[N_i\lambda_i P_i(0) - \mu_i P_i(1)] = 0, \qquad \text{for } \boldsymbol{j} \equiv 0 \tag{4d}$$

The upper boundary conditions on Eq. 2 cannot be as explicitly stated since more complex upper boundary surfaces are allowed. If $\boldsymbol{j} \in \bar{A}$, there exists at least one i such that $\boldsymbol{j} + \delta_i$ does not belong to A. Let $V(\boldsymbol{j})$ be those indices of $\boldsymbol{j} \in \bar{A}$ for which $\boldsymbol{j} + \delta_i \notin A$ (the number of elements in the set $V(\boldsymbol{j})$ must lie between 1 and K). For $\boldsymbol{j} \in \bar{A}$, both the forward difference as well as the $(\lambda, N - \boldsymbol{j})$ components at those coordinate indices contained in $V(\boldsymbol{j})$ go to zero. Then, substituting the product form into Eq. 2 modified for the upper boundary conditions on $\bar{A}$ and dividing by $P(\boldsymbol{j})$ yields for $\boldsymbol{j} \in \bar{A}$

$$\sum_{i \notin V(\boldsymbol{j})} P_i^{-1}(j_i)D_i[P_i(j_i)] = \sum_{i \in V(\boldsymbol{j})} P_i^{-1}(j_i)[\mu_i j_i P_i(j_i) - (N_i + 1 - j_i)\lambda_i P_i(j_i - 1)]. \tag{4e}$$

Note that the right-hand side of Eq. 4e is a *first-order* difference operator on indices $i \in V(\boldsymbol{j})$, while the usual second-order operator D_i is in effect on those $i \notin V(\boldsymbol{j})$.

Eq. 4 (with D_i defined by Eq. 3), together with the normalization

$$\sum_{\boldsymbol{j} \in A} \prod_{i=1}^{K} P_i(j_i) = 1, \tag{5}$$

provides a sufficient set of difference equations for a separable solution to Eq. 2.

Although a product form for $P(\boldsymbol{j})$ was assumed, the resulting equations of state (4) and (5) have not separated! At $\boldsymbol{j} \equiv 0$ arbitrarily* set each summand of Eq. 4d to zero individually so that $P_i(1) = N_i a_i P_i(0)$, $a_i = \lambda_i/\mu_i$; then it follows that the righthand side of Eq. 4c will be zero and $D_i[P_i(j_i)] = 0$ along each coordinate axis $i = 1, 2, \cdots, K$. Similarly, for $\boldsymbol{j}$ belonging to a coordinate plane, $A(i)$, the right-hand side of Eq. 4b will be zero from zeroing Eq. 4d and the left-hand side will zero from Eq. 4c. Furthermore, Eq. 4a will also be satisfied for an interior point of A. It is easily checked that the recursion relation, $P_i(j_i + 1) = (N_i - j_i)a_i P_i(j_i)$, satisfies the D_i operator and the boundary condition at $\boldsymbol{j} \equiv 0$. This recursion for each i implies the form of Eq. 1.

Thus one *forces* a separating solution at $\boldsymbol{j} \equiv 0$ and recurses first along each coordinate axis, into each lower boundary coordinate plane and then upward into A. Note that this solution technique requires that the rectilinear recursion process chosen traverse states all belonging to A, i.e., that A be coordinate-convex. Thus if Eq. 4e is also satisfied then a product form solution to $P(\boldsymbol{j})$ will have been found and since A is finite, $P(A) = 1$ (Eq. 5) and the solution will be unique (Ref. 6, p. 408).

Let $\boldsymbol{j}$ be any state on the boundary of $\bar{A}$. Then the recursion relation for P_i guarantees that for those $i \notin V(\boldsymbol{j})$, $D_i[P_i(j_i)] = 0$, and hence the left-hand side of Eq. 4e is zero. Consequently, it is only necessary that for each $i \in V(\boldsymbol{j})$ the corresponding sum on the right-hand side of Eq. 4e is zero. This is shown by direct substitution of Eq. 6 into the right side of Eq. 4e.

Summing $\Pi P_i(j_i)$ over all $\boldsymbol{j} \in A$ and setting the result equal to one produces the separated solution in *product form* given by Eq. 1.

To reemphasize, for access constraint sets A with coordinate-convexity and blocked-calls-cleared, the only portion of $P(\boldsymbol{j})$ that exhibits dependency on A is the normalization constant $P_A(0)$. The product portion is completely determined by the traffic parameters of the user classes, i.e., N_i, λ_i, μ_i, ($a_i = \lambda_i/\mu_i$). The parameters N_i, c_i, and C_0 establish the set Ω within which A must be contained. Thus, in a formal sense, *choice of system architecture A is primarily manifested in the calculation of $P_A(0)$.*

D. Statistical Dependence

A point of considerable importance is that apart from the factor $P_A(0)$, the product form of Eq. 1 might indicate that the user classes are independent. In a strict probabilistic sense, the user classes are statistically *dependent,* although the form of the state probability has the computational convenience as if the user classes were independent. Physically, one would expect that only when capacity is dedicated to a user class will it be statistically independent of the other K-1 classes. This is

* One can physically reason that the $P_i(0)$ are like arbitrary constants of integration depending on the normalization.

now proved as an example of the geometrical properties of an A-set. For a given $\mathbf{j}$, let the number $M_i(\mathbf{j})$ be the largest j_i value in A with the other j_l, $l \neq i$ fixed by the given $\mathbf{j}$.

Any class i (say $i = 1$) is statistically independent if, and only if

$$P(\mathbf{j}) = Q_1(j_1)P(j_2, j_3, \cdots, j_K); \qquad \text{for all } \mathbf{j} \tag{6}$$

where

$$Q_1(j_1) \equiv \text{Prob } [j_1 \text{ calls in progress}]$$

and

$$P(j_2, \cdots, j_K) \equiv \sum_{j_1=1}^{M_1(\mathbf{j})} \text{Prob } [\mathbf{j} \mid j_1].$$

Define $\sigma(j_i)$ to be the set of all $\mathbf{j} \in A$ with i^{th} coordinate equal to j_i. Then using Eq. 1,

$$Q_1(j_1) = P_A(0) \binom{N_1}{j_1} \sum_{\mathbf{j} \in \sigma(j_1)} \pi_2(N, a, \mathbf{j}) \tag{7}$$

$$P(j_2 \cdots j_K) = P_A(0)\pi_2(N, a, \mathbf{j})$$

$$\cdot \sum_{k_1=0}^{M_1(\mathbf{j})} \binom{N_1}{k_1} a_1^{k_1} \tag{8}$$

where

$$\pi_2(N, a, \mathbf{j}) \triangleq \prod_{i=2}^{K} \binom{N_i}{j_i} a_i^{j_i}.$$

Inserting Eqs. 7 and 8 into Eq. 6 and canceling terms produces the condition that for every $\mathbf{j} \in A$ (and hence unique pair $j_1, M_1(\mathbf{j})$)

$$P_A^{-1}(0) \equiv \sum_{j_1=0}^{M_1(\mathbf{j})} \left(\binom{N_1}{j_1} a_1^{j_1} \cdot \sum_{\mathbf{j} \in \sigma(j_1)} \pi_2(N, a, \mathbf{j}) \right)$$

$$= \left(\sum_{k=0}^{M_1(\mathbf{j})} \binom{N_1}{k_1} a_1^{j_1} \right) \left(\sum_{\mathbf{j} \in \sigma(j_1)} \pi_2(N, a, \mathbf{j}) \right) \tag{9}$$

where the left-hand side of Eq. 9 used the fact that A is the disjoint sum of $\sigma(j_1)$. Thus if $i(=1)$ is to be an independent class, Eq. 9 must hold for every $\mathbf{j} \in A$ or equivalently for every pair j_i, $M_i(\mathbf{j})$ associated with $\mathbf{j}$. *This can only happen if $\sigma(j_i)$ and $M_i(\mathbf{j})$ are constant as $\mathbf{j}$ ranges over A.* That is to say, the $\bar{A}$ surface normal to the i^{th} coordinate is a flat plane whose "outline" is congruent to all planar cuts through A at all $j_i \geqslant 0$ values.

For all K-user classes independent, this condition applied successively can only be satisfied by a K-dimensional rectangular solid or dedicated access per Fig. 2b. Consequently, even though Eq. 1 is a product of P_i, these P_i are not the marginal probabilities!

E. Performance Measures

The average percentage of capacity used gives a measure of facility utilization, a quantity of interest to the *system provider*. User interest centers on the percentage of time (i.e., the probability) he will be blocked (i.e., denied a circuit).

The average utilization of the system is given by

$$U_A \triangleq \sum_{\mathbf{j} \in A} C(\mathbf{j})P(\mathbf{j})/C_0 = \sum \frac{c_i}{C_0} \langle j_i \rangle$$

$$\langle j_i \rangle \triangleq \sum j_i Q_i(j_i) \tag{10}$$

where U_A depends on traffic characteristics, available capacity, and access strategy or A-set. Note that the blocked states, B_i, need not be mutually disjoint, but that $\cup_{i=1}^{K} B_i = \bar{A}$. Therefore,

$$P(B_i) \leqslant P(\bar{A}) \leqslant \sum_{i=1} P(B_i). \tag{11}$$

An illustrative example showing B_1 and B_2 in two dimensions is given in Fig. 3 for dedicated and fully shared access. Note that in fully shared access $B_1 \cup B_2 = \bar{A}$. Fig. 3 suggests the (obvious) fully shared ordering property (Ref. 5).

Order the user classes $i = 1, 2, \cdots, K$ so that $c_1 \leqslant \cdots \leqslant c_K$. Then, for fully shared access, the blocking states are ordered $B_i \subseteq B_{i+1}$, and consequently $P(B_i) \leqslant P(B_{i+1})$. (A simple proof of this property is provided in Ref. 2.) Physically, one would expect that those users requiring more capacity would be blocked more often. But note that this *ordering is independent* of the distribution in traffic levels a_i.

An important feature for finite population sources ($N_i < \infty$) is the difference between the probability of being blocked as seen by any given arriving customer (call congestion) in contrast to the probability of full capacity usage (time congestion). The calculations for $P(\mathbf{j})$ correspond to time congestion. For finite populations, an arriving customer implies knowledge that at least one traffic source (himself) is inactive. Thus, the active population in class i is reduced by one to $N_i - 1$. (For infinite populations the distinction is unnecessary.)

To calculate call congestion for a finite user population class i a simple guess would be to modify the marginal probability $Q_i(j_i)$ by replacing N_i with $N_i - 1$ (an $N_i - 1$ source model). In Ref. 7 this property is shown to hold for a one-dimensional birth-death process with quasi-random arrivals and general holding times (not necessarily exponential) and queueing strategies (e.g., Blocked Calls Held). However, Ref. 7 shows by example that for multidimensional quasi-random birth-death processes this desired property need not hold. Paralleling Ref. 7, it was verified (Ref. 2) that for exponentially distributed traffic and blocked-calls-cleared, the call congestion probabilities (arriving calls) can be computed by replacing N_i with $N_i - 1$.

F. Numerical Example, Fully Shared vs. Dedicated

A numerical comparison is provided here between fully shared ($A = \Omega$) and dedicated access. Computation for dedicated access is quite simple; use the Erlang B formula K times. Fully shared access, however, requires a computational algorithm.

There are computational challenges and limitations to evaluating the simple form of Eq. 1 for general A-sets. Serious problems can arise for the number of user classes K exceeding 4 and coping with large C_0. Both of these factors increase the number of $\mathbf{j}$ points in the space $A \subseteq \Omega$, resulting in a very rapid increase in computational steps, especially in determining the B_i. This issue is clearly appreciated by recent work in queueing networks (Refs. 8 through 14). However, models in queueing networks assume that all $c_i \equiv 1$, and $A \equiv \Omega$ which allows summing (Ref. 8) over Ω recursively along disjoint parallel planes as shown in Fig. 4. All $\mathbf{j}$ points are "captured"

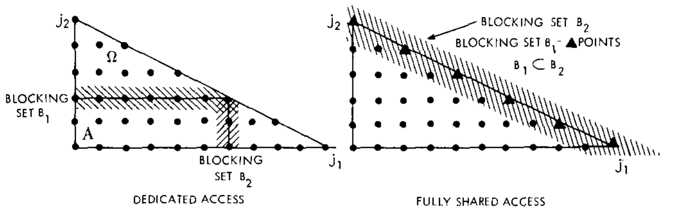

Fig. 3. Blocking Sets.

only because the c_i are equal. This summing technique (Ref. 9) considerably reduces the number of computational steps. If the c_i are integer valued, the state space can be augmented with dummy states with zero probability in order to obtain "equivalent" equal c_i and allow using the planar sum technique. However, since the number of state points is increased, it is not clear that any significant reduction in computational steps will be achieved.

Another computing technique that might be of interest is contained in Refs. 10 and 11. Conventional circuit theory is reviewed for possible application to queueing networks. These approaches may have more utility to the problem here, in that they do not depend inherently on the equality of the c_i.

In order to reduce the computational burden for fully shared access, $P(B_K)$ was used as an upper bound on the blocking probabilities, $P(B_i)$, where the user classes are ordered such that $c_1 \leqslant c_2 \leqslant c_3 \leqslant c_K$ and use is made of the order property on $P(B_i)$ for fully shared access. (Note that B_K is equal to the upper boundary of A, $\bar{A} \equiv \bar{\Omega}$, since $\bar{A} = \cup_i B_i = \bar{B}_K$.)

A computer program was written based on Eqs. 1, 10, and 11 in order to compare the dedicated and fully shared strategies with respect to the amount of total capacity, C_0, needed to provide a specified probability of blocking at a given traffic level. Three user classes are permitted (e.g., $K \leqslant 3$, small, medium, and large capacity terminals). For each user class, the traffic parameters (N_i, $a_i = \lambda_i/\mu_i$) and per circuit capacity need (c_i) as well as the blocking objective or grade of service (GOS) are specified. Using the conventional one-dimension Erlang B (or Engset) equations, the dedicated capacity C_i, needed for each user class to achieve the GOS objective is calculated. The total capacity, C_0, is then the sum of the C_i.

Using the C_0 value calculated for the dedicated strategy, the blocking probabilities that result with the fully shared strategy are then calculated. In all cases this results in a blocking probability which is less than the GOS objective. The initial value of C_0 is then reduced until the blocking objective is just met. Note that the examples assume that traffic parameters are known (i.e., measurable) and that capacity is allotted in a manner to *match* the traffic need. (Reference 1 provides other numerical examples of interest comparing traffic balance, heavy traffic and GOS dependence.)

As a purely hypothetical example, a mix of Naval task force users was postulated. Three classes of circuits were modeled as follows:

Class 1—Leader Type Class: Representative of circuits required by medium flagships (Destroyer Leader) with moderate traffic parameter a_1 and moderate bit rate/circuit and moderate terminal EIRP and G/T.

Class 2—Fleet Element Type Class: Typical circuits required by force element ship (Destroyer) with low a_2 and low bit rate/circuit and G/T.

Class 3—Major Flag Class: Typical circuits required by Major flag-type ships (Carrier or Cruiser Flagship) with high-duty factor a_3, high G/T and high bit rate.

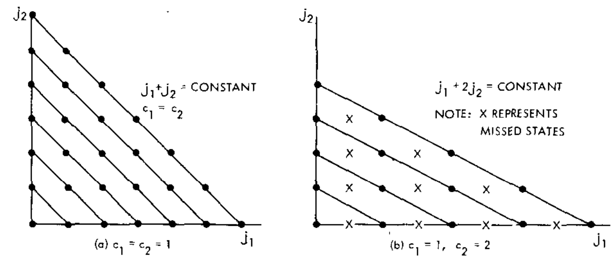

Fig. 4. Planar States Missed.

TABLE 1
EXAMPLE PARAMETERS FOR SATELLITE CIRCUIT SWITCHED SYSTEM

Case	Terminal Class: 1 Medium Flag	2 Fleet Element	3 Major Flag	Traffic Level*
No. of Sources	N_3	$8N_3$	N_3	--
A. c_i	1	4	4	$.37N_3$
a_i	0.1	0.01	0.25	
B. c_i	1	1	4	$.37N_3$
a_i	0.1	0.01	0.25	
C. c_i	1	1	4	$1.02N_3$
a_i	0.1	0.1	0.25	
D. c_i	1	4	4	$1.02N_3$
a_i	0.1	0.1	0.25	

*In Erlangs, $\Sigma\, a_i N_i/(1+a_i)$

In a task force, one might expect overall user traffic would tend to correlate with the number of sources N_3 on a Major flagship.** Consequently, required satellite capacity for dedicated versus fully shared allocation was computed as a function of the number of Major flag sources N_3. It is postulated that $N_1 = N_3$ and $N_2 = 8N_3$. The cases computed are summarized in Table 1. Case A represents the nominal parameter values. Case B, having the exact same traffic level as Case A, examines the consequences of improving the capability of the Fleet Element terminals to that of the Medium Flag by reducing c_2 from 4 to 1. Case C postulates the improved Fleet Element terminal capability of Case B but raises the traffic level by increasing the Fleet element call intensity, a_2, to that of a Medium Flag. Case D examines the effect of traffic elevated per Case C but with a terminal capacity of Case A.

The results for Cases A and B are plotted in Fig. 5; Cases C and D are plotted in Fig. 6. The plots show as a function of Major Flag traffic sources N_3, the amount of total capacity C_0 needed for a 5 percent grade of service with fully shared and dedicated access strategies. Also shown is the system utilization or percent of capacity in usage on the average. Also overlaid on Fig. 6 for purposes of comparing capacity need is Case A from Fig. 5.

Due to the discrete nature of circuits a granularity effect is

** There may be more than one user circuit per "flagship" and/or more than one flagship per "task force."

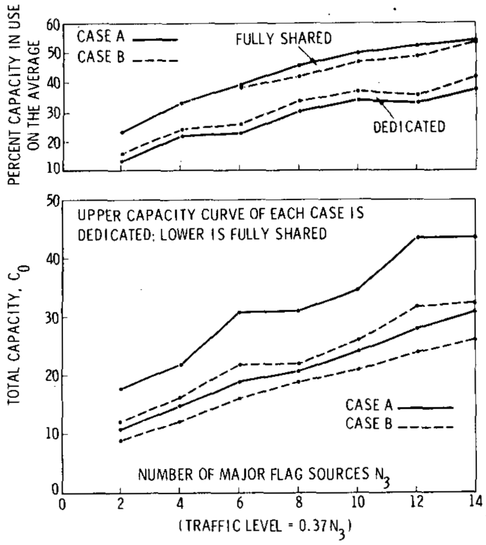

Fig. 5. Capacity and Utilization Cases A and B.

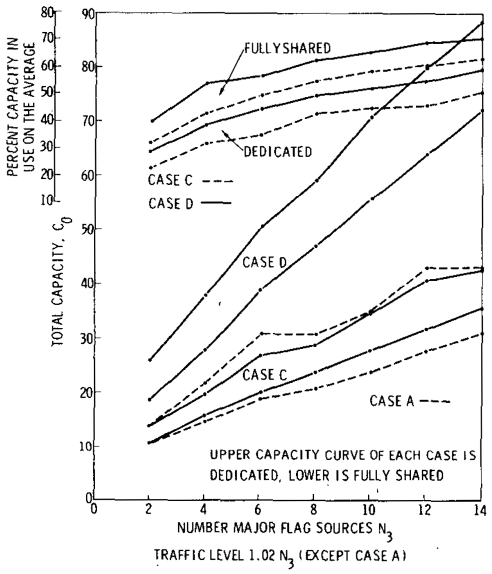

Fig. 6. Capacity and Utilization Cases C and D.

manifested in the C_0 curve especially for the dedicated strategy at low traffic levels. The fully shared strategy tends to average the needs over all users and thus smooths the C_0 curve. At higher traffic levels, the dedicated curve also becomes smoother. This effect on the utilization curve can even manifest as a slight negative slope over a small portion of the curve.

Given an available capacity limit, the curves also determine the number of sources that can be accommodated at a grade of service. At low levels of capacity and traffic as in Case A, the difference between fully shared and dedicated is quite dramatic. But notice that improving the Fleet Element terminal in Case B reduces the advantage. At higher traffic levels, as in Cases C and D, the strategy differences are lessened. The advantages of fully shared access to terminal aggregates with low levels of traffic are obvious. They are less so as traffic builds, provided adequate satellite capacity is available.

Another interesting comparison is the effect of varying G/T of the most numerous terminal (the Fleet Element). For fully shared strategy at low traffic levels (Cases A and B) the difference is small and the cost would probably not be warranted. *But* if traffic were to have a rather busy period involving the Fleet Element, then the difference is quite dramatic as shown

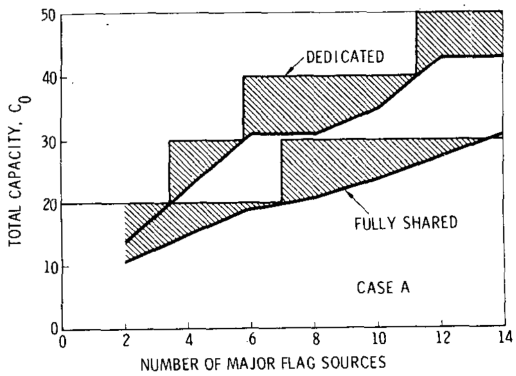

Fig. 7. Capacity Loss with Allocation by "Transponder".

in Fig. 6. Since terminal antennas are fixed quantities, these results point up *quantitatively* the system planning dilemma as to resource availability (satellite capacity), unpredictable communication load or growth thereto, and investment cost. In the alternative, information bandwidth could be (would have to be) reduced.

Finally, Fig. 7 (a reproduction of Case A) demonstrates the effects of allocating capacity in discrete steps. The least efficient use would be to permanently assign a fixed amount of capacity (e.g., a satellite transponder) to a user community with average traffic load well below what could be accommodated. Allocating transponders as opposed to circuits generates larger increments of capacity allocation. As an example, suppose a transponder would represent a capacity increment of 10 c_1 units. Consider Case A and start with 2 transponders (20 c_1 circuits) and then add capacity in whole transponders (10 c_1 circuit increments). This would generate the indicated staircase use of capacity. The shaded region shows the loss of capacity in allocating a whole transponder at a time.

III. CONCLUSION

A simple birth-death analytic model was presented for circuit-switched demand access systems. It was shown that a geometrical relationship exists between user access constraints, capacity, and user traffic needs. Access constraints directly affect call-blocking characteristics. These constraints can be represented as a solid body resting on the positive quadrant of K space. For blocked-calls-cleared, the variation in grade of service is then primarily determined by the upper boundaries of this body. Although the steady-state probability of the calls in progress has a product form, the user classes are statistically dependent unless they are provided dedicated access.

Much is left to be done. Of great importance is the extension to nonlinear functions of capacity in use, $C(j)$. A computationally efficient means of evaluating Eq. 1 and finding the blocking sets B_i is needed. Can the geometric approach be fruitfully extended to blocked-calls-delayed systems?

Within the confines of the linear blocked-calls-cleared system there is a need for the following: (a) theorems which characterize the properties of access constraints, such as optimal A-sets in the limits of heavy and light traffic.*** (b) introduction of interclass user correlation, (c) introduction of more general call-holding distributions (Markov properties are lost with nonexponentially-distributed arrival processes). Last,

*** It is interesting to note that preliminary results of Ref. 15 indicate that in the diffusion limit the fully shared strategy need not be optimal.

and perhaps least, the class of A-sets with product form $P(j)$ should be extended beyond those with coordinate-convexity.

REFERENCES

1. J. M. Aein and O. S. Kosovych, "Satellite Capacity Allocation," *Special Issue of the Proceedings of the IEEE,* vol. *65,* No. 3, March 1977, pp. 332-342.
2. J. M. Aein and O. S. Kosovych, "On Capacity Allocation Strategies For DoD Communication Satellites," IDA Paper, P-1155, January 1976.
3. J. M. Aein and O. S. Kosovych, "Efficient Transponder Utilization," Presented at EASCON 76, September 1976.
4. G. F. W. Fredrickson, "Analysis of Channel Utilization in Traffic Concentrators," *IEEE Trans. on Comm.* Vol. *COM-22,* No. 8, August 1974.
5. G. Frenkel, "The Grade of Service in Multiple-Access Satellite Communications Systems with Demand Assignments," *IEEE Trans. on Comm.,* Vol. *COM-22,* No. 10, October 1974, pp. 1681-1685.
6. W. Feller, *An Introduction to Probability Theory and its Applications,* Vol. *1,* John Wiley & Sons, New York, 1957.
7. R. B. Cooper, *Introduction to Queueing Theory,* MacMillan Company, New York, 1972.
8. J. P. Buzen, "Computational Algorithms for Closed Queueing Networks with Exponential Servers," *Communications of the ACM,* Vol *16,* No. 9, September 1973, pp. 527-531.
9. IBM Research Center, *Numerical Methods in Queueing Networks,* RC-5344, M. Reiser and H. Kobayashi, 27 March 1975.
10. K. M. Chandy, U. Herzog, and L. Woo, "Parametric Analysis of Queueing Networks," *IBM J. Res. Develop.,* January 1975, pp. 36-42.
11. K. M. Chandy, U. Herzog, and L. Woo, "Approximate Analysis of General Queueing Networks," *IBM J. Res. Develop.,* January 1975, pp. 43-49.
12. Department of Industrial Engineering, University of Michigan, Ann Arbor, *Network Models for Large-Scale Time-Sharing Systems,* Ph.D. Thesis, IR-71-1, C. G. Moore, III, April 1971.
13. IBM Research Center, *Horner's Rule for the Evaluation of General Closed Queueing Networks,* RC-5219, M. Reiser and H. Kobayashi, 16 January 1975.
14. H. Kobayashi, "Application of the Diffusion Approximation to Queueing Networks, I: Equilibrium Queue Distributions, II: Nonequilibrium Distributions and Applications to Computer Modeling," *J. Assoc. Comp. Mach.,* Vol. *21,* No. 2, April 1974, pp. 316-328, and No. 3, July 1974, pp. 459-469.
15. G. Gopinath, G. Foschini, and J. Hayes, "On Circuit Allocation for Satellite Communication Systems," talk presented at U.R.S.I. Meeting, Commission C–Session 1, Stanford University, Stanford, Connecticut, June 23, 1977.

Differential Encoding for Multiple Amplitude and Phase Shift Keying Systems

WILLIAM J. WEBER, III, MEMBER, IEEE

***Abstract*–Because of the symmetry in most two-dimensional signal constellations, ambiguities exist at the receiver as to the exact phase orientation of the received signal set. In PSK systems, this ambiguity is resolved by the use of differential encoding. This paper presents differential encoding techniques which can be used with a variety of symmetric signal sets to remove their phase ambiguity. While not proven to be optimum, the techniques do have low performance penalties relative to the uncoded performance. The key to reducing the performance penalty is to use the minimum amount of differential encoding necessary to resolve the ambiguity. Examples of encoding techniques for several common signal constellations are given, including their performance penalties.**

Paper approved by the Editor for Data Communication Systems of the IEEE Communications Society for publication without oral presentation. Manuscript received August 26, 1977; revised November 9, 1977. This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology under Contract NAS-7-100 sponsored by the National Aeronautics and Space Administration.

The author is with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91103.

I. INTRODUCTION

The recent increase in research in the area of bandwidth conserving modulation techniques has led to proposals for systems with a variety of two-dimensional signal structures. References 1 and 2 give good examples of the types of signal constellations which are possible in two dimensions. The signal design problem is one involving many tradeoffs of factors such as bandwidth reduction, power increase, complexity, and suitability for the user's channel. The very nature of this signal design, which is really a sphere packing type of problem, dictates the use of very symmetric patterns to reduce power requirements or, conversely, lower the probability of error. A problem, however, arises at the receiver in that the receiver can recognize the pattern of the signal points but cannot distinguish between the various symmetric phase orientations of the signal set. An L-fold rotational symmetry can be defined as a signal set for which the signal set pattern remains unchanged after a rotation of $\pm I \cdot (2\pi/L)$ radians, where I and L are integers. Thus, given $M = 2^K$ signal points, the receiver cannot properly assign the K bits to each detected signal point without first resolving the L-fold ambiguity. The simplest example, of course, is binary PSK. The receiver needs some absolute way of distinguishing between the two equal and opposite signal points before it can assign a binary one to one phase and a zero to the other. This ambiguity can be resolved in three ways: (1) a constant reference signal of some kind can be transmitted along with the modulated data signal; (2) an acquisition signal and/or the periodic insertion of a synchronization sequence can be used in the data stream; or (3) the use of differential encoding. The first two techniques have rather obvious advantages and disadvantages. The third technique, which is discussed in this paper, has the advantage that by properly encoding and decoding the signal points, the proper bit detection takes place regardless of rotational phase ambiguities. The disadvantage is that this coding procedure increases the probability of bit error over the conventional uncoded case. In the binary PSK case mentioned above the probability is increased by a factor of two at high signal to noise ratios.

The differential encoding problem is really a subset of the more general problem of the assignment of the blocks of K bits to the 2^K signal points. In the uncoded form, this assignment is a straight one-to-one mapping of the distinct K-bit blocks to particular signal points. This assignment is generally not made haphazardly, but rather made to minimize the number of bit errors given a symbol error. That is, symbol errors are most likely to occur between nearest neighbors or, put another way, the probability of error is dominated by the minimum distance, δ_{min}, between signal points. Thus, the assignment of bits to a signal point must be made such that its nearest neighbors differ in as few bits as possible. Gray coding [3] is an example in which each adjacent level or signal point differs by only one bit in a K bit block. This one bit difference